Propagation of uncertainty in physicochemical data to force field predictions

Ahmet Yildirim[ID]
Department of Physics, Siirt University, 56100 Siirt, Turkey

Mohammad Mehdi Ghahremanpour[ID]
Department of Chemistry, Yale University, New Haven, Connecticut 06520-8107, USA

David van der Spoel[ID][*]
Uppsala Centre for Computational Chemistry, Science for Life Laboratory, and Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, 75124 Uppsala, Sweden

The solvation free energy (SFE) is a key property in the thermodynamics of chemical processes. It can be evaluated using molecular simulations with good statistical accuracy. However, force field predictions exhibit systematic errors due to uncertainties in the parametrization. Here we evaluate how the uncertainty in physicochemical data underlying force fields propagates to SFE predictions. We find that the data contribution to the uncertainty in SFE is up to 25 times larger than the statistical uncertainty. The total uncertainty in the SFE in water is higher than in cyclohexane.

I. INTRODUCTION

Accurate prediction of absolute or relative free energies of molecular systems in the liquid phase remains a challenging endeavor. The multistate Bennett acceptance ratio [1] estimator based on equilibrium molecular dynamics (MD) simulations makes it possible to get converged results. Recently, there has been renewed interest in nonequilibrium methods based on the Jarzynski equality [2] or the Crooks Gaussian intersection method [3,4], although both the uncertainties in results and efficiencies of such methods remain under scrutiny [5–7]. While aleatory uncertainties due to fluctuations and sampling are now largely under control and standard protocols are available [1,8], studies of uncertainty quantification are now being focused on epistemic uncertainties, i.e., those due to a lack of knowledge of the potentials, the simulations methods, and the parameters [9–12]. Here we assess the epistemic uncertainty due to propagation of measurement errors in electric multipole moments and liquid density.

Most research on computing solvation free energies (SFEs) is based on classical empirical potentials because describing the electronic structure of molecules is too expensive in the liquid phase to make it feasible to obtain adequate sampling. However, the accuracy and reliability of classical potentials are largely dependent on the parameters in conjunction with the form of the potential. The parameters are typically fine-tuned to reproduce specific properties for a limited number of molecules [13–15]. This means that a unique set of parameters is chosen from a parameter space that is often high dimensional with a rugged landscape [16]. As a result, it is important to systematically quantify the uncertainty in the outcome of the molecular simulations that are caused by the uncertainty in the calibrated force field parameters. In this context, we consider here the finite accuracy of reference data for building force fields, a mostly overlooked aspect of predictions from force field simulations. Uncertainties in reference data limit the accuracy that can be obtained in SFE predictions, even in the hypothetical case of infinite sampling and (hypothetical) perfect models, that is, models that reproduce the experimental data exactly.

The level of accuracy in the SFE and other properties obtained from MD simulations depends on the force field chosen as well as on other factors concerning computational procedure such as simulation time and size of the system [17,18]. Modification of existing force fields or development of new force fields can in principle improve the level of accuracy in the free energy predictions [19]. The SFE in water, or hydration free energy (HFE), is an essential property in the development and assessment of the accuracy of the force fields [20–25]. The fast convergence and sampling of HFE calculations enables the widespread use of those calculations for the evaluation of the accuracy of sampling methods and comparison of different theoretical approaches [5].

Here we establish how the SFE of 11 small organic molecules in water and cyclohexane vary due to changes in the molecular volume (by scaling the van der Waals radius σ) and due to changes in the electric moment by scaling the

---

[*]david.vanderspoel@icm.uu.se

TABLE I. Experimental data of compounds. Quadrupole θ [in Buckingham (equivalent to 1 D Å)], liquid density ρ (g/l), and dipole μ in D.

IUPAC name	θ_{zz}	$\delta\theta_{\text{expt}}$	ρ	$\delta\rho_{\text{expt}}$
benzene	−8.69 [27]	0.51	873.43 [28]	0.37[a]
naphthalene	−13.52 [29]	1.50	976.72 [30]	[b]
anthracene	−7.74 [31]	0.49[a]	1280 [32]	[b]
IUPAC name	μ	$\delta\mu_{\text{expt}}$	ρ	$\delta\rho_{\text{expt}}$
methanol	1.70 [32]	0.02	786.47 [33]	0.14[a]
ethanol	1.69 [32]	0.03	785.13 [33]	0.17[a]
phenol	1.22 [32]	0.01	1071.6 [34,35]	0.42[a]
p-Cresol	1.48 [32]	[b]	1018.5 [32]	[c]
piperazine	1.47 [36]	0.02	1100 [37]	[c]
1-methylpiperazine	1.14 [36]	0.01	930[d] [38]	[c]
propionamide	3.85 [39]	[b]	926.2 [32]	[c]
hexanamide	3.93 [40]	[b]	999 [32]	[c]

[a]These errors are standard deviations over multiple experimental values.

[b]For compounds where no experimental error in the dipole is known, we assume the average relative error as in the other compounds, that is, 1.2%.

[c]We assume the average relative error in the density to be 0.000 31.

[d]Solid density.

atomic charges q relative to the force fieldcharges q_0. Both the magnitude of the molecular dipole μ and the quadrupole θ scale linearly with q,

$$\frac{\delta q}{q_0} = \frac{\delta\mu}{\mu_0} = \frac{\delta\theta}{\theta_0}, \tag{1}$$

which means the relative change in the charges is equal to the relative change in the electric moments. The molecular volume V is a nonlinear function of the van der Waals radii σ and it is computed using the atomic and bond contributions of van der Waals volume method [26]. The experimental observables for all solutes are listed in Table I. We then combine the simulation results with experimental data and use error propagation to show how the uncertainty in experimental properties affects the final SFE results.

II. METHODS

A. Uncertainty quantification

The variance $\text{var}(\hat{f})$ in SFE due to the uncertainty in the molecular van der Waals volume V and in the charge q can be

calculated by the error propagation formula [41]

$$\text{var}(\hat{f}) = \left[\frac{\partial f}{\partial V}\right]^2 \Sigma_{VV} + \left[\frac{\partial f}{\partial q}\right]^2 \Sigma_{qq}$$
$$+ 2\left[\frac{\partial f}{\partial V}\right]\left[\frac{\partial f}{\partial q}\right]\Sigma_{Vq}, \tag{2}$$

where f is the free energy function, $\frac{\partial f}{\partial V}$ is the derivative of the SFE with respect to V, Σ_{VV} is the variance of V, and $\frac{\partial f}{\partial q}$ is the derivative of the SFE with respect to q. Here Σ_{qq} follows from the experimental uncertainty in the dipole and quadrupole using Eq. (1), that is,

$$\frac{\Sigma_{qq}}{q_0^2} = \frac{\Sigma_{\mu\mu}}{\mu_0^2} \tag{3}$$

and similarly for quadrupolar compounds. Here Σ_{Vq} is the covariance between V and q and to estimate it we can apply the Cauchy-Schwarz inequality [42], which reads

$$\Sigma_{V,q}^2 \leqslant \Sigma_{VV}\Sigma_{qq}. \tag{4}$$

If there is a correlation between the variables then the upper limit is the number on the right-hand side of this equation, while the lower limit (no correlation) is zero by definition. We assume there is no correlation between the values of the electric moments, which are measured in the gas phase, and the liquid density. Force fields are often parametrized on a combination of the liquid density and the enthalpy of vaporization and in such cases the correlation may be important [43]. Finally, the variance of the molecular volume V is estimated from the variance in the density ρ of the solute in the liquid phase

$$\frac{\Sigma_{VV}}{V^2} = \frac{\Sigma_{\rho\rho}}{\rho^2}. \tag{5}$$

This notion connects the microscopic variable from the simulation system to macroscopic observables. With this, we can compute the uncertainty

$$\sigma(\hat{f}) = \sqrt{\text{var}(\hat{f})}. \tag{6}$$

B. Simulation details

All simulations were carried out with the molecular-dynamics software package GROMACS [44–48]. The generalized AMBER force field (GAFF) [13] topology parameters and the initial coordinates of compounds were taken from
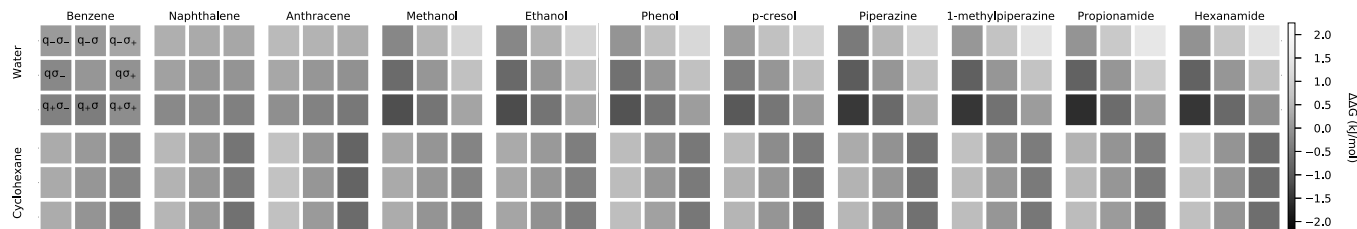

FIG. 1. Plot of $\Delta\Delta G_{solv}$ for all solute-solvent combinations with all combinations of $\pm q$ and $\pm\sigma$ as indicated in the benzene-water combination, averaged over three force fields. The Gibbs energy difference plotted is relative to the default q and σ.

TABLE II. CGenFF: compounds with their SFE with statistical uncertainty due to sampling σ_{SFE}, $q_0 \frac{\partial \Delta G}{\partial q}$, $V_0 \frac{\partial \Delta G}{\partial V}$, and $\sigma(\hat{f})$, all in kJ/mol.

IUPAC name	Water					Cyclohexane				
	SFE	σ_{SFE}	$q_0 \frac{\partial \Delta G}{\partial q}$	$V_0 \frac{\partial \Delta G}{\partial V}$	$\sigma(\hat{f})$	SFE	σ_{SFE}	$q_0 \frac{\partial \Delta G}{\partial q}$	$V_0 \frac{\partial \Delta G}{\partial V}$	$\sigma(\hat{f})$
benzene	-2.5	0.10	-10.8	-0.7	0.7	-17.0	0.09	-5.6	-9.8	0.4
naphthalene	-9.7	0.12	-41.0	-6.4	4.6	-29.9	0.11	-11.7	-13.6	1.3
anthracene	-16.0	0.15	-25.8	-9.8	1.7	-42.0	0.14	33.6	-21.1	2.2
methanol	-19.2	0.08	-71.7	18.0	0.9	-5.0	0.05	1.0	-3.5	0.1
ethanol	-19.2	0.09	-70.1	15.6	1.3	-8.6	0.07	-2.7	-5.8	0.1
phenol	-20.9	0.11	-66.5	11.8	0.6	-21.1	0.09	1.4	-13.1	0.1
p-Cresol	-20.0	0.12	-76.4	6.3	0.7	-24.2	0.10	9.8	-13.9	0.1
piperazine	-18.6	0.12	-64.2	16.8	0.9	-14.0	0.09	10.1	-9.6	0.2
1-methylpiperazine	-29.4	0.13	-72.1	18.1	0.7	-17.8	0.10	13.1	-11.7	0.2
propionamide	-33.5	0.10	-109.3	15.0	0.4	-16.5	0.08	-7.0	-7.2	0.1
hexanamide	-32.1	0.13	-76.7	11.1	0.3	-25.8	0.11	11.1	-16.6	0.1

the FreeSolv database [49,50]. We used the CHARMM general force field (CGenFF) server [51] for the CGenFF parameters and the LigParGen server [52,53] for the all-atom optimized potential for liquid simulations (OPLS-AA) parameters with localized bond-charge corrected CM1A charges for condensed-phase simulations [14]. In all simulations, we used the transferable intermolecular potential three-point water model [54], the linear constraint solver algorithm [55] for hydrogen bond constraints, and the SETTLE algorithm [56] to keep the water bonds and angle rigid. The leapfrog stochastic dynamics integrator [57] with a time step of 2 fs was used to integrate the equations of motion. Temperature was held constant with Langevin dynamics [57,58] with a coupling constant of 1.0 ps and a reference temperature of 298.15 K. At the equilibrium stage of the simulations the Berendsen barostat [59] was used to establish a pressure of 1 bar and the Parrinello-Rahman barostat [60] was used at production stage of the simulations, with a time constant of 1.0 ps and a compressibility of 4.5×10^{-5} bar^{-1}. To treat the electrostatic interactions, the particle mesh Ewald algorithm [61] was used with a cutoff of 1.2 nm, a grid spacing of 0.12 nm, and an interpolation order of 6 for long-range electrostatics. The van der Waals interactions were computed with a cutoff radius of 1.1 nm.

In the free energy simulations, the electrostatic and van der Waals interactions of compounds were modified using total 21 λ values: 5 λ values for electrostatic interactions (0, 0.25, 0.5, 0.75, and 1) and 16 λ values for van der Waals interactions (0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, and 1). A soft-core potential [62] was used for the van der Waals interactions. The simulation time was 5 ns for each λ window. The results of free energy simulations were analyzed using the alchemical analysis tool [63] with the multistate Bennett acceptance ratio (MBAR) method [64]. The molecular volume of compounds is computed using the atomic and bond contributions of van der Waals volume method [26].

III. RESULTS AND DISCUSSION

The SFE was computed for all combinations of the variables σ and q of the solutes modified by $\pm 1\%$ in both solvents in three different force fields. The MBAR method, which produces good statistical accuracy at a modest computational cost, was used for SFE calculations. The changes in SFE due to changing q and σ are plotted in Fig. 1. Interestingly, the two solvents behave qualitatively different from each other. In cyclohexane, increased molecular volume yields to lower

TABLE III. GAFF: compounds with their SFE with statistical uncertainty due to sampling σ_{SFE}, $q_0 \frac{\partial \Delta G}{\partial q}$, $V_0 \frac{\partial \Delta G}{\partial V}$, and $\sigma(\hat{f})$, all in kJ/mol.

IUPAC name	Water					Cyclohexane				
	SFE	σ_{SFE}	$q_0 \frac{\partial \Delta G}{\partial q}$	$V_0 \frac{\partial \Delta G}{\partial V}$	$\sigma(\hat{f})$	SFE	σ_{SFE}	$q_0 \frac{\partial \Delta G}{\partial q}$	$V_0 \frac{\partial \Delta G}{\partial V}$	$\sigma(\hat{f})$
benzene	-3.3	0.10	-46.4	4.5	2.8	-14.2	0.10	6.1	-11.1	0.4
naphthalene	-13.7	0.12	-22.4	-0.1	2.5	-26.9	0.13	13.7	-17.9	1.6
anthracene	-21.8	0.14	-58.2	-2.9	3.7	-39.6	0.16	1.1	-19.9	0.1
methanol	-14.5	0.07	-58.7	14.1	0.7	-3.6	0.06	-9.2	-2.6	0.2
ethanol	-14.3	0.08	-59.4	15.0	1.1	-7.5	0.08	6.2	-6.0	0.2
phenol	-23.6	0.11	-95.5	16.5	0.8	-18.9	0.11	-5.1	-11.4	0.1
p-Cresol	-23.6	0.12	-86.4	15.2	0.8	-22.3	0.12	-1.7	-13.1	0.1
piperazine	-35.2	0.12	-77.1	16.7	1.1	-23.0	0.11	-17.5	-9.8	0.3
1-methylpiperazine	-34.2	0.12	-113.3	16.5	1.0	-26.0	0.12	34.7	-13.0	0.4
propionamide	-34.9	0.10	-98.8	15.8	0.4	-15.5	0.09	14.3	-8.8	0.1
hexanamide	-35.0	0.13	-113.8	10.4	0.4	-24.8	0.13	8.2	-18.6	0.1

TABLE IV. OPLS: compounds with their SFE with statistical uncertainty due to sampling σ_{SFE}, $q_0\frac{\partial \Delta G}{\partial q}$, $V_0\frac{\partial \Delta G}{\partial V}$, and $\sigma(\hat{f})$, all in kJ/mol.

IUPAC name	Water					Cyclohexane				
	SFE	σ_{SFE}	$q_0\frac{\partial \Delta G}{\partial q}$	$V_0\frac{\partial \Delta G}{\partial V}$	$\sigma(\hat{f})$	SFE	σ_{SFE}	$q_0\frac{\partial \Delta G}{\partial q}$	$V_0\frac{\partial \Delta G}{\partial V}$	$\sigma(\hat{f})$
benzene	−5.4	0.10	−38.1	2.0	2.3	−16.8	0.10	−3.0	−9.2	0.2
naphthalene	−16.3	0.12	−37.0	−0.4	4.1	−30.1	0.12	9.1	−11.7	1.1
anthracene	−24.8	0.15	−73.2	1.2	4.7	−43.2	0.15	−14.5	−19.6	1.0
methanol	−19.7	0.08	−75.6	17.4	0.9	−3.2	0.06	−2.0	−4.1	0.1
ethanol	−20.1	0.09	−76.4	19.6	1.4	−7.0	0.07	−6.2	−5.1	0.2
phenol	−26.1	0.13	−82.7	17.4	0.7	−20.3	0.10	0.8	−11.3	0.1
p-Cresol	−26.4	0.14	−79.8	14.7	0.7	−23.0	0.11	6.5	−12.0	0.1
piperazine	−39.7	0.12	−114.6	22.2	1.6	−20.8	0.10	−12.7	−11.0	0.2
1-methylpiperazine	−37.2	0.13	−84.5	21.7	0.8	−24.4	0.12	4.3	−13.2	0.1
propionamide	−36.6	0.13	−108.8	34.9	0.4	−15.0	0.09	16.1	−9.8	0.1
hexanamide	−34.8	0.15	−123.3	33.7	0.4	−23.6	0.12	−13.9	−12.0	0.1

SFE, while the charge has less influence. This is expected [65] since the cyclohexane model does not include polarizability. In water, the SFE is dominated by electrostatic interactions and therefore both increasing the charge and reducing the volume leads to lower SFE, except for the large apolar solutes naphthalene and anthracene, where increased volume decreases the SFE, likely due to loss of water-water interactions [66].

With equations and data in place, we can now inspect the SFE, its statistical error due to sampling σ_{SFE}, and the uncertainty due to the data. Tables II–IV show that the SFE

in water varies from approximately −3 to −37 kJ/mol and in cyclohexane from approximately −3 to −42 kJ/mol, depending slightly on the force field used. The uncertainty due to sampling varies between 0.06 and 0.16 kJ/mol. However, the uncertainty due to the data $\sigma(\hat{f})$ used in model derivation is much larger (Fig. 2). For cyclohexane it varies from 0.1 to 2 kJ/mol and in water the data-derived uncertainties $\sigma(\hat{f})$ vary between 0.4 and 4 kJ/mol. The $\sigma(\hat{f})$ for the three force fields are similar in most cases (Fig. 2).

Figure 3 shows that the two SFE derivatives (numbers in Tables II–IV) are highly correlated (Pearson $r > 90\%$) and that the pattern is very similar for the three force fields. The slope of the fit is about −0.3, which can be interpreted to mean that a 1% change in molecular volume can be compensated by a −3% change in the charges; however, the scatter around the
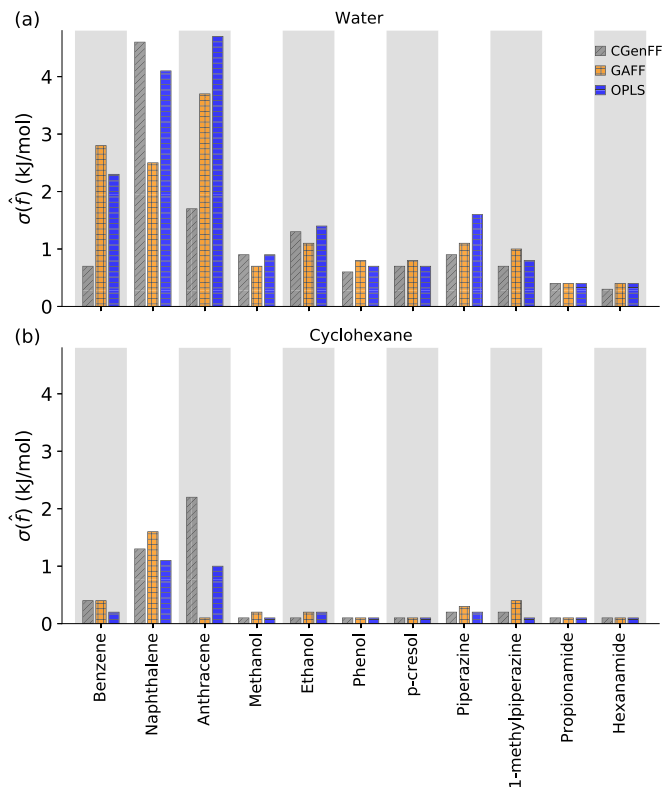

FIG. 2. Plot of $\sigma(\hat{f})$ corresponding to the SFE in (a) water and (b) cyclohexane. The diagonal, plus, and horizontal patterns in gray, orange, and blue bars refer to CGenFF, GAFF, and OPLS, respectively.
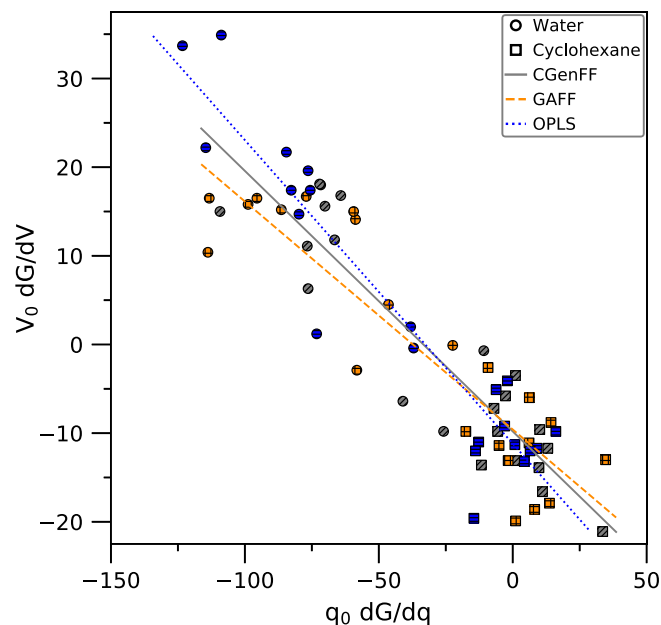

FIG. 3. Correlation between SFE derivatives (Tables II–IV). Circles denote the SFE in water and squares in cyclohexane. The diagonal, plus, and horizontal patterns in gray, orange, and blue symbols refer to CGenFF, GAFF, and OPLS data, respectively.

regression lines is significant. The sensitivity of the SFE with respect to the force field parameters is found to be high and it is indeed well established that small changes in charges can affect hydration free energies significantly [24].

Errors in the experimental liquid density are typically smaller and therefore the contribution to $\sigma(\hat{f})$ is smaller than that due to errors in experimental multipole moments (it should be noted that in many cases quantum chemistry is used to produce charges, adding an additional error source [40]). Nevertheless, the SFE is very sensitive to changes in the van der Waals radius σ in the Lennard-Jones 12-6 potential (Tables II–IV). Recent force fields employing softer potentials in combination with distributed charges [67,68] may be less sensitive to small changes in σ or the magnitude of the charges.

In the derivation of $\sigma(\hat{f})$ we have assumed that the force field reproduces the experimental data perfectly. However, it is well known that this assumption does not hold. Cailliez and Pernot have studied model uncertainty using Bayesian optimization [43]. They found that the uncertainty due to parametrization is larger than statistical uncertainty for both thermodynamic and transport properties. Ghahremanpour *et al.* have determined force field parameters based on the Alexandria library of quantum chemical properties [40]. They used bootstrapping to estimate uncertainties in, for instance, atomic polarizabilities and found these to vary between 1% and 4% [69]. These examples highlight that additional sources of error are present in force field calculations.

IV. CONCLUSION

We conclude that there are errors in SFE calculations, due to imperfect data, that are significant because of the high sensitivity of the SFE to changes in force field parameters (Tables II–IV). While a great deal of computational effort is typically used to produce well-converged free energy calculations with low error bars [5], the larger error sources discussed here are usually neglected. Finally, we note that the largest uncertainties are found for aromatic compounds, which is cumbersome since these are important in drug design and there is a lack of free energy data for such compounds [70].

[1] M. R. Shirts, J. W. Pitera, W. C. Swope, and V. S. Pande, J. Chem. Phys. **119**, 5740 (2003).

[2] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[3] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).

[4] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).

[5] A. Yildirim, T. Wassenaar, and D. van der Spoel, J. Chem. Phys. **149**, 144111 (2018).

[6] P. Procacci, J. Chem. Phys. **150**, 127101 (2019).

[7] A. Rizzi, T. Jensen, D. R. Slochower, M. Aldeghi, V. Gapsys, D. Ntekoumes, S. Bosisio, M. Papadourakis, N. M. Henriksen, B. L. de Groot, Z. Cournia, A. Dickson, J. Michel, M. K. Gilson, M. R. Shirts, D. L. Mobley, and J. D. Chodera, J. Comput. Aid. Mol. Des. **34**, 601 (2020).

[8] V. Gapsys, L. Pérez-Benito, M. Aldeghi, D. Seeliger, H. van Vlijmen, G. Tresadern, and B. L. de Groot, Chem. Sci. **11**, 1140 (2020).

[9] G. J. Rocklin, D. L. Mobley, and K. A. Dill, J. Chem. Theory Comput. **9**, 3072 (2013).

[10] A. V. Tran and Y. Wang, Comput. Mater. Sci. **127**, 141 (2017).

[11] A. Tran and Y. Wang, in *Uncertainty Quantification in Multiscale Materials Modeling*, edited by Y. Wang and D. L. McDowell, Elsevier Series in Mechanics of Advanced Materials (Woodhead Publishing, 2020), pp. 229–271.

[12] F. Cailliez, P. Pernot, F. Rizzi, R. Jones, O. Knio, G. Arampatzis, and P. Koumoutsakos, *Uncertainty Quantification in Multiscale Materials Modeling*, edited by Y. Wang and D. L. McDowell, Elsevier Series in Mechanics of Advanced Materials (Woodhead Publishing, 2020), pp. 169–227.

[13] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, J. Comput. Chem. **25**, 1157 (2004).

[14] W. L. Jorgensen and J. Tirado-Rives, Proc. Natl. Acad. Sci. USA **102**, 6665 (2005).

[15] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell Jr., J. Comput. Chem. **31**, 671 (2010).

[16] J. J. Mortensen, K. Kaasbjerg, S. L. Frederiksen, J. K. Nørskov, J. P. Sethna, and K. W. Jacobsen, Phys. Rev. Lett. **95**, 216401 (2005).

[17] J. Zhang, B. Tuguldur, and D. van der Spoel, J. Chem. Inf. Model. **55**, 1192 (2015).

[18] D. van der Spoel, M. M. Ghahremanpour, and J. Lemkul, J. Phys. Chem. A **122**, 8982 (2018).

[19] Z. Jing, C. Liu, S. Y. Cheng, R. Qi, B. D. Walker, J.-P. Piquemal, and P. Ren, Annu. Rev. Biophys. **48**, 371 (2019).

[20] C. F. Wong, J. Am. Chem. Soc. **113**, 3208 (1991).

[21] J. P. Jämbeck, F. Mocci, A. P. Lyubartsev, and A. Laaksonen, J. Comput. Chem. **34**, 187 (2013).

[22] D. Shivakumar, Y. Deng, and B. Roux, J. Chem. Theory Comput. **5**, 919 (2009).

[23] B. Hess and N. F. A. van der Vegt, J. Phys. Chem. B **110**, 17616 (2006).

[24] D. L. Mobley, E. Dumont, J. D. Chodera, and K. A. Dill, J. Phys. Chem. B **111**, 2242 (2007).

[25] D. L. Mobley, E. Dumont, J. D. Chodera, and K. A. Dill, J. Phys. Chem. B **115**, 1329 (2011).

[26] Y. H. Zhao, M. H. Abraham, and A. M. Zissimos, J. Org. Chem. **68**, 7368 (2003).

[27] M. R. Battaglia, A. D. Buckingham, and J. H. Williams, Chem. Phys. Lett. **78**, 421 (1981).

[28] A. M. Awwad and M. A. Abu-Daabes, J. Chem. Thermodyn. **40**, 645 (2008).

[29] R. L. Calvert and G. L. Ritchie, J. Chem. Soc., Faraday Trans. 2 **76**, 1249 (1980).

[30] J. Hales and R. Townsend, J. Chem. Thermodyn. **4**, 763 (1972).

[31] R. D. Johnson III, NIST standard reference database **69** (2013).

[32] W. M. Haynes, *CRC Handbook of Chemistry and Physics* (CRC press, Cleveland, Ohio, 2014).

[33] B. García, R. Alcalde, S. Aparicio, and J. M. Leal, Phys. Chem. Chem. Phys. **4**, 1170 (2002).

[34] M. Frenkel, X. Hong, R. C. Wilhoit, and K. R. Hall, *Landolt-Börnstein - Group IV Physical Chemistry, Volume 8 / Subvolume G / 2000, 3 Diols, 4 Triols* (Springer, Berlin, Heidelberg, 2000).

[35] Y. Marcus, *The Properties of Solvents* (Wiley, Chichester, UK, 1998).

[36] N. Allinger, J. Carpenter, and F. Karkowski, Tetrahedron Lett. **5**, 3345 (1964).

[37] SCI, *Piperazine. Summary risk assessment report.*, Tech. Rep. (Swedish Chemicals Inspectorate, Box 2. SE-17213 Sundby-berg, Sweden, 2005).

[38] C. L. Yaws, *Thermophysical Properties of Chemicals and Hydrocarbons* (William Andrew Inc., Beaumont, Texas, 2008).

[39] K. A. Kam, B. I. Tengan, C. K. Hayashi, R. C. Ordonez, and D. G. Garmire, Sensors **18**, 2774 (2018).

[40] M. M. Ghahremanpour, P. J. van Maaren, and D. van der Spoel, Sci. Data **5**, 180062 (2018).

[41] H. H. Ku, J. Res. Natl. Bur. Stand. **70C**, 263 (1966).

[42] R. W. Keener, *Theoretical Statistics. Topics for a Core Course.* (Springer, New York, Dordrecht, Heidelberg, London, 2010).

[43] F. Cailliez and P. Pernot, J. Chem. Phys. **134**, 054124 (2011).

[44] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).

[45] E. Lindahl, B. Hess, and D. Van Der Spoel, J. Mol. Model. **7**, 306 (2001).

[46] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, J. Comput. Chem. **26**, 1701 (2005).

[47] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008).

[48] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, Bioinformatics **29**, 845 (2013).

[49] G. Duarte Ramos Matos, D. Y. Kyu, H. H. Loeffler, J. D. Chodera, M. R. Shirts, and D. L. Mobley, J. Chem. Eng. Data **62**, 1559 (2017).

[50] D. L. Mobley and J. P. Guthrie, J. Comput. Aided Mol. Des. **28**, 711 (2014).

[51] https://cgenff.umaryland.edu.

[52] L. S. Dodda, J. Z. Vilseck, J. Tirado-Rives, and W. L. Jorgensen, J. Phys. Chem. B **121**, 3864 (2017).

[53] L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, and W. L. Jorgensen, Nucl. Acids Res. **45**, W331 (2017).

[54] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[55] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, J. Comput. Chem. **18**, 1463 (1997).

[56] S. Miyamoto and P. A. Kollman, J. Comput. Chem. **13**, 952 (1992).

[57] W. F. van Gunsteren and H. J. C. Berendsen, Mol. Simul. **1**, 173 (1988).

[58] N. Goga, A. Rzepiela, A. De Vries, S. Marrink, and H. Berendsen, J. Chem. Theory Comput. **8**, 3637 (2012).

[59] H. J. C. Berendsen, J. P. M. Postma, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).

[60] M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182 (1981).

[61] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).

[62] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, Chem. Phys. Lett. **222**, 529 (1994).

[63] P. V. Klimovich, M. R. Shirts, and D. L. Mobley, J. Comput. Aided Mol. Des. **29**, 397 (2015).

[64] M. R. Shirts and J. D. Chodera, J. Chem. Phys. **129**, 124105 (2008).

[65] K. A. Beauchamp, J. M. Behr, A. S. Rustenburg, C. I. Bayly, K. Kroenlein, and J. D. Chodera, J. Phys. Chem. B **119**, 12912 (2015).

[66] C. Caleman, J. S. Hub, P. J. van Maaren, and D. van der Spoel, Proc. Natl. Acad. Sci. U.S.A. **108**, 6838 (2011).

[67] P. T. Kiss and A. Baranyai, J. Chem. Phys. **141**, 114501 (2014).

[68] M. M. Walz, M. M. Ghahremanpour, P. J. van Maaren, and D. van der Spoel, J. Chem. Theory Comput. **14**, 5933 (2018).

[69] M. M. Ghahremanpour, P. J. van Maaren, C. Caleman, G. R. Hutchison, and D. van der Spoel, J. Chem. Theory Comput. **14**, 5553 (2018).

[70] H.-J. Schneider, New J. Chem. **43**, 15498 (2019).