PCCP

Expanded ensemble predictions of toluene–water partition coefficients in the SAMPL9 LogP challenge

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-09-2024-003621.R1
Article Type:	Paper
Date Submitted by the Author:	22-Jan-2025
Complete List of Authors:	Goold, Steven; Temple University, Department of Chemistry Raddi, Robert; Temple University, Department of Chemistry Voelz, Vincent; Temple University, Department of Chemistry

PCCP

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Expanded ensemble predictions of toluene–water partition coefficients in the SAMPL9 LogP challenge

Steven R. Goold,[a] Robert M. Raddi,[a] and Vincent A. Voelz*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

The logarithm of the partition coefficient (logP) between water and a nonpolar solvent is useful for characterizing a small molecule's hydrophobicity. For example, the water-octanol logP is often used as a predictor of a drug's lipophilicity and/or membrane permeability, good indicators of its bioavailability. Existing computational predictors of water-octanol logP are generally very accurate due to the wealth of experimental measurements, but may be less so for other non-polar solvents such as toluene. In this work, we participate in a Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) logP challenge to examine the accuracy of a molecular simulation-based absolute free energy approach to predict water-toluene logP in a blind test for sixteen drug-like compounds with acid-base properties. Our simulation workflow used the OpenFF 2.0.0 force field, and an expanded ensemble (EE) method for free energy estimation, which enables efficient parallelization over multiple distributed computing clients for enhanced sampling. The EE method uses Wang-Landau flat-histogram sampling to estimate the free energy of decoupling in each solvent, and can be performed in a single simulation. Our protocol also includes a step to optimize the schedule of alchemical intermediates in each decoupling. The results show that our EE workflow is able to accurately predict free energies of transfer, achieving an RMSD of 2.26 kcal/mol (1.65 logP units), and R^2 of 0.80. An examination of outliers suggests that improved force field parameters could achieve better accuracy. Overall, our results suggest that expanded ensemble free energy calculations provide reasonably accurate logP prediction for a general-purpose force field.

1 Introduction

The logarithm of the partition coefficient (logP) between water and a nonpolar solvent is a highly useful molecular property in medicinal chemistry and pharmacology. LogP measurements for water/octanol partitioning are commonly used to characterize the lipophilicity of drug-like molecules, which can strongly influence their bioavailability and affinity for their targets. Therefore, there is great interest in developing accurate computational methods to predict logP, either by empirical or physics-based methods.[1]

For several years, the Statistical Assessment of Modeling of Proteins and Ligands (SAMPL) Challenges have provided an opportunity to assess state-of-the-art methods for logP prediction, through blind prediction of unpublished logP measurements.[1–3] SAMPL has also hosted blind prediction challenges for host–guest affinities,[4–8] pK_a prediction,[9,10] and distribution coefficients (logD).[11]

Based on the results of the last two SAMPL LogP challenges (SAMPL6[3] and SAMPL7[1]), the state-of-the-art for predicting water-octanol logP values is within a mean absolute error (MAE) of about 0.5 logP units, which can be achieved with empirical machine learning approaches such as the multiple linear regression model TFE-MLR[12] or neural network models such as D-MPNN[13] or ClassicalGSG[14], or with quantum chemistry approaches such as COSMO-RS.[15]

Preliminary results from the current SAMPL9 LogP challenge, consisting of water-toluene partition coefficients for sixteen drug-like compounds with acid-base properties, suggest similar findings: empirical machine learning approaches[16] and quantum chemistry approaches[17] were able to achieve the best predictions, with an MAE of less than 1 logP unit. Interestingly, while molecular mechanics approaches have historically performed less accurately, the best root-mean-squared error (RMSE) prediction in the SAMPL9 challenge was made using an MM/PBSA approach in which surface tension parameters in the SA (surface area) model were tuned to reproduce experimental solvation free energies.[18]

In this article, we present the results of a molecular simulation-based absolute free energy approach to predict water-toluene logP in SAMPL9. One goal of this work is to test the performance of expanded ensemble free energy methods for this purpose. Another goal is assess the accuracy of the general-purpose molecular force field used, OpenFF 2.0. This model has been parameterized using QM energy surfaces and experimental thermochemi-

[a] Department of Chemistry, Temple University, Philadelphia, PA, USA.
* Corresponding Author. E-mail: voelz@temple.edu.

cal measurements, but not logP values in particular. Assessments such as these are crucial in assessing the state-of-the-art, as well as understanding particular shortcomings in these models which could be improved.

1.1 An expanded ensemble free energy approach to logP prediction

In our expanded ensemble (EE) free energy method, a double-decoupling approach is used to compute the free energy of transfer from vacuum to solvent, through an alchemical transformation in which the nonbonded interactions (electrostatics and van der Waals) are turned off. Given estimated values of ΔG_{tol} and ΔG_{w}, the solvation free energies in toluene and water, respectively, the toluene partition coefficient is calculated as

$$\log_{10} P_{\mathrm{tol/w}} = -\frac{\Delta G_{\mathrm{tol}} - \Delta G_{\mathrm{w}}}{RT(\ln 10)} \quad (1)$$

Expanded Ensemble method. The EE method used here is described in previous works.[19–21] The key feature of EE is the ability to sample multiple thermodynamic ensembles in a single simulation. A coupling parameter $\lambda_i \in [0,1]$ is used to define a series of $i = 1,\ldots,N$ thermodynamic ensembles where $\lambda_1 = 0$ defines an ensemble with fully-scaled nonbonded interactions, and $\lambda_N = 1$ defines an ensemble where the nonbonded interactions are scaled to zero. Soft-core potentials are used to avoid numerical singularities.

Throughout an EE simulation, a Markov Chain Monte Carlo (MCMC) procedure is used to accept or reject moves between thermodynamic ensembles defined by λ_i and λ_j. The Wang-Landau flat-histogram method[22,23] is used to adaptively learn the values of constant biases $-f_i$ which, when applied to each thermodynamic ensemble i, results in equal probabilities for $i \to j$ and $j \to i$ transitions. When this is achieved, the free energy of the $\lambda = 0 \to 1$ transformation is estimated as $f_N - f_1$. The EE algorithm is available in the GROMACS simulation package.[24] Recent extensions of the EE method have been proposed that combine replica exchange with expanded ensemble sampling,[25,26] but we do not utilize those approaches here.

Because the EE method estimates free energies using a single simulation replica, it is ideally suited for distributed computing applications with an asynchronous client-server model. Our group has recently leveraged Folding@home[27] to perform massively parallel virtual screening using EE methods.[20,21] Through this work, several methodological issues with have been identified, which are ongoing challenges to be actively addressed. One issue is that the Wang-Landau flat-histogram method results in "saturation of the error" that can lead to premature convergence for fast learning rates.[23] This issue can be compounded by slow convergence due to high-energy barrier conformational transitions in the molecules being decoupled. To deal with these issues, we have found that convergence times and uncertainties of free energies can be estimated by simulating multiple independent EE trajectories.[19,20]

The other major issue with the EE approach is its sensitivity to the chosen schedule of λ_i values. Poor selection of these values can lead to poor MCMC acceptance rates, which in turn causes the simulation to need more time to converge (and potentially error-saturate). To choose optimal schedules for λ_i, we have devised the *pylambdaopt* algorithm,[20] which uses a preliminary round of sampling to infer λ_i values that maximize transition rates between all neighboring thermodynamic ensembles.

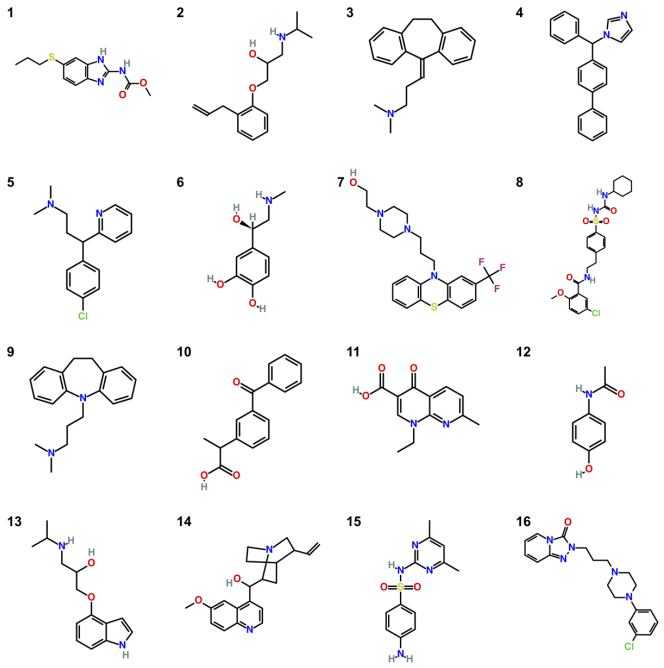

Fig. 1 Molecular structures of the molecules in the SAMPL9 logP blind challenge: (1) Albendazole, (2) Alprenolol, (3) Amitriptyline, (4) Bifonazole, (5) Chlorpheniramine, (6) Epinephrine, (7) Fluphenazine, (8) Glyburide, (9) Imipramine, (10) Ketoprofen, (11) Nalidixic acid, (12) Paracetamol, (13) Pindolol, (14) Quinine, (15) Sulfamethazine and (16) Trazodone .[28]

2 Methods

We performed blind predictions of $\log_{10} P_{\mathrm{tol/w}}$ for the sixteen molecules shown in Figure 1 as part of the SAMPL9 logP challenge. A three-part workflow was implemented to (1) prepare systems, (2) perform expanded ensemble simulations on Folding@home and Temple University high-performance computing (HPC) cluster, and (3) analyze the results. All simulations were preformed using GROMACS 2020.3 or GROMACS 2020.4.[24]

2.1 System preparation

Molecular topologies. SMILES strings provided by SAMPL9 were converted to three-dimensional chemical structures using the Openeye toolkit[29]. From these, molecule topologies using the OpenFF 2.0.0 force field[30] were constructed using the Open Force Field Toolkit.[31] Partial charges were assigned using AM1-BCC.[32] For simulations in aqueous solvent, the TIP3P water model was used.[33] For simulations in toluene, OpenFF 2.0.0 was used to parameterize the toluene molecule.

Simulation Preparation. An initial solvent box volume of (6 nm)^3 was filled with 7029 water molecules, and an initial solvent

box volume of $(3.38\ \text{nm})^3$ was filled with 216 toluene molecules. Solute molecules were inserted into each system and restrained in the center of the box using a harmonic restraint of 1000 kJ nm^{-2}. Systems were energy-minimized using steepest descent before undergoing 200 ps of NVT simulation followed by 200 ps of NPT simulation. All simulations were performed at a temperature of 298.15 K, using velocity Verlet integration with a 2.0-fs time step and with a velocity-rescaling thermostat. The sole exception was for quinine solutes, which used a 1.0-fs time step to avoid instabilities. NPT simulations used Berendsen pressure coupling. These steps were facilitated through in-house scripts utilizing the GromacsWrapper package.[34]

Optimization of lambda values. After NPT equilibration, a short EE simulation was run using a Metropolized-Gibbs MCMC criterion, with attempted moves restricted to nearest-neighbors ($i \to i \pm 1$). The starting bias was 10 k_BT where k_B is the Boltzmann constant and $T = 298.15$ K is the temperature. This choice of bias increment is designed to is to penalize staying in any one alchemical intermediate for very long, thus forcing sampling across all alchemical intermediates. Twenty lambda values were used, according to an initial schedule found to work satisfactorily in previous work.[20] Simulation snapshots were saved every 2 ps, with moves to neighboring ensembles attempted every 0.5 ps. Samples of energy differences of snapshots between current and neighboring ensembles ($\Delta u_{i,i-1}$ and $\Delta u_{i,i+1}$, stored in the `dhdl.xvg` output file of GROMACS) were used as input to the *pylambdaopt* algorithm to obtain an optimized schedule of lambda values.[35]

2.2 Production simulation

Production simulations were performed in GROMACS on the Folding@Home platform, with 100 independent EE replicas (with different randomized initial velocities) per calculation. All production runs used a velocity Verlet integrator with a 2 fs time step, with the exception of simulations for quinine, which used a time step of 1 fs , and for which 5 EE replicas were simulated each for 200 ns on the Owlsnest HPC cluster. A velocity-rescaling thermostat was used at a temperature of 298.15 K, and a Berendsen barostat with 2 ps time constant was used at a pressure of 1 bar. Particle-Mesh Ewald electrostatics were used (*pme-order* = 4, *fourierspacing* = 0.10) with nonbonded cutoffs of 0.9 nm and a long-range dispersion correction. Hydrogen bonds were constrained using the LINCS algorithm, and soft-core Lennard-Jones (LJ) interaction potentials were used according the formulation of Beutler et al.[36] In this scheme, a transformation between LJ potentials $V^A(r)$ and $V^B(r)$ described by alchemical parameter $\lambda = 0 \to 1$ is given by the "soft-core" potential $V_{sc}(r) = (1-\lambda)V^A(r_A) + \lambda V^B(r_B)$, where $r_A = (\alpha\sigma_A^6\lambda^p + r^6)^{\frac{1}{6}}$ and $r_B = (\alpha\sigma_B^6(1-\lambda)^p + r^6)^{\frac{1}{6}}$ are shifted distances designed to avoid the singularity at $r = 0$. Here, α is the soft-core parameter controlling the magnitude of the shift, σ_A and σ_B are interaction radii, and p is the soft-core power. These values were set to $\alpha = 0.5$, $p = 1$ and $\sigma_A = \sigma_B = 0.3$ nm, respectively, using GROMACS settings *sc-alpha* = 0.5, *sc-power* = 1, *sc-sigma* = 0.3). Coordinates and energies were saved every 50 ps.

The EE protocol used Wang-Landau flat-histogram sampling[22] with an initial bias increment of $\delta = 10\ k_BT$. When the histogram counts h_i for thermodynamic states i satisfied $\eta \leq h_i/\bar{h} \leq \eta^{-1}$ (where $\eta = 0.7$ and $\bar{h} = (1/N)\sum_{i=1}^{N} h_i$), the bias increment was scaled by 0.8 and all histogram counts were reset to zero. Scaling of the bias increment was discontinued when $\delta < 10^{-5}$.

2.3 Analysis of EE simulations

EE simulations were considered to have converged when the bias increment reached a value of 0.02 k_BT. Free energies for each EE replica were estimated as the sample mean of estimates collected after this convergence point. Final estimates of ΔG_{tol} and ΔG_{w} and their uncertainties were calculated as the sample means and standard deviations across the EE replicas. Convergence was typically reached within 50–100 ns of simulation, and trajectory lengths typically reached 100–200 ns (Figure 2). Unlike EE simulations performed in the SAMPL9 host-guest challenge,[20] and for relative binding free energy calculations,[19] we do not observe any significant convergence problems that can arise from slow conformational dynamics. This could be due to the particular set of molecules considered, or to our improved methods for optimizing alchemical intermediates.[35]

2.4 Quantum Mechanical Calculations

To better understand the force field accuracy for fluphenazine, quinine and trazodone, density functional theory (DFT) geometry optimization was performed on simulation snapshots, using the B3LYP functional[37] and cc-DZVP level of theory.[38] Calculations were performed *in vacuo* using WebMO[39] with the Gaussian 16 Revision A.03 engine.[40]

3 Results and Discussion

3.1 EE accurately predicts toluene–water partition coefficients

Predicted transfer free energies from water to toluene, $\Delta G = \Delta G_{\text{tol}} - \Delta G_{\text{w}}$, show accurate agreement with experimental measurements made by Zamora et al.[16] (Figure 3). Our submitted predictions for SAMPL9 (January 30, 2023) had a root-mean-squared deviation (RMSD) of 2.26 kcal/mol (1.65 logP units), a mean signed error (MSE) of –1.09 kcal/mol (0.80 logP units), a mean unsigned error (MUE) of 1.75 kcal/mol (1.28 logP units), and a correlation coefficient of $R^2 = 0.80$.

Due to the time constraints of the SAMPL9 challenge, we were unable to obtain a full set of trajectory data before the submission deadline. After nearly one month of additional simulation, we performed a final analysis (February 18, 2023), for which we report predicted solvation free energies, transfer free energies, logP values and their uncertainties (Table 1) The quality of the results did not significantly change with the additional simulation. Our final results had a root-mean-squared deviation (RMSD) of 2.29 kcal/mol (1.68 logP units), a mean signed error (MSE) of –1.16 kcal/mol (0.85 logP units), a mean unsigned error (MUE) of 1.77 kcal/mol (1.30 logP units), and a correlation coefficient of $R^2 = 0.81$.

Compared to the other blind predictions submitted by the

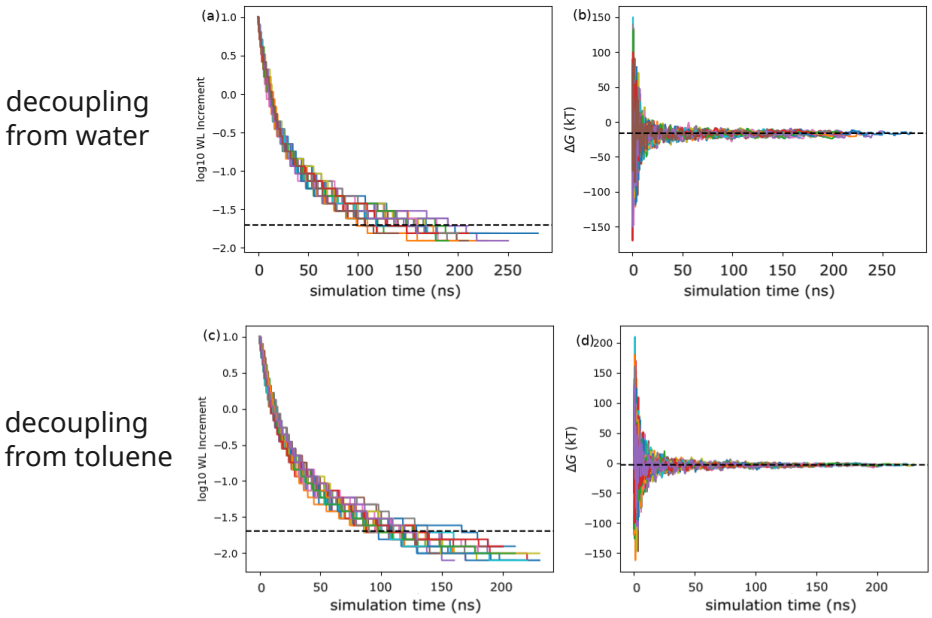

Fig. 2 Traces of (a) the Wang-Landau (WL) bias increment over time and (b) the decoupling free energy estimate over time for fifty independent EE trajectories of trazodone in water. Panels (c) and (d) show the corresponding traces for trazodone in toluene.

Table 1 Predicted solvation free energies, transfer free energies and $\log P_{\mathrm{tol/w}}$ values.

Molecule	ΔG_{tol}*	ΔG_{w}**	$\Delta G_{\mathrm{tol/w}}$†	$\Delta G^{\mathrm{exp}}_{\mathrm{tol/w}}$‡	$\log_{10} P_{\mathrm{tol/w}}$§	$\log_{10} P^{\mathrm{exp}}_{\mathrm{tol/w}}$‡
Albendazole	98.32 (0.78)	103.53 (0.26)	-5.20 (0.82)	-5.11	3.81 (0.60)	3.75
Alprenolol	-32.40 (0.66)	-26.22 (0.69)	-6.18 (0.96)	-3.26	4.53 (0.70)	2.39
Amitriptyline	-21.70 (0.93)	-13.22 (0.91)	-8.48 (1.30)	-7.49	6.22 (0.95)	5.49
Bifonazole	-25.16 (0.10)	-17.14 (0.13)	-8.02 (0.16)	-7.44	5.88 (0.12)	5.45
Chlorpheniramine	-16.78 (0.16)	-9.25 (0.13)	-7.54 (0.21)	-4.91	5.52 (0.15)	3.60
Epinephrine	-94.22 (0.31)	-96.23 (0.23)	2.01 (0.39)	1.67	-1.47 (0.29)	-1.22
Fluphenazine	-61.13 (0.74)	-51.24 (1.07)	-9.89 (1.30)	-5.94	7.25 (0.95)	4.35
Glyburide	110.93 (2.73)	115.71 (0.90)	-4.78 (2.87)	-3.79	3.50 (2.11)	2.78
Imipramine	-0.47 (0.21)	7.93 (0.29)	-8.39 (0.36)	-6.87	6.15 (0.26)	5.04
Ketoprofen	-34.51 (0.98)	-30.68 (0.08)	-3.83 (0.98)	-3.36	2.81 (0.72)	2.46
Nalidixic acid	180.29 (0.92)	180.30 (0.88)	-0.01 (1.27)	-1.99	0.01 (0.93)	1.46
Paracetamol	13.63 (1.10)	11.23 (0.85)	2.40 (1.39)	2.16	-1.76 (1.02)	-1.58
Pindolol	-25.12 (1.61)	-23.54 (1.52)	-1.58 (2.21)	-0.49	1.16 (1.62)	0.36
Quinine	-81.72 (1.23)	-74.38 (0.99)	-7.34 (1.58)	-1.92	5.38 (1.16)	1.41
Sulfamethazine	368.75 (1.21)	365.41 (1.23)	3.35 (1.72)	1.01	-2.45 (1.26)	-0.74
Trazodone	2.03 (1.10)	9.93 (1.60)	-7.90 (1.95)	-5.13	5.79 (1.43)	3.76

All ΔG values are in kcal mol^{-1}. Values in parentheses are uncertainties calculated as standard deviations across EE replicas.

* Predicted free energy solvation in toluene.

** Predicted free energy solvation in water.

† $\Delta G_{\mathrm{tol/w}} = \Delta G_{\mathrm{tol}} - \Delta G_{\mathrm{w}}$, with standard errors propagated as $\sigma_{\mathrm{tol/w}} = (\sigma^2_{\mathrm{tol}} + \sigma^2_{\mathrm{w}})^{1/2}$

‡ Experimental values from Zamora et al.[16]

§ $\log_{10} P_{\mathrm{tol/w}} = -\Delta G_{\mathrm{tol/w}}/RT \ln(10)$

participants in the SAMPL9 logP challenge, our EE predictions ranked eighth out of 18 total entries by RMSD (Figure 4a). As in the SAMPL7 logP challenge,[1] empirical methods and physics-based QM methods generally outperformed physics-based MM methods, although the top-ranked prediction (RMSD of 1.52 kcal/mol, 1.11 logP units) was from a MM-PBSA method.[41] Of the nine submitted predictions from physics-based MM methods, our EE predictions ranked third.

Our EE predictions use a general-purpose molecular mechanics force field with no special consideration or optimization for logP prediction. While we do not expect our results to compete with state-of-the-art logP predictors.[42] , our results suggest that

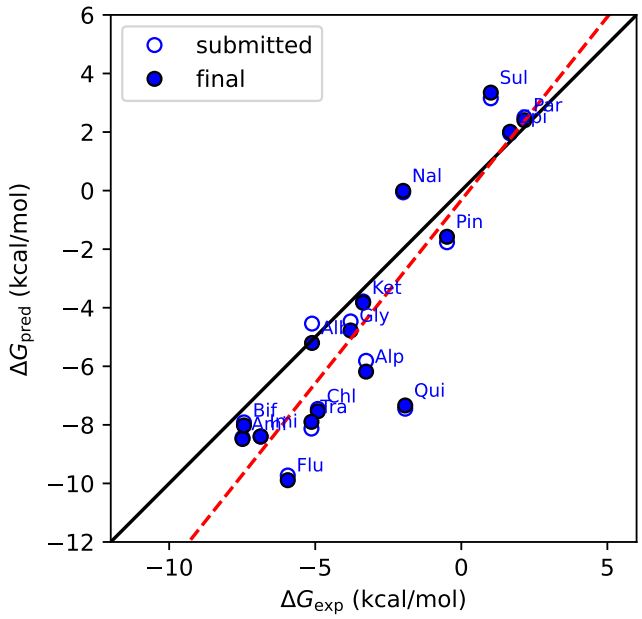

Fig. 3 Predicted free energies of transfer from water to toluene, ΔG_{pred}, versus experimental values, ΔG_{exp}, measured by Zamora et al.[16] Open circles represent our submitted SAMPL9 predictions; closed circles represent final predictions made after additional simulation. The red dashed line is a linear regression fit to the final results, and the black line represents perfect prediction.

the expanded ensemble free energy approach can provide reasonably accurate logP prediction when using a general-purpose force field.

Besides absolute prediction of partition coefficients, we also consider the ability of predictions to correctly rank-order the collection of logP values relative to each other. To quantify this, we used as a statistic the Spearman's rank correlation coefficient, r_s, which is defined as

$$r_s = 1 - \frac{6}{n\left(n^2-1\right)} \sum_{i=1}^{n} d_i^2, \tag{2}$$

where $n = 16$ is the number of ranked items (here, the number of molecules in the SAMPL9 logP challenge) and d_i is the integer difference between the predicted and actual rank.

For all the Physical MM entries submitted, we calculated the value of r_s and compared it with the null distribution, which was computed using 10^5 random permutations of rank orders (Figure 4b). The results show that MM methods with more accurate predictions tend to have more accurate rankings, although interestingly, none of the methods yielded rankings statistically significant enough to reject the null hypothesis (the smallest p-value calculated was 0.06, for the "MD (OPLS-AA/TIP4P)" submission).

3.2 An inspection of outliers reveals moderate force field inaccuracy for tertiary amines

An inspection of Figure 3 reveals that the three largest discrepancies between predicted and experimentally measured transfer free energies are for fluphenazine (7), quinine (14), and trazodone (16), with unsigned errors of 5.36, 6.31, and 4.35 kcal/mol, respectively. In all cases, EE predictions underestimate these transfer energies, which means that these molecules are predicted to more favorably partition into toluene than experimentally measurements show. A commonality shared by these outliers is the presence of non-aromatic nitrogen heterocycles. In contrast, bifonazole (4) and nalidixic acid (11) both contain aromatic nitrogen heterocyclics, and have predicted transfer free energies that are closer to experiment (0.48 and 1.93 kcal/mol unsigned error, respectively).

To examine the extent to which slow conformational sampling could be the cause of large discrepancies for these outliers, we examined the simulated trajectories for these molecules (data not shown). In all cases, dihedral angles incorporating non-aromatic nitrogens show chair-to-chair conversions on the 10–100 ns timescale. This observation, and the similar convergence across all EE simulations (see Figure 2), suggests that the discrepancies are not due to poorly converged sampling from slow conformational dynamics.

Next, we examined the possibility that the outliers might arise from inaccuracies in our chosen force field for tertiary amines. In their article describing the development and performance of the OpenFF Sage 2.0.0 force field, Boothroyd et al. (2023) mention large differences in improper torsion angles between MM- and QM-optimized minima.[30] The largest of these is for the nitrogen-centered improper i4, defined by SMIRKS string `([*:1]~[#7X3](*~[#6X3]):2](~[*:3])~[*:4]")` , although several others (i1, i3, i5) also show deviations. The i4 torsion parameter is difficult to generalize since it covers instances of both planar and pyramidal nitrogens. Of the three outliers, the i4 improper torsion is assigned for one of the nitrogens in fluphenazine, and two of the nitrogens in trazodone; it is not used for either of the quinine nitrogens.

To quantify the extent of nitrogen pyramidalization observed in the simulations of fluphenazine (7), quinine (14), and trazodone (16), we used the Dunitz χ_N parameter, defined as $\chi_N = \omega_1 - \omega_2 + 180°$, where ω_1 and ω_2 are two dihedral angles incorporating the (1,2)- and (1,3)-N-substituents, respectively (Figure 5a). For planar nitrogens, values of χ_N will be near zero, while for a perfectly tetrahedral nitrogen, χ_N values will be near $\pm 60°$.

For fluphenazine (Figure 5b), we found good agreement between the MM minima seen in EE simulations using OpenFF 2.0 and QM minima calculated using DFT (see Methods). The sampled distribution of χ_N for piperazine nitrogens in the EE simulations were peaked near $\pm 45°$, agreeing well with DFT geometry-optimized snapshots from these minima (Figure 5c). The sampled distribution of χ_N for the phenothiazine nitrogen showed planarity; the distribution was centered around zero with a standard deviation around 20°. DFT-optimized snapshots from these simulations indicate QM minima with $\chi_N \approx \pm 12°$ (Figure 5d).

For the nitrogen in the quinuclidine group of quinine, we found reasonably good agreement between MM and QM minima, which both showing a pyramidal nitrogen (Figure 6a). The distribution of χ_N is narrow ($\pm 10°$) and peaked around 55°, whereas DFT geometry-optimized snapshots have $\chi_N \approx 64°$.

The greatest disagreement in nitrogen pyramidalization be-

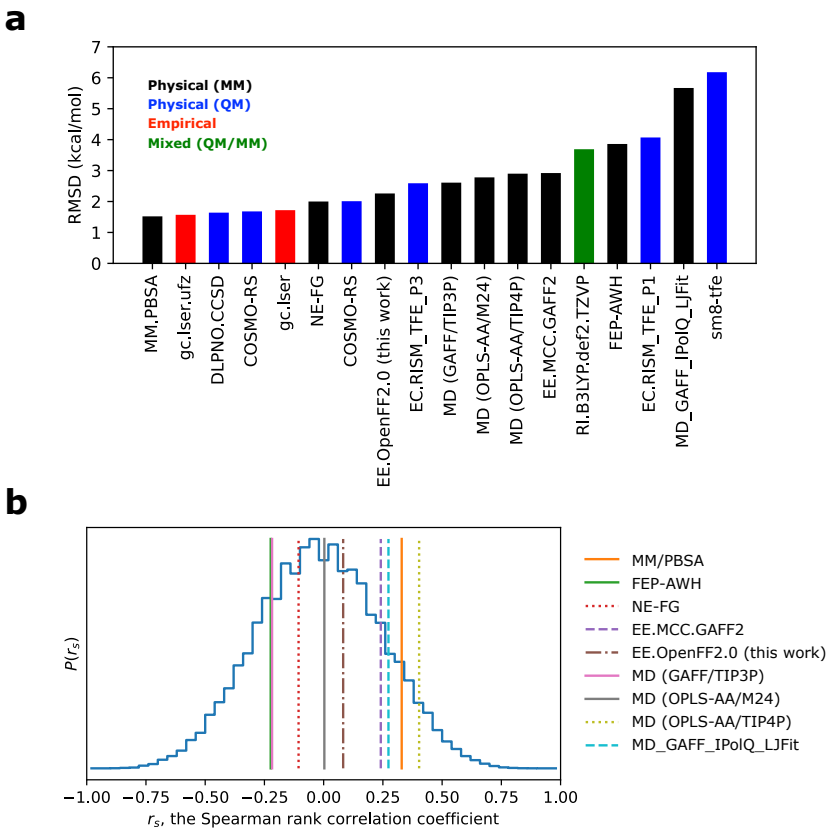

Fig. 4 Comparisons of the results of SAMPL9 logP submissions. (a) Prediction accuracies ranked by root-mean-squared deviations from experimental values. Colors denote the kind of method used in each prediction: (black) physical MM, (blue) physical QM, (red) empirical, and (green) mixed QM/MM methods. (b) Spearman's rank correlation coefficients r_s for Physical (MM) methods (vertical lines), plotted in relation to the computed null distribution (blue steps).

tween MM and QM minima was found for piperazine nitrogens in trazodone (Figure 6b). For the nitrogen with fully sp3-hybridized substituents, the sampled distribution of χ_N was peaked near $\pm 45°$, agreeing well with DFT geometry-optimized snapshots with χ_N near $+30°$ and $-38°$. For the nitrogen with the aromatic substituent, simulated χ_N distributions are broadly centered on zero, indicating a planar nitrogen. QM minima, however, suggest χ_N values near $\pm 53°$, suggesting a pyramidal nitrogen.

While these observations are anecdotal, they suggest that improved MM approaches to predicting partition coefficients and other properties may come from improvements in the bonded terms of force fields, specifically torsions. Boothroyd et al. (2023) note that incorrect puckering of small fused heterocycles is of particular concern, as it could lead to "erroneous intramolecular and intermolecular nonbonded interactions, especially in hydrogen bonding interactions and π-stacked configurations".[30] These are exactly the nonbonded interactions that dictate the extent to which molecules partition into nonpolar versus aqueous solvent. In their paper describing the parameterization and benchmarking of OpenFF 2.0.0, Boothroyd et al. (2023) note that the improper torsions were not optimized in the latest version, and remain the same as in OpenFF 1.0 "Parsley".[30] Inclusion of these terms for optimization in upcoming versions should be a priority. Boothroyd et al. also mention that terms such as the nitrogen-centered i4 impropers have small force constants ($\tilde{1}$ kcal/mol) that allow for a large range of geometries ranging from planar to pyramidal. Atom-typed force fields such as GAFF more strictly dictate the nitrogen geometry; OpenFF could similarly introduce more specific SMIRKS strings to capture a more specific hybridization scenarios.

Is it interesting to note that the distributions of Dunitz χ_N parameters are extremely similar for EE simulations performed in toluene versus water. This somewhat to be expected since geometry in molecular mechanics force fields is typically dominated by the bonded valence terms; also since EE simulations explore a range of thermodynamic ensembles that span the complete decoupling of non-bonded terms. For fluphenazine in water, we see a slight difference in the populations of positive and negative pyramidalization, but the locations of the low-energy conformational minima do not appear to change. We also see a slight decrease in the planar nitrogen population for trazadone in water versus toluene, but the effect is subtle.

In future work, it would be interesting to use our EE approach to compare logP predictions using general-purpose force fields like OpenFF and GAFF[43] against custom-fit potentials constructed using packages such as OpenFF BespokeFit[44] and AFFDO.[45]

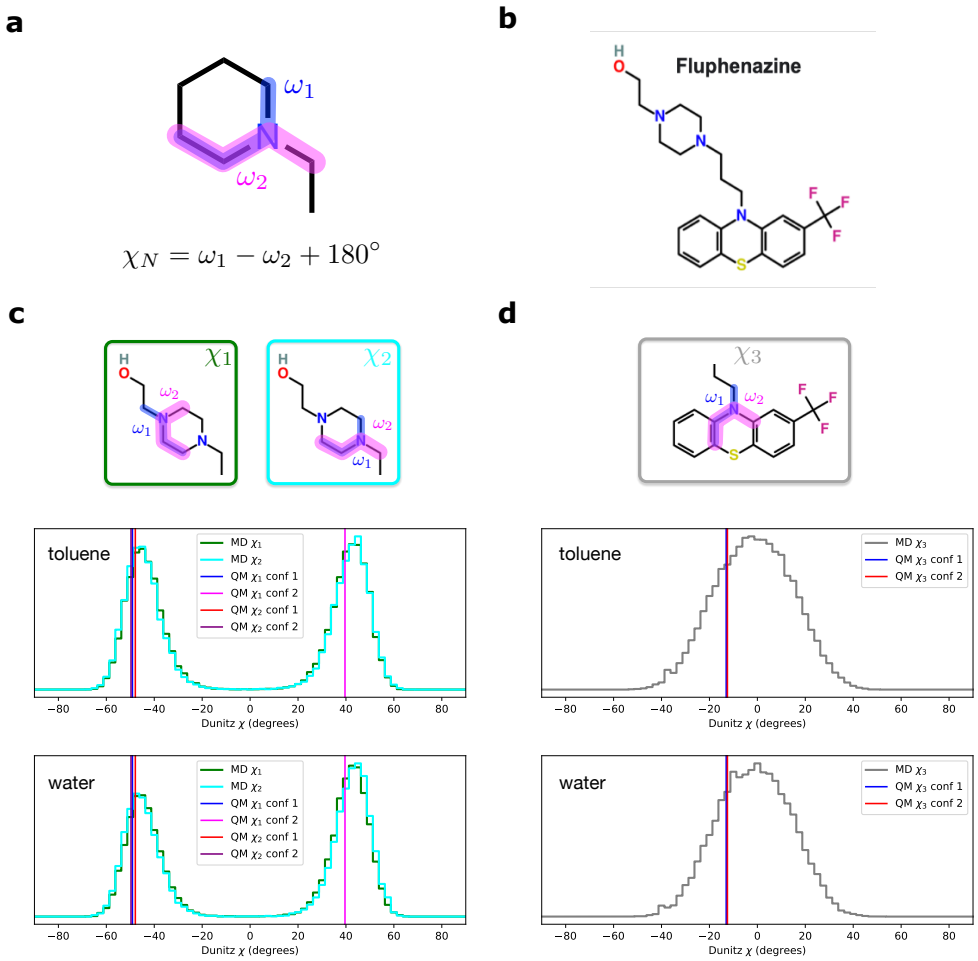

Fig. 5 (a) The Dunitz χ_N quantifies the extent of nitrogen pyramidalization using the difference of dihedral angle ω_1 (blue) from dihedral angle ω_2 (magenta). (b) The molecular structure of trazodone. (c) Distributions of Dunitz χ_N parameters χ_1 and χ_2, for the nitrogens in the piperazine group, sampled in expanded ensemble molecular dynamics (MD) simulations of fluphenazine in toluene and water. Vertical lines denote the χ_N values of two QM geometry-optimized conformations taken from the simulation. (d) Simulated Dunitz χ_N distributions for the phenothiazine nitrogen, for simulations in toluene and water. .

4 Conclusion

In this work, we have presented the results of an expanded ensemble (EE) free energy method to predict toluene–water partition coefficients in the SAMPL9 logP blind challenge. Our EE method achieved predictions within an RMSD of 2.26 kcal/mol, ranking third out of the nine submissions using physics-based molecular mechanics (MM) methods. Although the most accurate methods for logP prediction continue to be empirical or quantum-mechanical, our inspection of simulated MM versus QM geometries for nitrogen pyramidalization suggests force field improvements may continue to increase the accuracy of physics-based MM methods. The EE method is particularly well-suited for distributed computing platforms, and in the future we expect it to be used more widely for large-scale simulation-based virtual screening.

Author Contributions

Steven R. Goold: conceptualization, investigation, methodology, software, visualization, writing – original draft. Robert M. Raddi: conceptualization, data curation, formal analysis, investigation. Vincent A. Voelz: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, software, visualization, supervision, validation, writing – original draft preparation, writing – review editing.

Conflicts of interest

The authors declare no conflicts of interest.

Data Availability

SAMPL9 logP challenge instructions, experimental data, submissions and analysis are available at `https://github.com/samplchallenges/SAMPL9`. The expanded ensemble (EE) algorithm is implement and freely available in the open-source software package GROMACS (`https://gromacs.org`) Scripts for preparation of EE simulations and data analysis are available at `https://vvoelz.github.io/sampl9-voelzlab`.

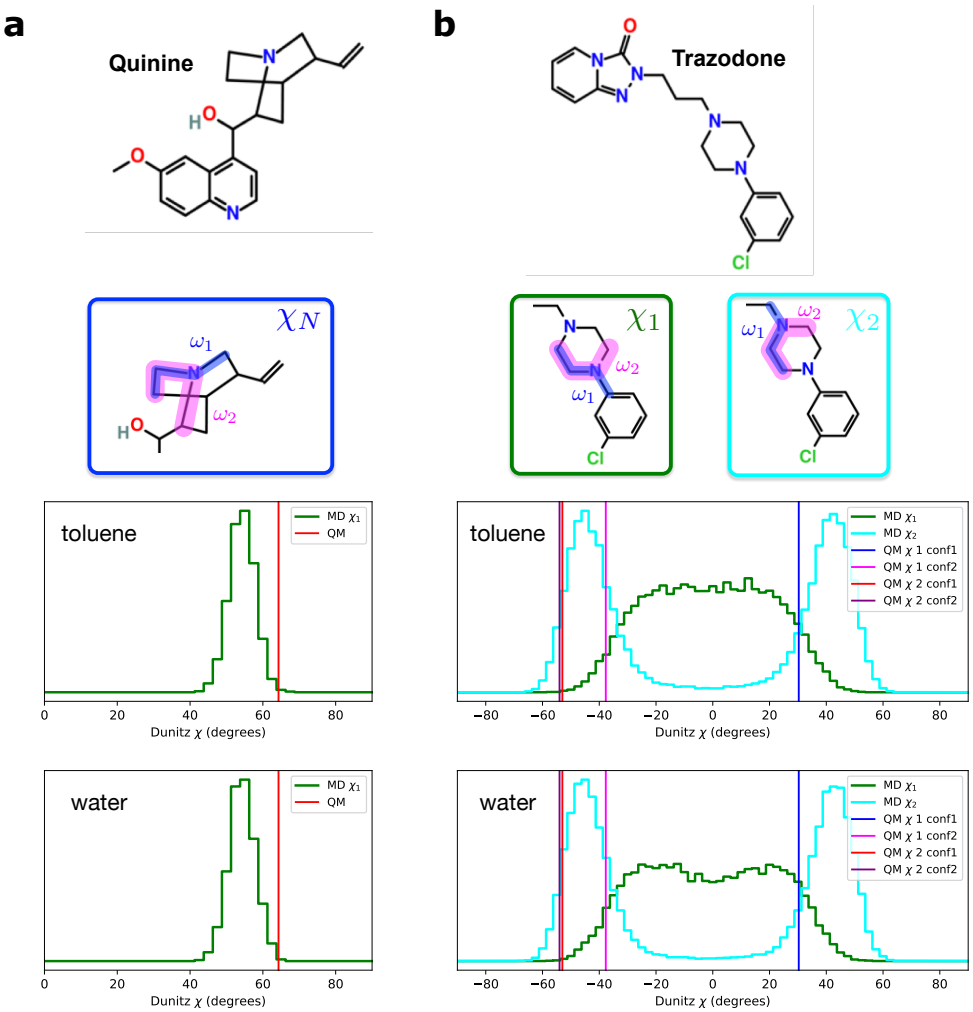

Fig. 6 (a) Distributions of Dunitz χ_N parameters χ_1 and χ_2, for the nitrogen in the quinuclidine group of quinine, sampled in expanded ensemble molecular dynamics (MD) simulations of quinine in toluene and water. Vertical line denotes the χ_N value of a QM geometry-optimized conformation taken from the simulations. (b) Distributions of Dunitz χ_N parameters χ_1 and χ_2, for piperazine nitrogens of trazodone, sampled in expanded ensemble molecular dynamics (MD) simulations in toluene and water. Vertical lines denote values of two QM geometry-optimized conformations taken from the simulations.

Acknowledgements

This work is supported by NIH R01GM123296. This research includes calculations carried out on HPC resources supported in part by the National Science Foundation through major research instrumentation grant number 1625061 and by the US Army Research Laboratory under contract number W911NF-16-2-0189. We thank the participants of Folding@home, who made this work possible. We appreciate the National Institutes of Health for its support of the SAMPL project via R01GM124270 to David L. Mobley (UC Irvine).

Notes and references

1 T. D. Bergazin, N. Tielker, Y. Zhang, J. Mao, M. R. Gunner, K. Francisco, C. Ballatore, S. M. Kast and D. L. Mobley, *Journal of computer-aided molecular design*, 2021, **35**, 771–802.

2 C. C. Bannan, K. H. Burley, M. Chiu, M. R. Shirts, M. K. Gilson and D. L. Mobley, *Journal of computer-aided molecular design*, 2016, **30**, 927–944.

3 M. Işık, T. D. Bergazin, T. Fox, A. Rizzi, J. D. Chodera and D. L. Mobley, *Journal of computer-aided molecular design*, 2020, **34**, 335–370.

4 H. S. Muddana, C. Daniel Varnado, C. W. Bielawski, A. R. Urbach, L. Isaacs, M. T. Geballe and M. K. Gilson, **26**, 475–487.

5 H. S. Muddana, A. T. Fenley, D. L. Mobley and M. K. Gilson, **28**, 305–317.

6 J. Yin, N. M. Henriksen, D. R. Slochower, M. R. Shirts, M. W. Chiu, D. L. Mobley and M. K. Gilson, **31**, 1–19.

7 A. Rizzi, S. Murkli, J. N. McNeill, W. Yao, M. Sullivan, M. K. Gilson, M. W. Chiu, L. Isaacs, B. C. Gibb, D. L. Mobley and J. D. Chodera, **32**, 937–963.

8 M. Amezcua, J. Setiadi, Y. Ge and D. L. Mobley, **36**, 707–734.

9 M. Işık, A. S. Rustenburg, A. Rizzi, M. R. Gunner, D. L. Mobley and J. D. Chodera, **35**, 131–166.

10 R. M. Raddi and V. A. Voelz, *Journal of computer-aided molec-*

ular design, 2021, **35**, 953–961.

11 M. N. Bahr, A. Nandkeolyar, J. K. Kenna, N. Nevins, L. Da Vià, M. Işık, J. D. Chodera and D. L. Mobley, **35**, 1141–1155.

12 K. Lopez, S. Pinheiro and W. J. Zamora, *Journal of computer-aided molecular design*, 2021, **35**, 923–931.

13 E. B. Lenselink and P. F. Stouten, *Journal of Computer-Aided Molecular Design*, 2021, **35**, 901–909.

14 N. Donyapour and A. Dickson, *Journal of computer-aided molecular design*, 2021, **35**, 819–830.

15 J. Warnau, K. Wichmann and J. Reinisch, *Journal of Computer-Aided Molecular Design*, 2021, **35**, 813–818.

16 W. J. Zamora, A. Viayna, S. Pinheiro, C. Curutchet, L. Bisbal, R. Ruiz, C. Ràfols and F. J. Luque, *Physical Chemistry Chemical Physics*, 2023, **25**, 17952–17965.

17 T. Nevolianis, R. A. Ahmed, A. Hellweg, M. Diedenhofen and K. Leonhard, *Physical Chemistry Chemical Physics*, 2023, **25**, 31683–31691.

18 T. Niu, X. He, F. Han, L. Wang and J. Wang, *Physical Chemistry Chemical Physics*, 2024, **26**, 85–94.

19 S. Zhang, D. F. Hahn, M. R. Shirts and V. A. Voelz, *Journal of Chemical Theory and Computation*, 2021, **17**, 6536–6547.

20 M. F. Hurley, R. M. Raddi, J. G. Pattis and V. A. Voelz, *Physical Chemistry Chemical Physics*, 2023, **25**, 32393–32406.

21 D. Novack, S. Zhang and V. A. Voelz, *bioRxiv*, 2024.

22 F. Wang and D. P. Landau, *Physical review letters*, 2001, **86**, 2050.

23 R. E. Belardinelli and V. D. Pereyra, **127**, 184105.

24 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19–25.

25 W.-T. Hsu and M. R. Shirts, *Journal of Chemical Theory and Computation*, 2024, **20**, 6062–6081.

26 A. J. Friedman, W.-T. Hsu and M. R. Shirts, *Multiple Topology Replica Exchange of Expanded Ensembles (MT-REXEE) for Multidimensional Alchemical Calculations*, 2024, `https://arxiv.org/abs/2408.11038`.

27 V. A. Voelz, V. S. Pande and G. R. Bowman, *Biophysical journal*, 2023, **122**, 2852–2863.

28 *The SAMPL9 Blind Prediction Challenges for Computational Chemistry*, `https://github.com/samplchallenges/SAMPL9`, original-date: 2021-08-17T21:01:38Z.

29 *OpenEye Toolkits 2021.1*, `http://www.eyesopen.com`.

30 S. Boothroyd, P. K. Behara, O. C. Madin, D. F. Hahn, H. Jang, V. Gapsys, J. R. Wagner, J. T. Horton, D. L. Dotson, M. W. Thompson, J. Maat, T. Gokey, L.-P. Wang, D. J. Cole, M. K. Gilson, J. D. Chodera, C. I. Bayly, M. R. Shirts and D. L. Mobley, **19**, 3251–3275.

31 J. Wagner, M. Thompson, D. L. Mobley, J. Chodera, C. Bannan, A. Rizzi, trevorgokey, D. L. Dotson, J. A. Mitchell, jaimergp, Camila, P. Behara, C. Bayly, JoshHorton, L. Wang, I. Pulido, V. Lim, S. Sasmal, SimonBoothroyd, A. Dalke, D. Smith, J. Horton, L.-P. Wang, R. Gowers, Z. Zhao, C. Davel and Y. Zhao, *openforcefield/openff-toolkit: 0.14.5 Minor feature release*, 2023, `https://doi.org/10.5281/zenodo.10103216`.

32 A. Jakalian, D. B. Jack and C. I. Bayly, *Journal of computational chemistry*, 2002, **23**, 1623–1641.

33 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of chemical physics*, 1983, **79**, 926–935.

34 O. Beckstein, *Becksteinlab/GromacsWrapper: release 0.9.0*, 2024, `https://zenodo.org/records/11682672`.

35 D. Novack, R. M. Raddi, S. Zhang, M. F. Hurley and V. A. Voelz, *ChemRxiv*, 2025.

36 T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber and W. F. Van Gunsteren, *Chemical physics letters*, 1994, **222**, 529–539.

37 C. Lee, W. Yang and R. G. Parr, *Physical review B*, 1988, **37**, 785.

38 T. H. Dunning Jr, *The Journal of chemical physics*, 1989, **90**, 1007–1023.

39 W. F. Polik and J. Schmidt, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2022, **12**, e1554.

40 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian~16 Revision A.03*, 2016, Gaussian Inc. Wallingford CT.

41 Y. Sun, T. Hou, X. He, V. H. Man and J. Wang, *Journal of computational chemistry*, 2023, **44**, 1300–1311.

42 T. Nevolianis, J. G. Rittig, A. Mitsos and K. Leonhard, *ChemRxiv*, 2024.

43 J. Wang, W. Wang, P. A. Kollman and D. A. Case, *Journal of molecular graphics and modelling*, 2006, **25**, 247–260.

44 J. T. Horton, S. Boothroyd, J. Wagner, J. A. Mitchell, T. Gokey, D. L. Dotson, P. K. Behara, V. K. Ramaswamy, M. Mackey, J. D. Chodera *et al.*, *Journal of chemical information and modeling*, 2022, **62**, 5622–5633.

45 A. Blanco-Gonzalez, W. Betancourt, R. Snyder, S. Zhang, T. J. Giese, A. W. Goetz, K. M. Merz, Jr., D. M. York, H. M. Aktulga and M. Manathunga, *ChemRxiv*, 2024.

SAMPL9 logP challenge instructions, experimental data, submissions and analysis are available at https://github.com/samplchallenges/SAMPL9. The expanded ensemble (EE) algorithm is implement and freely available in the open-source software package GROMACS (https://gromacs.org) Scripts for preparation of EE simulations and data analysis are available at https://vvoelz.github.io/sampl9-voelzlab .