

RESEARCH ARTICLE | MARCH 28 2023

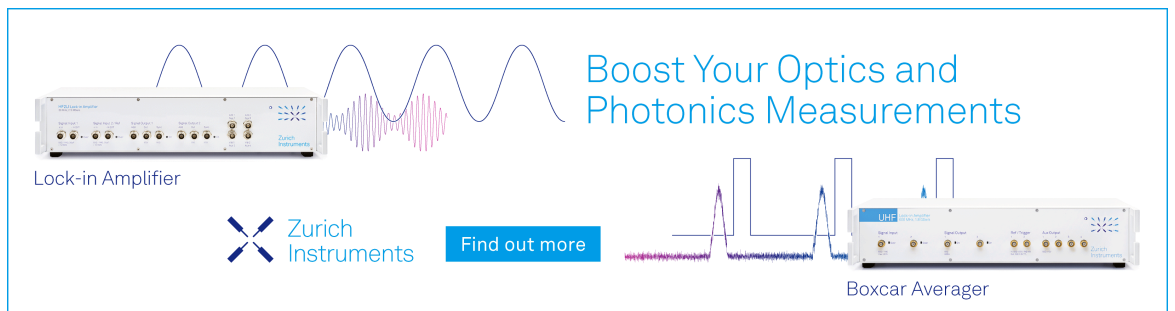
## SAMPL9 blind predictions for toluene/water partition coefficients using nonequilibrium alchemical approaches

Piero Procacci   ; Guido Guarnieri 



*J. Chem. Phys.* 158, 124117 (2023)

<https://doi.org/10.1063/5.0143824>



Boost Your Optics and Photonics Measurements

Lock-in Amplifier

Zurich Instruments

Find out more

Boxcar Averager

# SAMPL9 blind predictions for toluene/water partition coefficients using nonequilibrium alchemical approaches

Cite as: J. Chem. Phys. 158, 124117 (2023); doi: 10.1063/5.0143824

Submitted: 26 January 2023 • Accepted: 9 March 2023 •

Published Online: 28 March 2023



View Online



Export Citation



CrossMark

Piero Procacci<sup>1,a)</sup>  and Guido Guarneri<sup>2,b)</sup> 

## AFFILIATIONS

<sup>1</sup>Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, I-50019 Sesto Fiorentino, Italy

<sup>2</sup>ENEA, Portici Research Centre, DTE-ICT-HPC, P.le E. Fermi, 1, I-80055 Portici (NA), Italy

<sup>a)</sup> Author to whom correspondence should be addressed: [piero.procacci@unifi.it](mailto:piero.procacci@unifi.it)

<sup>b)</sup> Tel.: +39 081 7723576.

## ABSTRACT

We present our blind prediction of the toluene–water partition coefficients in the context of the SAMPL9 challenge. For the calculation of the solvation free energies in water, toluene, and 1-octanol, we used an efficient MD-based nonequilibrium alchemical technique relying on the GAFF2 non-polarizable force field. The method is based on the fast-growth of an initially decoupled solute. Canonical sampling of the associated end-state is efficiently obtained by performing a Hamiltonian replica exchange simulation of the gas-phase solute molecule alone, combined with equilibrium configurations of the solvent. Before submitting the prediction, a pre-assessment of the method and of the force field was made by comparing with the known experimental counterpart the calculated octanol–water partition coefficients using different set of atomic charges. The analysis allowed to optimize our blind prediction for the toluene–water partition coefficients, providing at the same time valid clues for improving the performance and reliability of the non-polarizable force field in free energy calculations of drug-receptor systems.

© 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0143824>

## I. INTRODUCTION

The Statistical Assessment of the Modeling of Proteins and Ligands (SAMPL) is a well-established international initiative for advancing the computational techniques in drug design.<sup>1–5</sup> New experimental data, such as host–guest dissociation free energies, hydration free energies, acid–base dissociation constants, or partition coefficients, are undisclosed to participants until the prediction submission deadline so that the true predictive power of methods can be assessed.

In modern methodologies based on atomistic molecular dynamics (MD) simulations,<sup>6–11</sup> SAMPL challenges are aimed to a two-fold scope, i.e., that of testing the accuracy of the physical model or “force field,”<sup>12–20</sup> often leading to systematic errors, and that of assessing the reliability of a computational procedure or protocol, likely producing random errors. In the latest SAMPL9 challenge,<sup>21</sup> participants were required to predict the toluene/water partition coefficients,  $\text{Log}P_{\text{tw}}$ , for a series of compounds with disparate flexibility or molecular weight and coarsely spanning a significant portion

of the “drug-like” chemical space, including moieties such as carboxyl, carbonyl, sulfonic, oxydryl, amino, amide, halogen, phenyl, hetero-cyclic, and alkyl (see Fig. 1). Given that the drug–receptor binding affinity is implicitly driven by the difference of two solvation free energies, namely, that of the drug when bound to the receptor and that of the drug in bulk solvent,<sup>22–24</sup> the capability of accurately predicting partition coefficients, strictly related to solute transfer free energy between the two solvents, is an essential requirement for applying any methodology to the challenging case of drug design.

Partition coefficients are generally computed using very efficient knowledge-based (KB) empirical techniques, such as XlogP3,<sup>25–27</sup> or using *ab initio* (QM) approaches mixed with trained parametric models for the polarization continuum, as in the Consortium for Small Scale Modelling (COSMO)<sup>28</sup> or the solvation model based on density (SMD)<sup>29</sup> implicit solvation models. In SAMPL challenges, these approaches generally outperform<sup>30,31</sup> much more costly MD methodologies, based on the *explicit* representation of the solvent to account for micro-solvation phenomena and largely

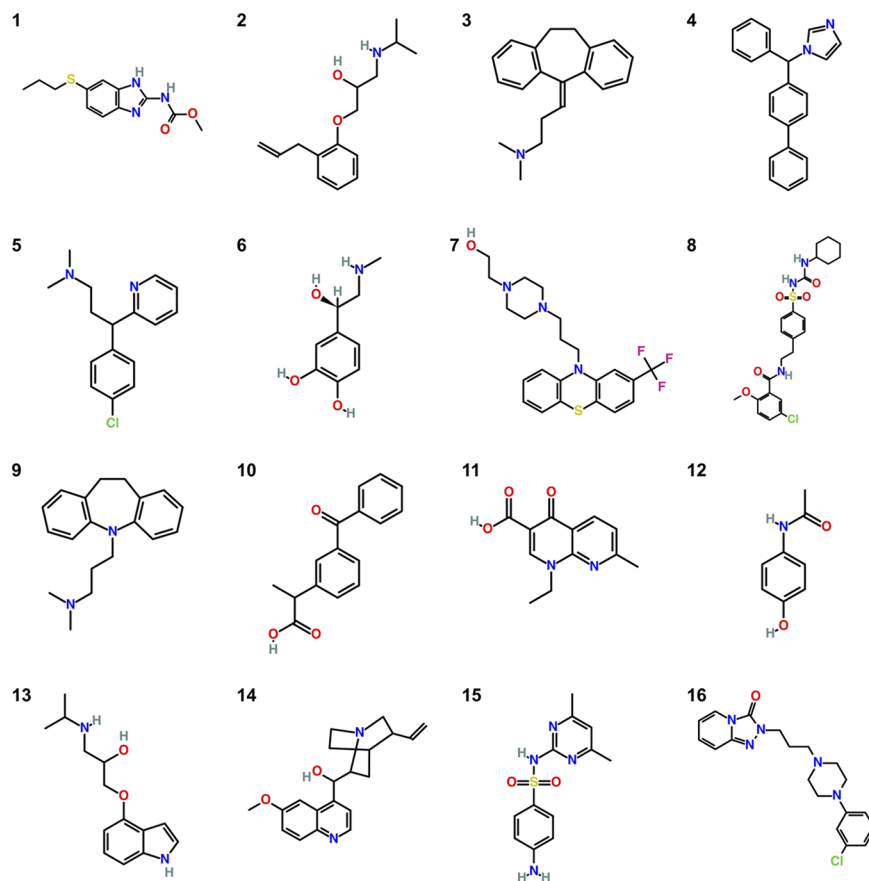


FIG. 1. Compounds for the SAMPL9 blind challenge for toluene/water partition coefficient.

relying on the so-called alchemical protocol.<sup>6,7,32</sup> However, when the KB methodology or QM-based techniques are applied to the calculation of host-guest or drug-receptor binding affinities, their performances are significantly degraded.<sup>3,5</sup> On the contrary, the reliability of the MD-based approaches is much less affected when passing from LogP predictions to drug-receptor or host-guest binding affinities calculations,<sup>33–36</sup> systematically resulting, in this last case, as the best performing methods.<sup>3,5</sup> Apparently, as demonstrated in SAMPL challenges, QM or KB methods are presently unable to effectively cope with the heterogeneity of the solute environment in the calculation of the solvation free energy of a host-bound solute. Hence, testing an MD-based approach on a LogP blind prediction involving bulky and flexible compounds with diverse chemical groups is important since a successful protocol in this context has a good chance to produce reliable predictions when transferred to the more challenging case of drug-receptor binding affinities.

Most of the MD submissions in SAMPL challenges are based on the well-established alchemical approach, whereby the free energy between the two end-states (solvated solute and gas-phase solute) is recovered by setting up a so-called  $\lambda$  stratification of  $n$  intermediate non-physical states where the interaction between the solute molecule and the explicit solvent molecules is progressively decoupled via the alchemical parameter  $\lambda$ .<sup>6,37</sup> The  $n$  intermediate states

at the fixed coupling parameter  $\lambda$  are chosen so that the potential energy distributions of contiguous  $\lambda$  states have a significant overlap,<sup>6</sup> allowing the calculation of the solvation free energy as a sum of  $n - 1$  individual free energy contributions along the  $\lambda$ -stratification by way of free energy perturbation<sup>38</sup> (FEP). FEP is an *equilibrium* technique requiring canonical sampling on the end-states and on each of the intermediate  $\lambda$  states. For flexible compounds with many rotatable bonds and high torsional barriers, canonical sampling can be problematic especially when dealing with solute molecules that are characterized by a complex conformational landscape, requiring long simulations whose convergence depends in an unknown way on the level of the  $\lambda$  coupling parameter.<sup>39–41</sup> In such cases, it may be necessary to combine the  $\lambda$ -stratification approach with costly enhanced sampling techniques, such as Hamiltonian Replica Exchange<sup>41,42</sup> (HREM).

In Refs. 33 and 34, we have recently analyzed and discussed the reliability of unidirectional or bidirectional nonequilibrium techniques (NE) in the calculation of solvation free energies. In the alchemical NE approach, the end-states are connected, rather by a stratification of *equilibrium* states, by a swarm of fast NE independent trajectories (initiated from canonically sampled end states) where the solute-solvent coupling parameter is continuously varied. The solvation free energy is recovered from the NE work data, exploiting the Crooks theorem<sup>43</sup> via the Bennett acceptance ratio<sup>44,45</sup>

(BAR) if the NE transitions are performed in both senses or using the Jarzynski theorem<sup>46</sup> or the Gaussian assumption<sup>47,48</sup> if the NE transformations are conducted in one sense. Unlike in FEP, in the NE variant of the alchemical approach, canonical sampling is required *only at the end states*. This is a huge advantage with respect to FEP, especially if the NE growth (NE-FG), i.e., starting from an end-state where the solute is completely decoupled from the solvent. In this case, canonical sampling can be reliably and cheaply obtained by performing an enhanced sampling of the gas-phase solute molecule alone, combined with equilibrium configurations of the solvent.

In this paper, we present our blind predictions on  $\text{LogP}_{tw}$  for the 16 compounds shown in Fig. 1 proposed in the latest SAMPL9 challenge. Given the expected confidence interval for the NE-FG method,<sup>33,34</sup> the availability of experimental  $\text{LogP}_{ow}$ <sup>49</sup> provided the opportunity to attempt to disentangle the systematic error due to the force field (mostly affected by that due to the fixed-charge representation<sup>50</sup>) from the methodological error, hence improving our blind prediction for the SAMPL9  $\text{LogP}_{tw}$ . Our submission had a good ranking among all submitted blind predictions (mostly based on empirical or QM methodologies), revealing as the best-performing and, at the same time, most efficient approach among the MD-based submissions.

## II. THEORETICAL ASPECTS

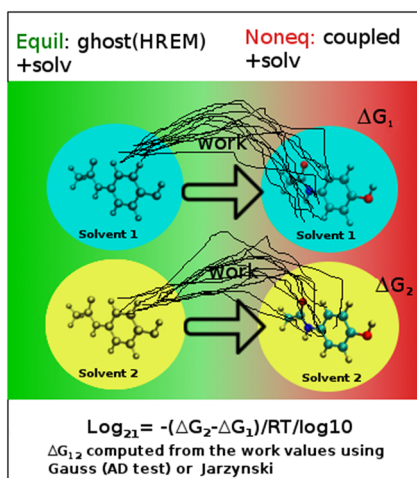
The theory of NE alchemical variant for the calculation of solvation free energy in the context of LogP determination is described in full detail in Ref. 33 for the bidirectional, BAR-based approach and in Ref. 34 for unidirectional methodologies. Here, we sketch out the crucial methodological aspects of the latter with emphasis on the expected accuracy and precision and on efficiency. Unidirectional methodologies imply that the NE alchemical transitions are performed in one sense, i.e., recoupling an initially decoupled

solute molecule in the solvent or decoupling an initially fully interacting solute in a swarm of independent fast-growth trajectories with a common time protocol. Both these NE processes must be started from a *canonical* sampling of the initial end-state, ending in a *nonequilibrium* sampling final end-state. In the annihilation direction, one needs to prepare a canonical sampling of the solvated compound, a task that for flexible and bulky solutes, such as those of Fig. 1, may require the usage of HREM on the full system. In the growth sense, the initial equilibrium states can be prepared by sampling the conformations of the isolated (gas-phase) molecule using HREM and then by combining such gas-phase configurations with equilibrium configurations of the pure solvent. This approach has the advantage to apply the HREM to a single molecule system with a very limited computational cost, while a *single* ordinary MD simulation can be used for the sampling of the solvent configurations to be combined with the HREM sampling of any given solute. The fast-growth procedure for the calculation of the LogP is illustrated in Fig. 2. For each NE-FG trajectory in a given solvent, the externally driven  $\lambda$  parameter does on the system the work  $W = \int_0^\tau \frac{\partial H(\lambda)}{\partial \lambda} \dot{\lambda} dt$ , where  $\tau$  is the duration of the NE trajectories. The corresponding solvation free energy  $\Delta G_s$  is then computed from the work histogram values using

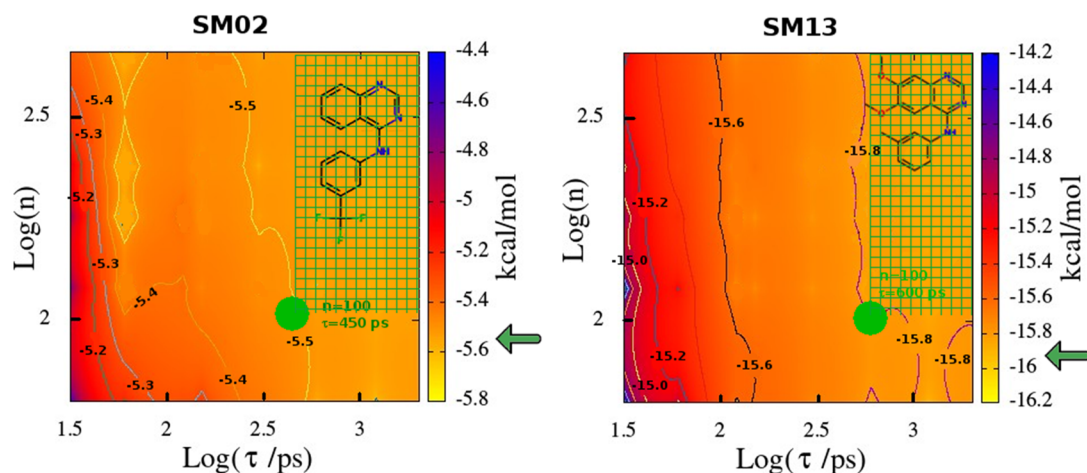
$$\Delta G_s = \langle W \rangle - \frac{\sigma^2}{2RT}, \quad (1)$$

$$\Delta G_s = -RT \ln \langle e^{-W/RT} \rangle + B_J(n, \sigma), \quad (2)$$

where  $T$  is the temperature and  $R$  is the gas constant. Equation (1) is valid and *exact* if and only if the work distribution is normal and can be straightforwardly proved using the Crooks theorem.<sup>47,51</sup> If the distribution is non-Gaussian, Eq. (2) must be used. The latter is based on the Jarzynski exponential average plus the bias correction  $B_J(n, \sigma)$  due the finite size of the work sample.<sup>52</sup> Whenever normality of the work distribution can be reliably assumed, Eq. (1) provides an *unbiased* estimates of the solvation energy and should hence be preferred to the Jarzynski-based estimate [Eq. (2)]. For work samples with  $n \geq 100$ , normality can be ruled with 95% confidence if the quantity  $A^2$  of the Anderson Darling (AD) test is larger than 0.752.<sup>53,54</sup> A probability exceeding 50% from the  $p$ -value<sup>55</sup> associated with AD that the distribution is normal is obtained when  $A^2 \leq 0.34$ . Given a  $n$  value work histogram with variance  $\sigma^2$ , the bias  $B_J$  in Eq. (2) can be estimated by sampling  $n$  values from a normal distribution  $\mathcal{N}$  with the same variance and zero mean and by comparing the exact value for the associated free energy,  $\Delta G_{\mathcal{N}} = -\beta\sigma^2/2$ , with that obtained from the exponential average on the normally distributed  $n$  points with  $\sigma^2$  variance,  $\Delta G_J(n, \sigma) = -RT \ln(\frac{1}{n} \sum_{i \in \mathcal{N}} e^{-\beta W_i})$ , yielding the bias as  $B_J(n, \sigma) = \Delta G_{\mathcal{N}} - \Delta G_J(n, \sigma)$ . In the assumption that the initial sampling of the fast-growth process is canonical (see Fig. 2), the parameters affecting accuracy and precision of the method are the number of NE-FG trajectories and the time duration  $\tau$  of the NE-FG trajectories, related to the resolution of the work distribution entering the solvation free energy. In the contour plots of Fig. 3, we show the convergence of the solvation free energy for the SM02 molecule in water (left) and the SM13 molecule in octanol [computed with Eq. (1) or Eq. (2) depending on the character of  $P(W)$ ] as a function of  $\tau$  and of the number of sampled work values.



**FIG. 2.** Scheme for NE-FG calculation of the LogP: NE-FG transitions are done from left (equilibrium state of the ghost solute in the solvent to NE state of the fully coupled solute in the solvent).



**FIG. 3.** Convergence of NE-FG solvation free energy of in SM02 in water (left) and of SM13 in octanol as a function of  $\tau$  [time duration of NE-FG trajectories and  $n$  (number of NE trajectories or work values)].

The work data were taken from Ref. 33. The green arrow on the z-scale indicate the reference value computed with BAR using 480 trajectories lasting 2.6 ns. The rectangle marked with a green grid corresponds to a plateau where the solvation free energy estimate is basically constant and coincident with the BAR reference value. We can see that NE-FG estimates of then solvation free energies for these two molecules are already close to the reference value for  $n = 100$  and  $\tau = 450$  ps for SM02 in water and for  $n = 100$  and  $\tau = 600$  ps SM13 in octanol.

Finally,  $\text{LogP}_{ab}$  is computed from the difference of the solvation free energies of the solute in the two solvents  $a, b$  as

$$\text{LogP}_{ab} = \log \left[ \frac{a}{b} \right] = - \frac{(\Delta G_a - \Delta G_b)}{RT \ln(10)}. \quad (3)$$

### III. METHODS

The starting structures for the solute was prepared with OpenBabel, by converting the SMILES code to a PDB file, using the `-gen3D` option.<sup>56</sup> OpenBabel creates first a non-optimal structure using fragment templates followed by a steepest geometry optimization with the MMFF96 force field.<sup>57</sup> Fixed atomic charges in our blind submission are readily computed on the OpenBabel-generated structure at the popular AM1-BCC level with the prospect of an easy and automatable implementation of our NE-FG protocol for LogP and host-guest predictions. We used the GAFF2<sup>17</sup> force field for the compounds of Fig. 1 and for the 1-octanol and toluene solvent molecules. The force field assignment and AM1-BCC charges were obtained from the web tool “PrimaDORAC”<sup>19</sup> ([www1.chim.unifi.it/orac/primadorac](http://www1.chim.unifi.it/orac/primadorac)). Given the sensitivity<sup>50</sup> to atomic charges of the solvation free energies, we tested two sets of atomic charges for the solutes, namely, the AM1-BCC charges<sup>58</sup> from PrimaDORAC and the Electrostatic Potential charges (ESP) computed at the HF/6-31Gd level according to the Merz-Kollman scheme with dipole restraint.<sup>59</sup> For water, we used the OPC3 model.<sup>60</sup>

We must warn that the choice of the initial 3D structure for *fixed* atomic charge determination may have a non-negligible impact on the finally accuracy. In this respect, iterative refitting or adaptive parameterization of the atomic charges based on the gas-phase and bulk conformational sampling may be of help in the selection of the best conformation-dependent fixed atomic charges, as shown in a recent SAMPL challenge study.<sup>61</sup> However, part of the emphasis on the present sample LogP challenge was on the benefit/cost ratio. Participants were, in fact, required to provide details on the computational resources used for their predictions. Smart adaptive methodologies for optimization of 3D-dependent atomic charges, such as those described in Ref. 61, are expensive and system dependent and, hence, are not suited for low-cost automatable screening protocol in drug design.

HREM of the gas-phase of the solute in the NVT ensemble<sup>62</sup> at 300 K was done by scaling only the potential energy of the isolated molecule with a minimum scaling factor of 0.1, corresponding to a “temperature” of 3000 K using eight replicas with scale factor progression<sup>63</sup> given by  $c_m = c^{(m-1)/8}$ ,  $m = 1 \dots 8$ . HREM simulations lasted 8 ns, collecting 96 solute conformations on the target state ( $T = 300$  K).

Pure solvents were equilibrated in the constant temperature constant pressure ensemble (NPT) ensemble for 1 ns with  $P = 1$  atm and  $T = 300$  K, imposed via a Parrinello–Rahman Langrangian with isotropic stress tensor<sup>64</sup> and Nosé thermostats coupled to the translational and internal degrees of freedom of the system. 96 configurations (to be combined with the decoupled 96 configuration of the solute) were collected in the following 1 ns of production run. Solvents were simulated in cubic boxes under periodic boundary conditions (PBCs) with 1728 for water, 343 for toluene, and 216 for 1-octanol, yielding a mean side-length of 37.43, 39.38, and 38.62 Å for water, toluene, and 1-octanol, respectively. The calculated densities for 1-octanol and toluene were 0.81 g/cm<sup>3</sup> (experimental value 0.82 g/cm<sup>3</sup>) and 0.86 g/cm<sup>3</sup> (experimental value 0.87 g/cm<sup>3</sup>), respectively. The density of OPC3 water was 0.992 compared to the experimental value of 0.997.

The ( $n = 96$ ) NE-FG simulations, starting from the canonical sampling obtained by combining the HREM gas-phase configurations of the solute molecules with that of the pure solvent, were launched in the NPT ensemble. The  $\lambda$  alchemical coupling parameter was switched on in  $\tau = 450$  ps for toluene and water and in 600 ps for 1-octanol, according to a common protocol whereby the Lennard–Jones interactions were first turned on using a soft-core regularization<sup>65</sup> in 300 ps for water and toluene and in 390 ps for 1-octanol, followed by the electrostatic interactions in the last 150 ps for water and toluene, and in the last 210 ps for 1-octanol. The choice of  $n$  and  $\tau$  for the parameters in the NE-FG runs was dictated by the results on past SAMPL challenges on drug-like molecule, as shown in Fig. 3.

In all simulations, the equations of motion were integrated using a multiple time-step r-RESPA scheme<sup>64,66</sup> imposing constraints only on the X–H bonds with X being any heavy atom. The long range cut-off for Lennard–Jones interactions was set to 13 Å in all cases. Long range electrostatic were treated under PBC using the smooth particle mesh Ewald method,<sup>67</sup> with an  $\alpha$  parameter of 0.38 Å<sup>-1</sup>, a grid spacing in the direct lattice of about 1 Å, and a fourth order B-spline interpolation for the gridded charge array.

The NE-FG calculations were run on the CRESCO6 cluster<sup>68</sup> using the ORAC program.<sup>69</sup> For the calculation of a solvation energy, the associated parallel job engaged 576 cores (96 MPI process each using six OpenMP threads) lasting about 1.5 and 2 wall-time clock hours for water/toluene and 1-octanol, respectively.

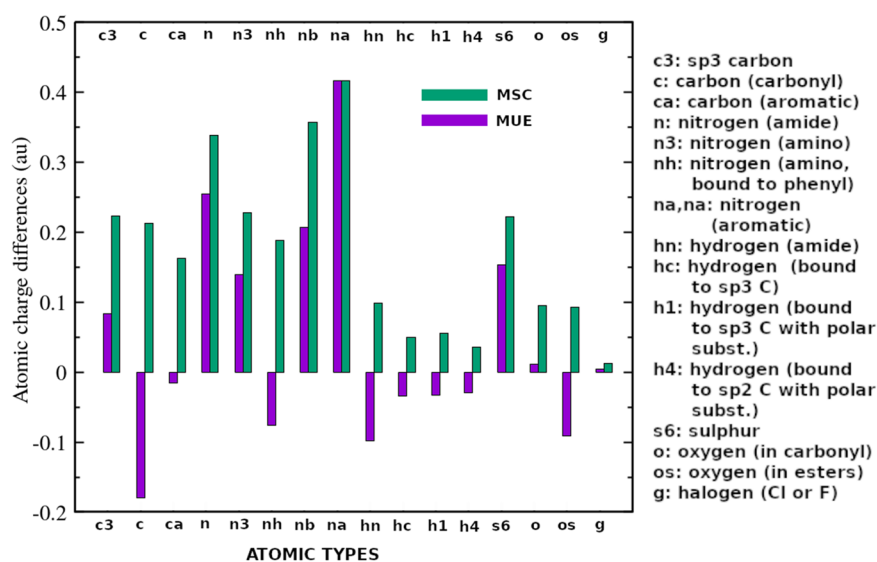
#### IV. RESULTS AND DISCUSSION

The SAMPL9 challenge calls for the prediction of the toluene/water partition coefficients of the compounds of Fig. 1 for which no experimental results were available. Given the sensitivity of solvation free energies to the selected atomic charges in non-polarizable force fields,<sup>50</sup> before submitting our prediction, we made a pre-assessment of the NE-FG performance with the GAFF2 atomic types force field on the LogP<sub>ow</sub> coefficients (for which experimental

data are available from the PUBCHEM database<sup>49</sup>) using two sets of atomic charges, namely, (i) the AM1-BCC charges, consisting in the Mulliken charges computed using the semiempirical AM1 method<sup>70</sup> plus a bond-charge correction tabulated in Ref. 58, and (ii) the ESP charges computed at the HF/6-31Gd level. Both these approaches are widely adopted in free energy calculations based on fixed-charge force fields

In Fig. 4, we report the mean unsigned (MUC) and signed charge (MSC) difference between the AM1-BCC and HF/6-31Gd charges averaged on all the atoms in the SAMPL9 compounds belonging to a given atomic GAFF2 type. Atoms belonging to the same atomic type are characterized by a common chemical environment and are hence expected to bear similar charges. Hence, when MSC is positive, it implies that the HF/6-31Gd average charge for that atomic type is systematically lower than the corresponding AM1-BCC charge. On the other hand, where the difference was random within a given atomic type, we should see MSC close to zero and MUC of the order of the mean unsigned deviation between the two sets of charges. In some cases, MUC and MSC are significant ( $>0.2 e$ ) and almost coincident, indicating that the HF/6-31Gd charges are systematically more *negative* with respect to the AM1-BCC counterpart. This occurs, in particular, for the nitrogen atoms in aromatic systems, in amino and amide groups, bearing, in general, a much more pronounced negative charge when ESP charges are evaluated at the HF/6-31Gd level. For the carbonyl carbon, MUC and MSC are similar in magnitude but with opposite sign MSC, implying the HF/6-31Gd charge is systematically more positive with respect to the AM1-BCC charge. In general, we may say that when using the HF/6-31Gd, we observe a higher charge separation within the molecule. We may therefore expect that the solvation free energy is affected by these differences in the two sets of charges in manner that depends on the polarity of the solvent.

In Tables S1 and S2 of the supplementary material, we report the detailed results for the NE-FG-computed solvation free energy estimates in the three solvents for the compounds of Fig. 1 using the AM1-BCC atomic charges and the HF/6-31Gd ESP charges,

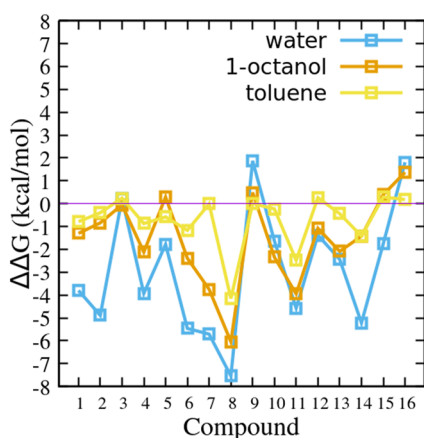


**FIG. 4.** Mean signed MSC =  $\langle q_{\text{am1-bcc}} - q_{\text{6-31Gd}} \rangle_{\text{type}}$  and mean unsigned charge difference  $\text{MUC} = \langle |q_{\text{am1-bcc}} - q_{\text{6-31Gd}}| \rangle_{\text{type}}$  in electron units for the GAFF2 atomic types of the compounds of Fig. 1. MUC and MUE are averaged over all atoms belonging to a given GAFF2 atomic type.<sup>17</sup>

respectively. The reported estimates in Tables S1 and S2 are based on Eq. (1) if  $A^2 < 0.34$ , implying a  $p$ -value  $> 50\%$  (i.e., the null hypothesis referring to the normal distribution is true with a probability exceeding 50%), and on Eq. (2) otherwise. The errors (95% confidence interval) have been computed by bootstrap with resampling from the 96 work data.

In Fig. 5, we show the differences in the solvation energy with the two sets of charges in the three solvents. We may say, in general, that the use of the HF/6-31Gd charges produces more negative solvation free energies in all solvents. Differences tend to be higher in water with respect to less polar solvents, such as 1-octanol or toluene. More in detail, the free energy gain when switching from AM1-BCC to HF/6-31Gd charges significantly increases in water (by up to 7 kcal/mol in compound 8) in all cases except for compound 9 and 16 where the solvation free energy computed with the HF/6-31Gd charges is similar to that evaluated using the AM1-BCC charges. This result is consistent with the fact that, in general, charge separation is higher in the HF/6-31Gd charges with respect to the AM1-BCC charges, and the free energy gains are expected to increase with increasing polarity of the solvent, hence preventing a compensating effect in the calculated LogP between a non-polar solvent and water when switching between the two sets of atomic charges.

The reliability of NE-FG calculated  $\text{LogP}_{ow}$  can be assessed in Fig. 6 where we report the correlation plots between experimental and NE-FG-calculated  $\text{LogP}_{ow}$  [computed from the solvation energies of Tables S1 and S2 by way of Eq. (3)] with the different charge sets. In one case (plot on the top left in Fig. 6), we used the AM1-BCC charges of the solute in both water and 1-octanol. In a second case (top central plot), we used the HF/6-31Gd ESP charge in both water and 1-octanol. Finally, in the third case (right top plot), we used the HF/6-31Gd ESP charges for water and the AM1-BCC charges for 1-octanol. The usage of two different set of charges in the two solvents is justified by the fact that charge separation on a molecule is a function of the polarity of the solvent so that the large HF/6-31Gd charges should be appropriate for water, while the smaller AM1-BCC charges should be used in the much less polar 1-octanol



**FIG. 5.** Solvation free energy differences computed using the HF/6-31Gd and AM1-BCC charges for the 16 compounds of Fig. 1, computed as  $\Delta\Delta G = \Delta G_{6-31Gd} - \Delta G_{am1-bcc}$ . Lines are drawn as a guide for the eye.

**TABLE I.** Agreement between experimental and calculated  $\text{LogP}_{ow}$  for the NE-FG approach with three different sets of solute atomic charges and for the xLogP3 empirical method as measured by  $R$  (Pearson correlation coefficient),  $a$  and  $b$  (in kcal/mol) (slope and intercept of the best fitting line), MAE and MSE [the mean absolute and signed errors (in kcal/mol)], and  $\tau$  (Kendall rank coefficient).

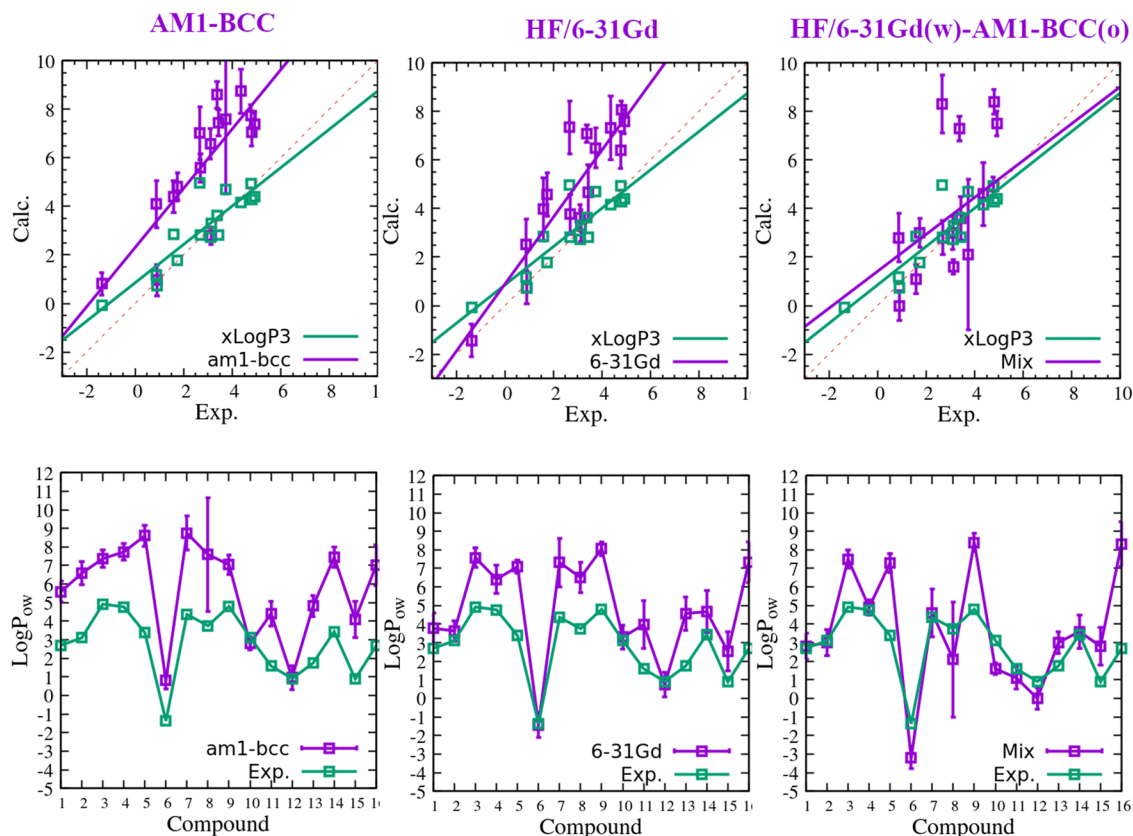
Charge set	$R$	$a$	$b$	MAE	MSE	$\tau$
AM1-BCC	0.83	1.22	2.30	2.97	-2.92	0.62
HF/6-31Gd	0.88	1.38	0.88	1.98	-1.95	0.60
Mix	0.76	1.41	-0.33	1.63	-0.81	0.55
xLogP3	0.89	0.79	0.85	0.57	-0.27	0.62

solvent. Correlation metrics for the plots shown in Fig. 6 is collected in Table I.

We can see that while correlation, as expressed by the Pearson and rank Kendall coefficients, is good and comparable to that of the reliable xLogP3 estimate in all cases, the mean absolute error (MAE) and MSE are found to critically depend on the selected charge set, highlighting the sensitivity of the solvation energies and LogP to electrostatic interactions. In this regard, we note that both MAE and MSE improve significantly when switching from AM1-BCC to HF/6-31Gd charges with correlation still being very strong. The slope and intercept of the best fitting line, on the other hand, are still far from their corresponding ideal value of 1 and 0, values that are more closely approached by the xLogP3 predictions. The best MAE and MSE were obtained when we used the mixed approach, i.e., HF/6-31Gd charges in water and the AM1-BCC charges in 1-octanol.

In the bottom plots of Fig. 6, we show the experimental and corresponding NE-FG-calculated  $\text{LogP}_{ow}$  for each of the 16 compounds of Fig. 1. For the AM1-BCC charge set in both solvents (bottom left plot in Fig. 6), all  $\text{LogP}_{ow}$  of the nitrogen containing compounds are overestimated except for compounds 10 and 12, which are in good agreement with experiment. Compound 10 does not contain nitrogen, while compound 12 has one amide nitrogen. Switching to the HF/6-31Gd charges in both solvents (bottom central plot in Fig. 6), overestimation is tamed and the calculated  $\text{LogP}_{ow}$  of compound 6 (besides 10 and 12) is now in excellent agreement with the experimental counterpart.

Finally, when using HF/6-31Gd charges for the solute in water and AM1-BCC charges in 1-octanol, the agreement with experiment becomes satisfactory for all compounds except for four clear outliers, namely, compounds 3, 5, 9, and 16, where still we have a significant overestimation of  $\text{LogP}_{ow}$ , comparable to that observed in the two previously discussed charge sets. On the other hand, the decisive improvement of  $\text{LogP}_{ow}$  of compounds containing one or more  $sp^2/sp^3$  nitrogen atoms in cycles (1, 4, 5, 7, 11, 13, 14, 15, and 16) with respect to  $\text{LogP}_{ow}$  computed with the AM1-BCC set leads us to believe that AM1-BCC charges on nitrogen containing cycles do not, in general, reflect the real electron density in hetero-cycles in water, possibly due to an effect of the nitrogen lone pair<sup>71</sup> and/or polarization of the solvent. Such systematic overestimation of  $\text{LogP}_{ow}$  when using AM1-BCC charges in both solvents for solutes with nitrogen containing cycles (SM41-SM46) was also detected in the SAMPL7 batch of compounds in our NES-1 (GAFF2/OPC3)G submission as well as in other FEP-based alchemical submissions.<sup>31</sup>



**FIG. 6.** Top: correlation diagrams between experimental<sup>49</sup> and calculated octanol–water partition coefficients using different set of atomic charges (see text). Bottom: details of the corresponding differences for the compounds of Fig. 1 (lines are drawn as a guide for the eye).

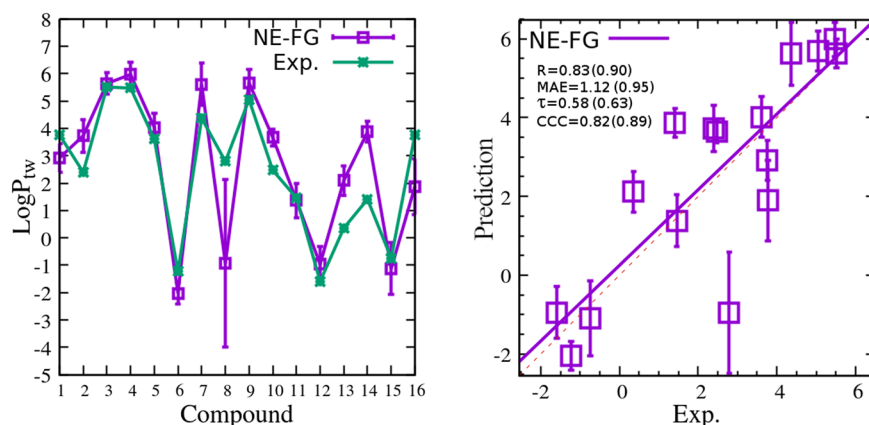
The outliers in the mixed charge set, 3, 5, 9, and 16 all share a tertiary amino group where the GAFF2 atomic type of the nitrogen is n3 (see Fig. 4). Compound 16 has five nitrogen atoms in cycles and a chlorine atom where the role of the positive charge on the extra center<sup>72</sup> due to the so-called “sigma-hole” (not included in our model for compound 16) may be, in part, responsible for the observed discrepancy. It is worth mentioning that substantial overestimation of MD-based LogP<sub>ow</sub> with the GAFF2 parameterization for tertiary amine solutes with n3 type was also detected in Ref. 50.

The mixed approach can be considered as a sort of zero-order polarizable model, where we use a stronger charge separation (see Fig. 4) and, hence, higher electric moments in the solvent with a high dielectric constant where the mean field induction effects are supposed to be more important. Such an approach has, however, drawbacks since we still have four clear outliers (see Fig. 6) and  $R$  and  $\tau$  correlation coefficients inferior to those observed using the same charge set in both solvents (see Table I). While most of the solvation free energy in non-polar solvents is recovered when switching on the Lennard-Jones potential<sup>36</sup> and the electrostatic contribution is minor, in water, the recharging stage produces by far the major contribution to the free energy.<sup>36</sup> If the solute atomic charges do not reflect with sufficient accuracy the electron density in liquid water,

large errors in LogP may arise. We can hence infer that most of the observed discrepancies in the mixed approach are probably due to a non-optimal charge parameterization for the solute *in water*, in particular, for the four outliers when using the mixed charge set. On the overall, the analysis of LogP<sub>ow</sub> data reported in Fig. 6 appears to confirm<sup>31,50</sup> that the parameterization of electrostatic interactions in the popular GAFF2 force field for tertiary amino compounds or for solute bearing sp<sub>2</sub> or sp<sub>3</sub> nitrogen containing cycles should be carefully revised. Based on these results, we decided to submit our prediction for the unknown LogP<sub>tw</sub>, exploiting the knowledge of experimental LogP<sub>ow</sub> and using the solvation energies computed in the less polar solvents toluene and 1-octanol with the AM1-BCC charges set as reported in Table S1 of the supplementary material. The LogP<sub>tw</sub> toluene–water partition coefficients in our submission have been computed as  $\text{LogP}_{tw} = \text{LogP}_{to} + \text{LogP}_{ow}(\text{Exp})$ .

Results for this prediction (named logP<sub>NE-FG</sub> on the SAMPL9 GitHub site<sup>21</sup>) are collected in Fig. 7, where, on the left, we report the experimental and calculated LogP<sub>tw</sub> with 95% confidence interval for the SAMPL9 compounds and, on the right, the corresponding correlation plot. Experimental and NE-FG-calculated values are well correlated, yielding a Pearson coefficient ( $R$ ) of 0.83, a Kendall rank coefficient ( $\tau$ ) of 0.58, a mean absolute





**FIG. 7.** Blind prediction for the toluene–water partition coefficients in SAMPL9. Left: experimental values vs NE-FG-calculated with 95% confidence interval for all compounds of Fig. 1 (lines are drawn as a guide for the eye). Right: corresponding correlation plot.

error (MAE) of 1.12 LogP units, and a Lin concordance coefficient (CCC) of 0.82.

The latter metrics combines<sup>73</sup> in a single coefficient a measure of the accuracy, as assessed by the MAE, and of precision in ranking as assessed by the correlation coefficient  $R$ . Our blind prediction ranks fifth for the MAE and sixth for the CCC among the 20 submitted predictions in the challenge. Compound 8 (Glyburide), in our case as well as in most of the MD-based predictions,<sup>74</sup> is a clear outlier, with a discrepancy between experimental and calculated value by more than four LogP units. In parentheses in the correlation plots of Fig. 7, we have reported the values of  $R$ ,  $\tau$ , MAE, and CCC computed by discarding the outlier glyburide, with  $R$  hitting 0.90 and with a MAE below 1 LogP unit.

It should be stressed that the deviations from the experimental values in our blind prediction mirror the discrepancies between the calculated and experimental transfer free energy from 1-octanol to toluene (evaluated from the knowledge of experimental LogP<sub>ow</sub> and LogP<sub>tw</sub>; see the caption of Fig. S1 in the supplementary material). In nine cases, these differences (see Fig. S1 of the supplementary material) are below 1 kcal/mol, and in only one case (compound 8), they exceed 2 kcal/mol. Glyburide (8) is a challenging test for unidirectional NE-FG calculations. Such a compound is the bulkiest (61 atoms) among all SAMPL9 molecules, bearing amide, halogen, sulfonic, phenyl, oxo, and amino moieties and a rotatable bond count of eight. The main source of the error for compound 8 is methodological, as it can be assessed by the unusually large confidence intervals found for the solvation energy in 1-octanol and in water (see Table S1 of the supplementary material). The statistical uncertainty of the solvation energy is strictly related to the variance of the fast-growth work distribution. The latter, in turn, is directly proportional to the dissipation of the NE unidirectional process. Hence, the NE growth in 1-octanol of the conformationally complex and bulky Glyburide is a strongly dissipative process because of the many disparate metastable conformations that the solute can exhibit in solution. In the case of 1-octanol, such large dissipation is due to the fact that the solvent is slow in rearranging the solvation shell, while the solute–solvent interaction is being switched on. In other words, 1-octanol as a solvent experiences a larger resistance to movement or diffusion of neighboring portions relative to one another

especially for large solutes, hence exhibiting a much larger viscosity with respect to toluene.

In water, whose viscosity is only marginally larger than that of toluene, the large dissipation is likely due to a combined effect of the conformational disorder and the polar nature of many glyburide moieties engaging in a variable number of strong H-bonds upon recharging, hence widening the work distribution. Very likely, the  $\approx 4, 3.5$  kcal/mol wide growth work distribution of compound 8 in 1-octanol and in water, respectively, is given by a mixtures of normal distributions due to disparate gas-phase starting conformations.<sup>75</sup> The number of work values (96) is too low to afford a reliable estimate of the coefficients of a mixture of Gaussian via, e.g., the expectation-maximization algorithm,<sup>76,77</sup> and the bias of the Jarzynski unidirectional estimate (see the section titled “Theoretical Aspect”), in this case, is overestimated, leading to a corresponding strong overestimation of the solvation energy in 1-octanol, negatively affecting the Log<sub>to</sub> estimate used in our blind prediction.

## V. COMPARATIVE ANALYSIS OF THE LOGP SAMPL9 SUBMISSIONS

Performances of all SAMPL9/LogP submissions are reported in Table S3 of the supplementary material. As expected, the best-performing methods are low-cost knowledge-based empirical approaches. Most of the QM-based approaches had an excellent correlation as measured by  $R$  and  $\tau$ , but they show an accuracy (measured by the MAE) somewhat below expectations compared to previous SAMPL results,<sup>30,31,78,79</sup> probably due to the not sufficiently optimized parameterization of the implicit model for the unusual toluene solvent. A full analysis including correlation plots for all submissions and histograms of deviations for all compounds can be found in Ref. 74

Here, we are interested in comparing MD-based submissions by assessing accuracy and precision in ranking (i.e., MAE and  $R$ ), with an eye on the benefit/cost ratio, an implicit requirement in this challenge, given that participants were requested to detail the central processing unit (CPU)/graphics processing unit (GPU) cost of their protocol in the submission template. We recall that MD-based approaches are systematically found among the best-performing

**TABLE II.** Comparative analysis of the MD-based blind predictions in the SAMPL0  $\text{LogP}_{\text{tw}}$  challenge (see text for details).

Subm.	Method	FF	Charges	Ranked	$t_{\text{sim}}/\text{ns}$	CCC	MAE	R
VoeltzLab	FEP/alchemy/EE	OpenFF-2.0	AM1-BCC	T	$\approx 3 \times 10^4$	0.82	1.26	0.90
NE-FG	NE/alchemy	GAFF2	AM1-BCC	T	90	0.82	1.12	0.83
Beckstein-Iorga	FEP/alchemy	GAFF/TIP3P	AM1-BCC	F	$\approx 250$	0.75	1.50	0.79
Beckstein-Iorga	FEP/alchemy	OPLS/M24	mol2ff	T	$\approx 250$	0.73	1.72	0.88
Beckstein-Iorga	FEP/alchemy	OPLS/TIP4P	CM1A	F	$\approx 250$	0.72	1.88	0.91
Sprick	FEP/alchemy/HREM	GAFF/TIP3P	IpolQ-Mod	T	240	0.47	3.03	0.66
Oxford	E-S/MCC	GAFF2/TIP3P	AM1-BCC	T	1200	0.42	1.78	0.44
MD (Patel)	FEP/alchemy	CGenFF/TIP4P	cGenFF	F	216	0.21	2.63	0.25

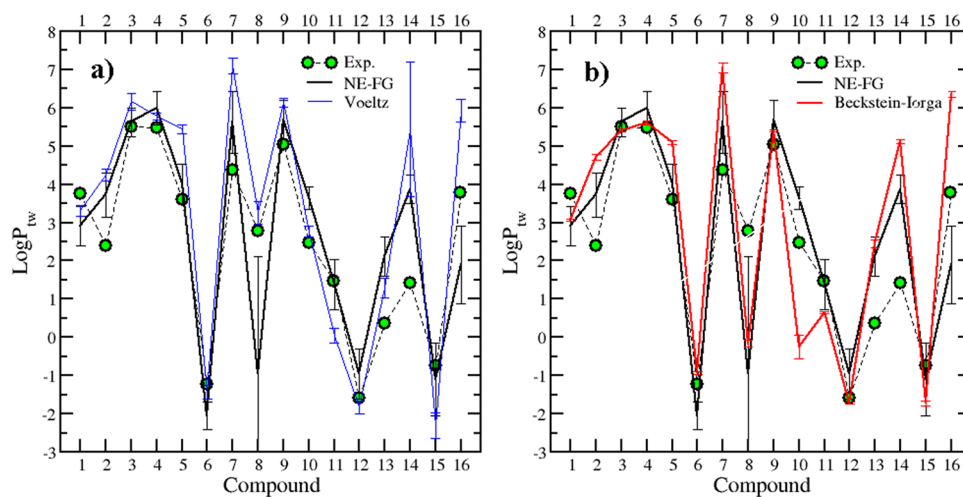
methods when dealing with drug–receptor or host–guest complexes. In Table II, we succinctly report the main methodological aspects of the eight MD-based submissions, along with their performances. Submissions are ranked according to the Lin concordance correlation coefficient, CCC. CCC ranges from  $-1$  and  $1$  and cannot exceed the Pearson correlation coefficient  $R$ . CCC effectively<sup>73</sup> measures the inter-rater reliability among two independent observables, combining the  $R$  metrics regarding ranking and the MAE outcome sensing the accuracy.

Among the eight submissions reported in Table II, six are done using the standard alchemical FEP method, based on  $\lambda$ -stratification. Two of these (VoeltzLab and Sprick) are implemented as  $\lambda$ -hopping techniques.<sup>80,81</sup> Our blind prediction is the only one produced using the nonequilibrium alchemical methodology. Finally, the Oxford submission relies on the end-state determination of the energy and entropy with the latter computed according to the so-called multi-scale cell correlation.<sup>82</sup> As to the force fields, all are based on the fixed-charge approach, with GAFF or GAFF2 used in four cases, the Optimized Potentials for Liquid Simulations (OPLS)<sup>18</sup> in two cases and CHARMM/CGenFF<sup>83</sup> and OpenFF<sup>84</sup> in one case. The cost reported in Table II is measured in invested simulation time (in ns) on a per-solute basis, deduced from the information reported in the methodological sections of the submissions csv files.<sup>85</sup> Such a measure is independent of the power of the MD engine and of the

hardware configuration. The best-performing methods according to CCC are the NE-FG and VoeltzLab. The latter, combining standard FEP with the so-called extended ensemble<sup>81</sup>  $\lambda$ -hopping approach, required, for a single  $\text{LogP}$  determination, a total simulation time in the order of the tens of microseconds, i.e., a cost  $\approx 300$  times higher than that of our NE-FG method.

In Fig. 8, we compare our prediction, based on the GAFF2 force field and AM1-BCC charge set, with the FEP predictions relying on the AM1-BCC parameterization for the solutes. As it can be seen, the three predictions sets, NE-FG, VoeltzLab, and Beckstein–Iorga, are strongly correlated one to the other with mutual  $R$  exceeding in all cases 0.85. This mutual correlation and the similar agreement with the experimental data occur despite the usage of three different force fields for the intramolecular and intermolecular Lennard–Jones interactions, highlighting the prominent role played by the fixed-charge modeling for the electrostatic contribution to the solvation free energies.

Figure 8 appears to confirm that the discrepancy observed for glyburide (compound 8) in our case is methodological due to the wide and non-Gaussian nature of the work distribution in the 1-octanol solvent used in our prediction (see Table S1 of the supplementary material), as the VoeltzLab and Beckstein–Iorga submissions predict a similar value for compound 8 much closer to the experiment. On the other hand, the error provided by the

**FIG. 8.** FEP or NE alchemical predictions based on the GAFF/AM1-BCC FF for all compounds. (a) NE-FG-calculated and VoeltzLab prediction. (b) Experimental values vs NE-FG-calculated and Beckstein-Iorga prediction. Lines are drawn as a guide for the eye.

NE-FG method via work bootstrapping, essentially proportional to the width of the work distribution, appears to be credible also in the case of compound 8, where the upper bound of the confidence interval is not far from the experimental value. The LogP SAMPL9 challenge has once again shown that the error determination is clearly a weak point in FEP calculations based on the equilibrium simulations of the intermediate  $\lambda$  states. Mean reported FEP errors<sup>74,85</sup> range from 0.05 (Beckstein–Iorga) to 0.43 (Patel) kcal/mol, with no apparent relation with the invested simulation time. The errors are computed in most cases by block averaging or bootstrapping in the intermediate  $\lambda$  states and summing in quadrature the resulting errors and in one case (Patel) by running the whole FEP calculation in triplicate.

## VI. CONCLUSION

We have presented our blind prediction for the toluene–water partition coefficients in the context of the SAMPL9 challenge using a very efficient MD-based nonequilibrium alchemical technique relying on the popular GAFF2 non-polarizable force field for the solutes. A pre-assessment was made based on the knowledge of the experimental octanol–water partition coefficients for all 16 compounds included in the challenge. The calculation of the solvation free energies in the various solvents were done using different approaches for the determination of the atomic charges on the SAMPL9 compounds. In particular, we used in our MD-based calculations of the solvation free energies the standard AM1-BCC atomic charges and the ESP atomic charges computed at the HF/6-31Gd level of theory. The comparison of calculated and experimental  $\text{LogP}_{ow}$  with various combination of the charge sets revealed that the HF/6-31Gd charges, yielding, in general, a stronger charge separation between atoms with disparate electronegativity, appear to be, in most cases, more appropriate for the water solvent. The analysis allowed to identify some important critical issues in the GAFF2 parameterization connected to the nitrogen atoms in cycles or in tertiary amine, providing a valid clue for improving the performance and reliability of non-polarizable force field in free energy calculations of drug–receptor systems. Based on the outcome of the pre-assessment, we therefore decided to use only the calculated solvation free energies of the non-polar toluene and 1-octanol solvents with the AM1-BCC charge set, exploiting the knowledge of the experimental value of the octanol–water partition coefficients to arrive at our blind submission for the SAMPL9 challenge. Our submission, based on nonequilibrium alchemy, was the best-performing and most efficient MD-based approach among the eight MD-based submissions.

## SUPPLEMENTARY MATERIAL

See the supplementary material for solvation free energies in water, 1-octanol, and toluene using AM1-BCC and ESP charges. Compact analysis of all ranked and not ranked LogP challenge submissions.  $\text{LogP}_{io}$  correlation plot.

## ACKNOWLEDGMENTS

The computing resources and the related technical support used for this work have been provided by the CRESCO/ENEAGRID

High Performance Computing infrastructure<sup>68</sup> and its staff. The CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and Sustainable Economic Development, and by Italian and European research programs (see [www.cresco.enea.it](http://www.cresco.enea.it) for information). The authors thank MIUR-Italy (“Progetto Dipartimenti di Eccellenza 2023–2027” allocated to Department of Chemistry “Ugo Schiff”) and the National Recovery and Resilience Plan, Mission 4 Component 2 - Investment 1.4 - NATIONAL CENTER FOR HPC, BIG DATA AND QUANTUM COMPUTING - funded by the European Union - NextGenerationEU - CUP B83C22002830001. We finally appreciate the National Institutes of Health for its support of the SAMPL project via R01GM124270 to David L. Mobley (UC Irvine).

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

### Author Contributions

**Piero Procacci:** Conceptualization (lead); Data curation (lead); Formal analysis (lead); Funding acquisition (equal); Project administration (lead); Software (equal); Writing – original draft (lead); Writing – review & editing (lead). **Guido Guarnieri:** Funding acquisition (equal); Project administration (supporting); Software (equal); Writing – original draft (supporting); Writing – review & editing (supporting).

## DATA AVAILABILITY

PDB trajectory files, raw work data, and force field parameter files are available at the general-purpose open-access repository <https://zenodo.org/record/7645899>. A complete analysis of the submissions of the LogP SAMPL9 challenge can be found at the GitHub site <https://procacci.github.io/LogP-SAMPL9/>.

The ORAC program (v6.1) is available for download under the General Public License at the website <http://www1.chim.unifi.it/orac/>.

Third-party software xLogP3 (v3.2.2) can be downloaded for free for academic users from the website <http://www.sioc-ccbg.ac.cn>.

## REFERENCES

- H. S. Muddana, A. T. Fenley, D. L. Mobley, and M. K. Gilson, “The SAMPL4 host–guest blind prediction challenge: An overview,” *J. Comput.-Aided Mol. Des.* **28**(4), 305–317 (2014).
- J. Yin, N. M. Henriksen, D. R. Slochower, M. R. Shirts, M. W. Chiu, D. L. Mobley, and M. K. Gilson, “Overview of the SAMPL5 host–guest challenge: Are we doing better?,” *J. Comput.-Aided Mol. Des.* **31**, 1–19 (2016).
- 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, “Overview of the SAMPL6 host–guest binding affinity prediction challenge,” *J. Comput.-Aided Mol. Des.* **32**(10), 937–963 (2018).
- A. Rizzi, T. Jensen, D. R. Slochower, M. Aldeghi, V. Gapsys, D. Ntekoimes, S. Bosisio, M. Papadourakis, N. M. Henriksen, B. L. de Groot, Z. Courmia, A. Dickson, J. Michel, M. K. Gilson, M. R. Shirts, D. L. Mobley, and J. D. Chodera, “The

- SAMPL6 SAMPLing challenge: Assessing the reliability and efficiency of binding free energy calculations,” *J. Comp. Aided Mol. Des.* **34**, 601–633 (2020).
- <sup>5</sup>M. Amezcua, L. El Khoury, and D. L. Mobley, “SAMPL7 host-guest challenge overview: Assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations,” *J. Comput.-Aided Mol. Des.* **35**(1), 1–35 (2021).
- <sup>6</sup>A. Pohorille, C. Jarzynski, and C. Chipot, “Good practices in free-energy calculations,” *J. Phys. Chem. B* **114**(32), 10235–10253 (2010).
- <sup>7</sup>J. D. Chodera, D. L. Mobley, M. R. Shirts, R. W. Dixon, K. Branson, and V. S. Pande, “Alchemical free energy methods for drug discovery: Progress and challenges,” *Curr. Opin. Struct. Biol.* **21**, 150–160 (2011).
- <sup>8</sup>P. Procacci, “Alchemical determination of drug-receptor binding free energy: Where we stand and where we could move to,” *J. Mol. Graphics Modell.* **71**, 233–241 (2017).
- <sup>9</sup>P. Procacci and M. Macchiagodena, “On the NS-DSSB unidirectional estimates in the SAMPL6 SAMPLing challenge,” *J. Comput.-Aided Mol. Des.* **35**(10), 1055–1065 (2021).
- <sup>10</sup>Z. Cournia, B. K. Allen, T. Beuming, D. A. Pearlman, B. K. Radak, and W. Sherman, “Rigorous free energy simulations in virtual screening,” *J. Chem. Inf. Model.* **60**, 4153–4169 (2020).
- <sup>11</sup>Z. Cournia, C. Chipot, B. Roux, D. M. York, and W. Sherman, *Free Energy Methods in Drug Discovery—Introduction* (American Chemical Society, Washington, DC, 2021), Chap. 1, pp. 1–38.
- <sup>12</sup>V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, “Comparison of multiple Amber force fields and development of improved protein backbone parameters,” *Proteins: Struct., Funct., Bioinf.* **65**(3), 712–725 (2006).
- <sup>13</sup>R. Salomon-Ferrer, D. A. Case, and R. C. Walker, “An overview of the amber biomolecular simulation package,” *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **3**(2), 198–210 (2013).
- <sup>14</sup>A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, “All-atom empirical potential for molecular modeling and dynamics studies of proteins,” *J. Phys. Chem. B* **102**(18), 3586–3616 (1998).
- <sup>15</sup>C. Zhang, C. Lu, Z. Jing, C. Wu, J.-P. Piquemal, J. W. Ponder, and P. Ren, “AMOEBa polarizable atomic multipole force field for nucleic acids,” *J. Chem. Theory Comput.* **14**(4), 2084–2108 (2018).
- <sup>16</sup>J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, “Development and testing of a general amber force field,” *J. Comput. Chem.* **25**, 1157–1174 (2004).
- <sup>17</sup>GAFF and GAFF2 are public domain force fields and are part of the AmberTools distribution, available for download at <https://amber.org> internet address (accessed January 2022). According to the AMBER development team, the improved version of GAFF, GAFF2, is an ongoing project aimed at “reproducing both the high quality interaction energies and key liquid properties such as density, heat of vaporization and hydration free energy.” GAFF2 is expected “to be an even more successful general purpose force field and that GAFF2-based scoring functions will significantly improve the successful rate of virtual screenings.”
- <sup>18</sup>L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, and W. L. Jorgensen, “LigParGen web server: An automatic OPLS-AA parameter generator for organic ligands,” *Nucleic Acids Res.* **45**(W1), W331–W336 (2017).
- <sup>19</sup>P. Procacci, “PrimaDORAC: A free web interface for the assignment of partial charges, chemical topology, and bonded parameters in organic or drug molecules,” *J. Chem. Inf. Model.* **57**(6), 1240–1245 (2017).
- <sup>20</sup>Y. Qiu, D. G. A. Smith, S. Boothroyd, H. Jang, D. F. Hahn, J. Wagner, C. C. Bannan, T. Gokey, V. T. Lim, C. D. Stern, A. Rizzi, B. Tjanaka, G. Tresadern, X. Lucas, M. R. Shirts, M. K. Gilson, J. D. Chodera, C. I. Bayly, D. L. Mobley, and L. P. Wang, “Development and benchmarking of open force field v1.0.0—the parsley small-molecule force field,” *J. Chem. Theory Comput.* **17**(10), 6262–6280 (2021).
- <sup>21</sup>See <https://github.com/samplchallenges/SAMPL9> for SAMPL9: Experimental results; accessed 13 January 2022.
- <sup>22</sup>M. K. Gilson, J. A. Given, B. L. Bush, and J. A. McCammon, “The statistical-thermodynamic basis for computation of binding affinities: A critical review,” *Biophys. J.* **72**, 1047–1069 (1997).
- <sup>23</sup>H.-X. Zhou and M. K. Gilson, “Theory of free energy and entropy in noncovalent binding,” *Chem. Rev.* **109**, 4092–4107 (2009).
- <sup>24</sup>P. Procacci and R. Chelli, “Statistical mechanics of ligand–receptor noncovalent association, revisited: Binding site and standard state volumes in modern alchemical theories,” *J. Chem. Theory Comput.* **13**(5), 1924–1933 (2017).
- <sup>25</sup>T. Cheng, Y. Zhao, X. Li, F. Lin, Y. Xu, X. Zhang, Y. Li, R. Wang, and L. Lai, “Computation of octanol–water partition coefficients by guiding an additive model with knowledge,” *J. Chem. Inf. Model.* **47**(6), 2140–2148 (2007).
- <sup>26</sup>A. K. Ghose, V. N. Viswanadhan, and J. J. Wendoloski, “Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: An analysis of ALOGP and CLOGP methods,” *J. Phys. Chem. A* **102**(21), 3762–3772 (1998).
- <sup>27</sup>See <https://www.molinspiration.com/> for Molinspiration cheminformatics software; accessed 6 May 2019.
- <sup>28</sup>A. C. Spiess, E. Werner, M. Peters, M. F. Eckstein, L. Greiner, and J. Büchs, “Prediction of partition coefficients using COSMO-RS: Solvent screening for maximum conversion in biocatalytic two-phase reaction systems,” *Chem. Eng. Process.: Process Intensif.* **47**(6), 1034–1041 (2008), *Biothermodynamics*.
- <sup>29</sup>A. V. Marenich, C. J. Cramer, and D. G. Truhlar, “Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions,” *J. Phys. Chem. B* **113**(18), 6378–6396 (2009).
- <sup>30</sup>M. Isik, T. D. Bergazin, T. Fox, A. Rizzi, J. D. Chodera, and D. L. Mobley, “Assessing the accuracy of octanol–water partition coefficient predictions in the SAMPL6 Part II log P Challenge,” *J. Comput.-Aided Mol. Des.* **34**(4), 335–370 (2020).
- <sup>31</sup>T. D. Bergazin, N. Tielker, Y. Zhang, J. Mao, M. R. Gunner, K. Francisco, C. Ballatore, S. M. Kast, and D. L. Mobley, “Evaluation of log P, pK<sub>a</sub>, and log D predictions from the SAMPL7 blind challenge,” *J. Comput.-Aided Mol. Des.* **35**(7), 771–802 (2021).
- <sup>32</sup>W. L. Jorgensen, J. K. Buckner, S. Boudon, and J. Tirado-Rives, “Efficient computation of absolute free energies of binding by computer simulations. application to the methane dimer in water,” *J. Chem. Phys.* **89**, 3742–3746 (1988).
- <sup>33</sup>P. Procacci, “Accuracy, precision, and efficiency of nonequilibrium alchemical methods for computing free energies of solvation. I. Bidirectional approaches,” *J. Chem. Phys.* **151**(14), 144113 (2019).
- <sup>34</sup>P. Procacci, “Precision and computational efficiency of nonequilibrium alchemical methods for computing free energies of solvation. II. Unidirectional estimates,” *J. Chem. Phys.* **151**(14), 144115 (2019).
- <sup>35</sup>P. Procacci, M. Guarrasi, and G. Guarnieri, “SAMPL6 host–guest blind predictions using a non equilibrium alchemical approach,” *J. Comput.-Aided Mol. Des.* **32**, 965–982 (2018).
- <sup>36</sup>P. Procacci and G. Guarnieri, “SAMPL6 blind predictions of water-octanol partition coefficients using nonequilibrium alchemical approaches,” *J. Comput.-Aided Mol. Des.* **34**, 371–384 (2019).
- <sup>37</sup>P. Procacci, “Methodological uncertainties in drug-receptor binding free energy predictions based on classical molecular dynamics,” *Curr. Opin. Struct. Biol.* **67**, 127–134 (2021).
- <sup>38</sup>R. W. Zwanzig, “High-temperature equation of state by a perturbation method. I. Nonpolar gases,” *J. Chem. Phys.* **22**, 1420–1426 (1954).
- <sup>39</sup>Z. X. Sun, X. H. Wang, and J. Z. H. Zhang, “BAR-based optimum adaptive sampling regime for variance minimization in alchemical transformation,” *Phys. Chem. Chem. Phys.* **19**, 15005–15020 (2017).
- <sup>40</sup>P. Procacci, “Solvation free energies via alchemical simulations: let’s get honest about sampling, once more,” *Phys. Chem., Chem. Phys.* **21**, 13826–13834 (2019).
- <sup>41</sup>P. Procacci, “Does Hamiltonian replica exchange via lambda-hopping enhance the sampling in alchemical free energy calculations?,” *Molecules* **27**(14), 4426 (2022).
- <sup>42</sup>H. Fukunishi, O. Watanabe, and S. Takada, “On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction,” *J. Chem. Phys.* **116**, 9058–9067 (2002).
- <sup>43</sup>G. E. Crooks, “Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems,” *J. Stat. Phys.* **90**, 1481–1487 (1998).
- <sup>44</sup>C. H. Bennett, “Efficient estimation of free energy differences from Monte Carlo data,” *J. Comput. Phys.* **22**, 245–268 (1976).

- <sup>45</sup>M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, "Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods," *Phys. Rev. Lett.* **91**, 140601 (2003).
- <sup>46</sup>C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.* **78**, 2690–2693 (1997).
- <sup>47</sup>P. Procacci, S. Marsili, A. Barducci, G. F. Signorini, and R. Chelli, "Crooks equation for steered molecular dynamics using a Nosé-Hoover thermostat," *J. Chem. Phys.* **125**, 164101 (2006).
- <sup>48</sup>S. Park and K. Schulten, "Calculating potentials of mean force from steered molecular dynamics simulations," *J. Chem. Phys.* **120**(13), 5946–5961 (2004).
- <sup>49</sup>S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang, and S. H. Bryant, "PubChem substance and compound databases," *Nucleic Acids Res.* **44**(D1), D1202–D1213 (2016).
- <sup>50</sup>D. Vassetz, M. Pagliai, and P. Procacci, "Assessment of GAFF2 and OPLS-AA general force fields in combination with the water models TIP3P, SPCE, AND OPC3 for the solvation free energy of druglike organic molecules," *J. Chem. Theory Comput.* **15**(3), 1983–1995 (2019).
- <sup>51</sup>G. Hummer, "Fast-growth thermodynamic integration: Error and efficiency analysis," *J. Chem. Phys.* **114**, 7330–7337 (2001).
- <sup>52</sup>J. Gore, F. Ritort, and C. Bustamante, "Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements," *Proc. Natl. Acad. Sci.* **100**(22), 12564–12569 (2003).
- <sup>53</sup>T. W. Anderson and D. A. Darling, "A test of goodness of fit," *J. Am. Stat. Assoc.* **49**, 765–769 (1954).
- <sup>54</sup>N. M. Razali and Y. B. Wah, "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *J. Stat. Model. Anal.* **2**, 21–33 (2011).
- <sup>55</sup>J. Lorentz and S. D. Bolboacă, "Computation of probability associated with Anderson-Darling statistic," *Mathematics* **6**(6), 88 (2018).
- <sup>56</sup>N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open Babel: An open chemical toolbox," *J. Cheminf.* **3**(1), 33 (2011).
- <sup>57</sup>J. Adam, R. Adam, I. Ladjanski, and G. Cserey, "Optimized GPU implementation of Merck Molecular Force Field and Universal Force Field," *J. Mol. Struct.* **1188**, 227–233 (2019).
- <sup>58</sup>A. Jakalian, D. B. Jack, and C. I. Bayly, "Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation," *J. Comput. Chem.* **23**(16), 1623–1641 (2002).
- <sup>59</sup>C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model," *J. Phys. Chem.* **97**(40), 10269–10280 (1993).
- <sup>60</sup>S. Izadi and A. V. Onufriev, "Accuracy limit of rigid 3-point water models," *J. Chem. Phys.* **145**(7), 074501 (2016).
- <sup>61</sup>Z. Sun, Z. Huai, Q. He, and Z. Liu, "A general picture of cucurbit[8]uril host-guest binding," *J. Chem. Inf. Model.* **61**(12), 6107–6134 (2021).
- <sup>62</sup>S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *J. Chem. Phys.* **81**(1), 511–519 (1984).
- <sup>63</sup>S. Marsili, G. F. Signorini, R. Chelli, M. Marchi, and P. Procacci, "ORAC: A molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level," *J. Comput. Chem.* **31**, 1106–1116 (2010).
- <sup>64</sup>M. Marchi and P. Procacci, "Coordinates scaling and multiple time step algorithms for simulation of solvated proteins in the NPT ensemble," *J. Chem. Phys.* **109**, 5194–5202 (1998).
- <sup>65</sup>T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, "Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations," *Chem. Phys. Lett.* **222**, 529–539 (1994).
- <sup>66</sup>P. Procacci, T. A. Darden, E. Paci, and M. Marchi, "ORAC: A molecular dynamics program to simulate complex molecular systems with realistic electrostatic interactions," *J. Comput. Chem.* **18**, 1848–1862 (1997).
- <sup>67</sup>U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," *J. Chem. Phys.* **103**, 8577–8593 (1995).
- <sup>68</sup>F. Iannone, F. Ambrosino, G. Bracco, M. De Rosa, A. Funel, G. Guarnieri, S. Migliori, F. Palombi, G. Ponti, G. Santomauro, and P. Procacci, "CRESCO ENEA HPC clusters: A working example of a multifabric GPFS spectrum scale layout," in *2019 International Conference on High Performance Computing Simulation (HPCS)*, (2019), pp. 1051–1052.
- <sup>69</sup>P. Procacci, "Hybrid MPI/OpenMP implementation of the ORAC molecular dynamics program for generalized ensemble and fast switching alchemical simulations," *J. Chem. Inf. Model.* **56**(6), 1117–1121 (2016).
- <sup>70</sup>M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, "AM1: A new general purpose quantum mechanical model," *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).
- <sup>71</sup>M. Pagliai, G. Funghi, D. Vassetz, P. Procacci, R. Chelli, and G. Cardini, "Imidazole in aqueous solution: Hydrogen bond interactions and structural reorganization with concentration," *J. Phys. Chem. B* **123**(18), 4055–4064 (2019).
- <sup>72</sup>I. Soteras Gutiérrez, F.-Y. Lin, K. Vanommeslaeghe, J. A. Lemkul, K. A. Armacost, C. L. Brooks, and A. D. MacKerell, "Parameterization of halogen bonds in the CHARMM general force field: Improved treatment of ligand-protein interactions," *Bioorg. Med. Chem.* **24**(20), 4812–4825 (2016), *Advances in Computational and Medicinal Chemistry*.
- <sup>73</sup>L. I.-K. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics* **45**, 255–268 (1989).
- <sup>74</sup>See <https://procacci.github.io/LogP-SAMPL9/> for reported errors in SAMPL9 blind MD-based submissions; accessed 19 February 2023.
- <sup>75</sup>P. Procacci, "Unbiased free energy estimates in fast nonequilibrium transformations using Gaussian mixtures," *J. Chem. Phys.* **142**(15), 154117 (2015).
- <sup>76</sup>A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc., B* **39**, 1–38 (1977).
- <sup>77</sup>M. R. Gupta and Y. Chen, *Theory and use of the EM algorithm* (Foundations and Trends in Signal Processing, 2011), Vol. 4, issue 3, pp. 223–296.
- <sup>78</sup>C. C. Bannan, K. H. Burley, M. Chiu, M. R. Shirts, M. K. Gilson, and D. L. Mobley, "Blind prediction of cyclohexane-water distribution coefficients from the SAMPL5 challenge," *J. Comput.-Aided Mol. Des.* **30**(11), 927–944 (2016).
- <sup>79</sup>J. Warnau, K. Wichmann, and J. Reinisch, "COSMO-RS predictions of logP in the SAMPL7 blind challenge," *J. Comput.-Aided Mol. Des.* **35**(7), 813–818 (2021).
- <sup>80</sup>C. J. Woods, J. W. Essex, and M. A. King, "Enhanced configurational sampling in binding free-energy calculations," *J. Phys. Chem. B* **107**(49), 13711–13718 (2003).
- <sup>81</sup>S. Zhang, D. F. Hahn, M. R. Shirts, and V. A. Voelz, "Expanded ensemble methods can be used to accurately predict protein-ligand relative binding free energies," *J. Chem. Theory Comput.* **17**(10), 6536–6547 (2021).
- <sup>82</sup>H. S. Ali, J. Higham, and R. H. Henchman, "Entropy of simulated liquids using multiscale cell correlation," *Entropy* **21**(8), 750 (2019).
- <sup>83</sup>K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell, "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," *J. Comput. Chem.* **31**(4), 671–690 (2010).
- <sup>84</sup>D. L. Mobley, C. C. Bannan, A. Rizzi, C. I. Bayly, J. D. Chodera, V. T. Lim, N. M. Lim, K. A. Beauchamp, D. R. Slochower, M. R. Shirts, M. K. Gilson, and P. K. Eastman, "Escaping atom types in force fields using direct chemical perception," *J. Chem. Theory Comput.* **14**(11), 6076–6092 (2018).
- <sup>85</sup>See <https://github.com/samplchallenges/SAMPL9/tree/main/logP> for SAMPL9: Analysis and experimental results for logp(tw); accessed 20 February 2023.