Comparison of Machine-Learning and Classical Force Fields in Simulating the Solvation of Small Organic Molecules in Acetonitrile

Sangni Xun, Fang Liu[1]

Department of Chemistry, Emory University, Atlanta, Georgia, USA, 30322

Abstract: Machine learning force fields (MLFFs) have emerged as a new method for molecular simulation that combines the accuracy of ab initio approaches with the computational efficiency of classical force fields. However, the performance of MLFFs in describing the solvation configuration has yet to be explored. Here, we compare and contrast the performance of ANI-1ccx MLFF, the GAFF classical force field, and the ab initio molecular dynamics (AIMD) in simulating nine organic solutes in acetonitrile solvents. We examine the solvent-solute interaction described by these methods from four aspects: the solute conformation landscape, the solvation shell structure, the structure and dynamics of the O-H···N hydrogen bond, and the dynamics of the first solvation shell. For solute conformation description, ANI-1ccx and GAFF both yield minima that agree with density functional theory optimization for rigid solutes. However, their results diverge for flexible solutes. For solvation

[1] Electronic mail: fang.liu@emory.edu

shell structure description, ANI-1ccx agrees better with AIMD on the location of the first solvent shell than GAFF does. For the description of the O-H···N hydrogen bond formed between acetonitrile and the solute, ANI-1ccx generates stronger hydrogen bonds with shorter bond lengths, wider bond angles, and longer hydrogen bond lifetimes, agreeing better with DFT-optimized structure. ANI-1ccx also describes a more frequent exchange of acetonitrile molecules in and out of the first solvation shell than GAFF. Our study demonstrates the potential benefits of utilizing MLFF for simulating solution-phase dynamics and generating solvation configurations.

I. INTRODUCTION

The study of the interactions between solute and solvents is fundamental to understanding phenomena such as solvation, molecular recognition, and biological function.[1] For a solvated molecule, determining its solvation configuration is the prerequisite for quantum chemistry calculations to predict its properties and activities, such as redox potential,[2] pKa,[3] fluorescence,[4] and catalytic efficiency.[5] Various computational approaches have been developed to determine molecules' solvation configurations based on molecular dynamics (MD) simulations,[6, 7] or the direction generation of microsolvated conformations from global optimization or stochastic algorithms.[8-13] Among them, MD simulations are still widely used due to their flexibility in simulating solvation under different temperatures, pressures, and concentrations.[14, 15]

MD simulations of solvated molecules can be classified based on the theoretical methods for evaluating the forces (gradients of the system's potential energy function), including

classical force fields (FFs), ab initio methods, and machine learning force fields (MLFF). Classical FFs have the highest computational efficiency, but their accuracy strongly depends on the simulation parameters for the potential energy functions, which are typically generated from experimental data of nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, or quantum mechanical calculations. Tuning classical FFs' parameters can be highly time-consuming and system-dependent. In contrast, ab initio methods do not rely on predefined simulation parameters and can be directly applied to simulate various chemical systems with high accuracy. However, the high computational costs of ab initio methods limit the MD simulation's system size and time scale.

In recent years, machine learning force fields (MLFFs) have emerged as a new alternative method to evaluate forces in MD simulations.[16-22] Trained on massive quantum mechanical calculation data, MLFFs aim at bridging the accuracy of ab initio methods and the efficiency of classical force fields. Unlike the classical force fields that use parameterization to describe atomic interactions, MLFFs employ an artificial intelligence (AI) framework to learn the energy of atoms based on their specific chemical environment.[20] MLFFs enhance the adaptability across diverse molecules and improve transferability. In recent years, MLFFs have been widely used in the simulations of gas-phase molecules,[23-25] condensed-phase materials[26, 27] and liquid solutions.[28, 29] Specifically, TorchANI is one of the most widely used MLFF packages for molecular systems using the PyTorch framework.[30] TorchANI facilitates the training/inference of ANI (or ANAKIN-ME: Accurate NeurAl networK engINe for Molecular Energies) deep learning models for the computation of potential energy surfaces and various

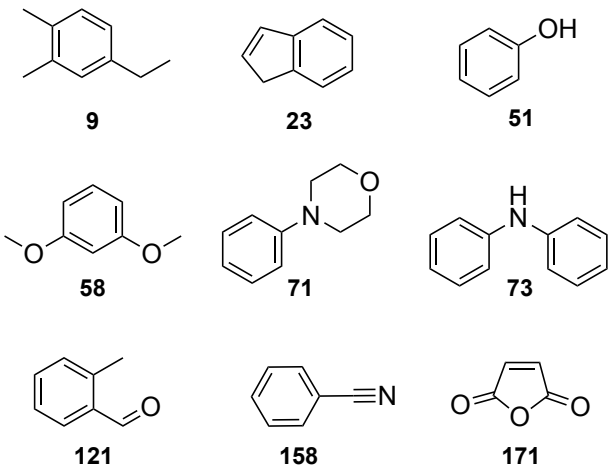
molecular physical properties.[30, 31] Currently, the model zoo of TorchANI has three ANI models including ANI-1x,[32] ANI-1ccx,[33] and ANI-2x.[34] Various studies have demonstrated the superior accuracy of ANI models compared to classical force fields, including the comparison of peptide conformations generated by ANI-2x vs. two classical FFs by Rosenberger *et al*.,[35] and the benchmark of torsional potential energy surface of biaryl drug fragments by. Lahey *et al*.[36] However, few studies[37] have investigated the performance of ANI models in simulating the solvation of small organic molecules in organic solvents.

In this work, we aim to understand the differences between classical FFs and MLFFs in simulating solvation by performing MD simulations of nine organic solutes in acetonitrile solvents with classical FFs and MLFFs. We focus on the acetonitrile-solvated systems because acetonitrile is among the most commonly used organic solvents in chemical studies due to its unique physical properties such as high dielectric constant and high dipole moment.[38] For our classical FF simulations, General Amber Force Field (GAFF)[39, 40] is chosen to describe the solutes because of its auto-parameterization[41] using semi-empirical quantum chemistry methods and the resulting broad applications in diverse organic compounds.[41-44] The six-site acetonitrile model by Gravuleda *et al*. is used for the solvent because of its good agreement with experimental data in density, heat of vaporization and isothermal compressibility.[45] Our MLFF simulations are all performed with the ANI-1ccx model consisting of an ensemble of 8 different networks, which were developed utilizing transfer learning techniques on the ANI-1ccx dataset.[46, 47] The aimed accuracy of ANI-1ccx corresponds to the CCSD(T)/CBS (coupled cluster considering single, double, and perturbative triple excitations utilizing a complete basis

set) level of theory. [33, 48, 49] The trained potentials of ANI-1ccx implicitly include dispersion interaction, which might give a better description of solute-solvent interactions.[35]

To compare the conformations sampled from GAFF and ANI-1ccx, we measured the closeness between the solute and solvent through minimum distance distribution function (MDDF) analysis. We also assessed the dynamic residence of the hydrogen bond formed between phenol and acetonitrile. Further, we examined the dynamic behaviors of solvents within the primary solvent shell surrounding the solutes, drawing from the trajectories of both ANI-1ccx and GAFF. The GAFF and ANI-1ccx solvation configurations were also compared with ab initio MD simulations or ab initio geometry optimization results. Our findings underscore the promise of MLFFs to generate solvation configurations with a better agreement with ab initio methods than the classical FFs. [37]

II. COMPUTATIONAL METHODS

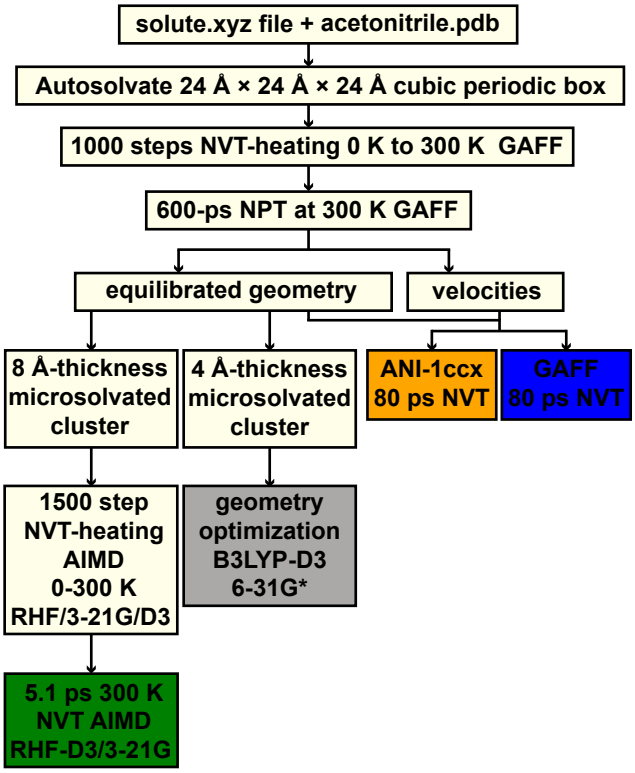


Scheme 1. Nine neutral solutes composed of C, H, N, O from ROP313 data. The bold numbers below the structures are the serial numbers in ROP313 data including **9**: 1,2-Dimethyl-4-ethylbenzene, **23**: Indene, **51**: Phenol, **58**: 1,3-Dimethoxybenzene, **71**: 4-Phenylmorpholine,

Classical MD simulation. Nine neutral solutes composed of C, H, N, O from ROP313 data[50] (see Scheme 1) were selected for MD simulations. All classical MD simulations were carried out using Amber GAFF force field[40] in AmberTool22.[39] As Scheme 2 shows, each solute was solvated in a 24 Å × 24 Å × 24 Å cubic periodic box of acetonitrile packed by Autosolvate.[43] The atomic charge was determined by AMI-BCC method.[51] The parameters for the acetonitrile solvent model were adopted from the six-site model detailed in reference.[45] The cutoff of non-bonded interactions was set to 8 Å. Bonds involving hydrogens were constrained by the SHAKE algorithm.[52] In the stage of energy minimization, a steepest descent algorithm was employed within 2000 steps. In the following heating stage, the temperature of the system was increased from 0 to 300 K, using Langevin thermostat with a collision frequency of 2.0 ps^{-1}. A total of 10,000 steps were performed with a time step of 2 fs under the NVT ensemble. After heating, the system was equilibrated under the NPT ensemble, running for 300000 steps with a time step of 2 fs at 300 K. The Berendsen method was used for pressure coupling at 1 atm. Starting from the last frame of the equilibrated NPT stage as the initial condition, an 80-ps production run was performed under the NVT ensemble at 300 K using Langevin dynamics with a collision frequency of 2.0 ps^{-1}. The coordinates and energy values were saved every 10 steps.

MLFF simulations All the initial coordinates and velocities were obtained from the last frames of equilibrated NPT classic simulations as described in the previous subsection. For each solute, an 80-ps NVT MD simulation was performed using the TorchANI.ANI-1ccx module[30] in the Atomic Simulation Environment (ASE).[53] No restraints were applied to bonds involving hydrogens. The time step was 0.2 fs. Temperature control was achieved using the Langevin thermostat, with a friction coefficient set at 0.02, as implemented in ASE.

Ab initio MD simulations Ab-into molecular dynamics simulations (AIMD) were carried out using the GPU-accelerated quantum chemistry package, TeraChem.[54] For each acetonitrile-



Scheme 2. The flowchart of MD and QM simulations in this work. The geometries generated in the colored blocks (green: AIMD, gray: DFT geometry optimization, orange: MD with ANI-1 ccx, and blue: MD with GAFF) are used for comparison.

solvated system, the microsolvated cluster, including the solute and the solvent molecules within 8 Å of the solute, was extracted from the final frame of the classical FF NPT simulation to be used as the initial structure for AIMD simulations. AIMD simulations were performed at Restricted Hartree-Fock (RHF)/3-21G level of theory with Grimme's D3 dispersion correction[55] and a time step of 0.3 fs. It is worth noting that despite being a small basis set, the 3-21 G basis set was shown to capture most of the important chemistry of organic molecules in AIMD in previous studies.[56-58] Our choice of RHF-D3/3-21G allows us to save computational resources while obtaining a modest description of these microsolvated clusters with more than 327 atoms. The AIMD simulations were initialized at 10 K, gradually increasing to a target temperature of 300 K through a Langevin thermostat within 1500 steps. The spherical boundary condition with a force constant of 10.0 kcal/(mol×Å^2) and target density of 0.78 g/mL matching the density of acetonitrile was set surrounding the cluster throughout the simulation to prevent the solvents from evaporating into the vacuum. After the heating process, a 5.1-ps NVT AIMD was conducted for each solute-acetonitrile system at 300 K through a Langevin thermostat.

Quantum mechanics (QM) geometry optimization To obtain a more accurate assessment of solute-solvent interactions using quantum mechanics, we performed geometry optimization of microsolvated clusters at the B3LYP[59]/6-31G* level of theory with D3 dispersion correction,[55] hereby denoted as B3LYP-D3, by using TeraChem. Our choice of B3LYP-D3/6-31G* is based on its good performance in predicting solution-phase properties in our previous explicit solvent calculations,[2] and that dispersion correction is needed for the solute-solvent

interaction. We did not use higher-level methods (e.g., CCSD(T)) because of the large size of the microsolvated clusters (more than 91 atoms) and the unavailability of analytical energy gradients for the low-scaling alternative DLPNO-CCSD(T).[60] From the final frame of the classical FF NPT simulations, the microsolvated cluster, including the solute and any solvent within a 4-Å-thick shell around the solute, was extracted for QM geometry optimization. The Graphics Processing Unit (GPU) accelerated[61, 62] conductor-like polarizable continuum model (CPCM)[63-66] with a dielectric constant of 35.7 was used to account for the electrostatic interactions between the microsolvated cluster and the solvents outside.

Principal Component Analysis of solutes We use principal component analysis (PCA) to investigate the high-dimensional movements of solutes in NVT simulations. Each production MD trajectory of the solute obtained from GAFF and ANI-1ccx was saved every 20 fs and resulted in 4000 frames. After removing the rotational and translational motions, the Cartesian coordinates of the solutes were expressed as 4000 × 3N matrix

$$\mathbf{A} = \begin{bmatrix} x_{1,1} & y_{1,1} & z_{1,1} & \cdots & x_{1,N} & y_{1,N} & z_{1,N} \\ x_{2,1} & y_{2,1} & z_{2,1} & \cdots & x_{2,N} & y_{2,N} & z_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{4000,1} & y_{4000,1} & z_{4000,1} & \cdots & x_{4000,N} & y_{4000,N} & z_{4000,N} \end{bmatrix} \quad (1)$$

$$\mathbf{A}_{\mathrm{ANI-1ccx}} = \begin{bmatrix} \mathbf{x}_1^{\mathrm{ANI}T} \\ \mathbf{x}_2^{\mathrm{ANI}T} \\ \vdots \\ \mathbf{x}_{4000}^{\mathrm{ANI}\ T} \end{bmatrix}, \ \mathbf{A}_{\mathrm{GAFF}} = \begin{bmatrix} \mathbf{x}_1^{\mathrm{GAFF}T} \\ \mathbf{x}_2^{\mathrm{GAFF}T} \\ \vdots \\ \mathbf{x}_{4000}^{\mathrm{GAFF}T} \end{bmatrix} \quad (2)$$

Here, N is the total number of atoms in each solute. The row vectors in $\mathbf{A}$ represent the 3-dimentional coordinates of atoms from the solutes at a specific frame during simulations. $\mathbf{A}_{\mathrm{GAFF}}$ and $\mathbf{A}_{\mathrm{ANI-1ccx}}$ denote the Cartesian coordinates matrices of NVT simulations from GAFF and ANI-1ccx, respectively. We use $\mathbf{x}_i^{\mathrm{GAFF}T}$ and $\mathbf{x}_i^{\mathrm{ANI}T}$ to represent the *i*-th row

vectors in $\mathbf{A}_{\mathrm{GAFF}}$ and $\mathbf{A}_{\mathrm{ANI-1ccx}}$, respectively. Then, we define the covariance matrix of $\mathbf{A}_{\mathrm{GAFF}}$, $\mathbf{C}_{\mathrm{GAFF}} \in \mathbb{R}^{3N\times 3\mathrm{N}}$, as the following.

$$\mathbf{C}_{\mathrm{GAFF}} = \mathrm{cov}(\mathbf{A}_{\mathrm{GAFF}}) \tag{3}$$

By performing eigen decomposition of $\mathbf{C}_{\mathrm{GAFF}}$, we obtain

$$\begin{gathered}\mathbf{C}_{GAFF}\,\mathbf{Y} = \mathbf{Y}\boldsymbol{\Lambda} \\ \mathbf{Y} = \left(\mathbf{y}_1^{\mathrm{GAFF}}, \mathbf{y}_1^{\mathrm{GAFF}}, \cdots, \mathbf{y}_1^{\mathrm{GAFF}}\right) \\ \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1^{\mathrm{GAFF}}, \lambda_2^{\mathrm{GAFF}}, \ldots, \lambda_{3\mathrm{N}}^{\mathrm{GAFF}})\end{gathered} \tag{4}$$

Here, $\{\lambda_i^{\mathrm{GAFF}}\}$ are the eigen values ordered by the magnitude from large to small, and the column vectors $\{\mathbf{y}_1^{\mathrm{GAFF}}\}$ are the corresponding eigenvectors. We then project the cartesian coordinates of the solute in the *i*-th frame of the GAFF simulation to $\vec{y}_1^{\mathrm{GAFF}}$ and $\vec{y}_2^{\mathrm{GAFF}}$ to represent the solute conformation in the PCA space:

$$\mathrm{PCA1}_i^{\mathrm{GAFF}} = \mathbf{x}_i^{\mathrm{GAFF}^T}\mathbf{y}_1^{\mathrm{GAFF}}, \;\; \mathrm{PCA2}_i^{\mathrm{GAFF}} = \mathbf{x}_i^{\mathrm{GAFF}^T}\mathbf{y}_2^{\mathrm{GAFF}} \tag{5}$$

To get a side-to-side comparison between the GAFF and ANI-1ccx conformation, we also project the ANI-1ccx conformations to the same set of PCA vectors:

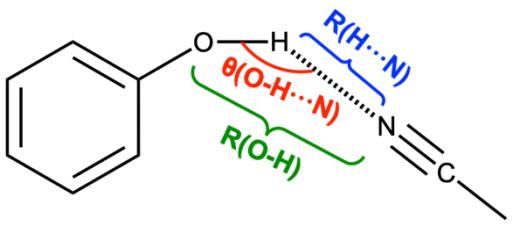
$$\mathrm{PCA1}_i^{\mathrm{ANI-1ccx}} = \mathbf{x}_i^{\mathrm{ANI-1ccx}^T}\mathbf{y}_1^{\mathrm{GAFF}}, \;\; \mathrm{PCA2}_i^{\mathrm{ANI-1ccx}} = \mathbf{x}_i^{\mathrm{ANI-1ccx}^T}\mathbf{y}_2^{\mathrm{GAFF}} \tag{6}$$

Analysis of Hydrogen Bonds Given that there are few references specifying the criteria for intermolecular hydrogen bonding between phenol and acetonitrile in an explicit solvent environment, we have adopted the criteria for intermolecular hydrogen bonds (O-H···O) in water-water interatcions[67] to determine the existence of hydrogen bonds between phenol and acetonitrile solvent in molecular dynamics simulations:

$$R(\mathrm{H}\cdots\mathrm{N}) \leq 2.45\ \text{Å},$$

$$R(\mathrm{N}-\mathrm{O}) \leq 3.50\ \text{Å},$$

$$\theta(\mathrm{O}-\mathrm{H}\cdots\mathrm{N}) \geq 150°.$$



Scheme 3. Schematic diagram of a hydrogen bond between the phenol and an acetonitrile

Here, R is the distance of the atom-atom pair between the acetonitrile solvent and the phenol (**Scheme 3)**. N denotes the nitrogen in the acetonitrile. H and O denote the hydrogen and oxygen atom in the hydroxyl group of phenol. The angle $\theta(\mathrm{O}-\mathrm{H}\cdots\mathrm{N})$ represents the oxygen-hydrogen-nitrogen angle between the hydrogen donor and acceptor.

The autocorrelation function has been widely used to describe the dynamic behaviors in MD simulations including hydrogen-bonding interaction, solvent-solute interactions, etc.[6, 68] The integration of the autocorrelation function over time yields the lifetime of the interactions. To characterize the dynamic hydrogen bonds between the solvent (acetonitrile) and solute (phenol) in system 51, we evaluate their relevant lifetime (τ_{HB}) in both ANI-1ccx's and GAFF's production trajectories. Here, τ_{HB} of the i-th acetonitrile-phenol pair engaged in hydrogen bonding is defined as[68]

$$\tau_i^{\mathrm{HB}} = \int_0^{\infty} C_i^{\mathrm{HB}}(t)dt\,, \tag{7}$$

where the $C_i^{\mathrm{HB}}(t)$ is the autocorrelation function for the i-th hydrogen bond, defined as

$$C_i^{\mathrm{HB}}(t) = \frac{h_i^{\mathrm{HB}}(t_0)h_i^{\mathrm{HB}}(t_0+t)}{h_i^{\mathrm{HB}}(t_0)^2}. \tag{8}$$

Here, $h_i^{\mathrm{HB}}(t)$ indicates whether a hydrogen bond between the phenol solute and the *i*-th solvent is present at time *t*. If the hydrogen bond is present, $h_{HBi}(t) = 1$; otherwise $h_{HBi} =$

0. t_0 denotes the first time the hydrogen bond appears. The average lifetime, $\overline{\tau^{\mathrm{HB}}}$, is defined as

$$\overline{\tau^{\mathrm{HB}}} = \frac{\sum_{i=1}^{n^{\mathrm{HB}}} \tau_i^{\mathrm{HB}}}{n^{\mathrm{HB}}} \tag{9}$$

Where n^{HB} is the total number of acetonitrile-phenol pairs engaged in hydrogen bonding.

Analysis of the First-Solvent Shell To estimate the lifetime of solvents in the first-solvent shell from GAFF and ANI-1ccx NVT MD simulations, we define the autocorrelation function as

$$C(t) = \left\langle \frac{\sum_{i=1}^{n} h_i(t_0) h_i(t_0 + t)}{\sum_{i=1}^{n} h_i(t_0)^2} \right\rangle. \tag{10}$$

Here, *n* represents the total number of solvents within the periodic boundary condition (PBC) box. The function $h_i(t)$ is set to 1 if the solvent is present in the first-solvent shell at time *t* and 0 otherwise. We use the minimum distance d_i between the solvent and solute as the criteria to define the presence in the first-solvent shell if $d_i \leq$ cutoff, where the cutoff is estimated from the minimum distance distribution functions (MDDF). The brackets in Eq (9) indicate taking the average over 40-ps windows throughout an 80-ps NVT simulation. The lifetime of the solvent in the 1st shell, τ, is defined as

$$\tau = \int_0^{40\ \mathrm{ps}} C(t)dt\,. \tag{11}$$

III. RESULTS AND DISCUSSIONS

1. Principal Component Analysis of solutes' motions

Different force fields yield varied potential energy surfaces, leading to divergent motion patterns for the solute during simulations. The dynamics of solutes in MD simulations occur in a high-dimensional space. To assess the differences in these motions and the conformations

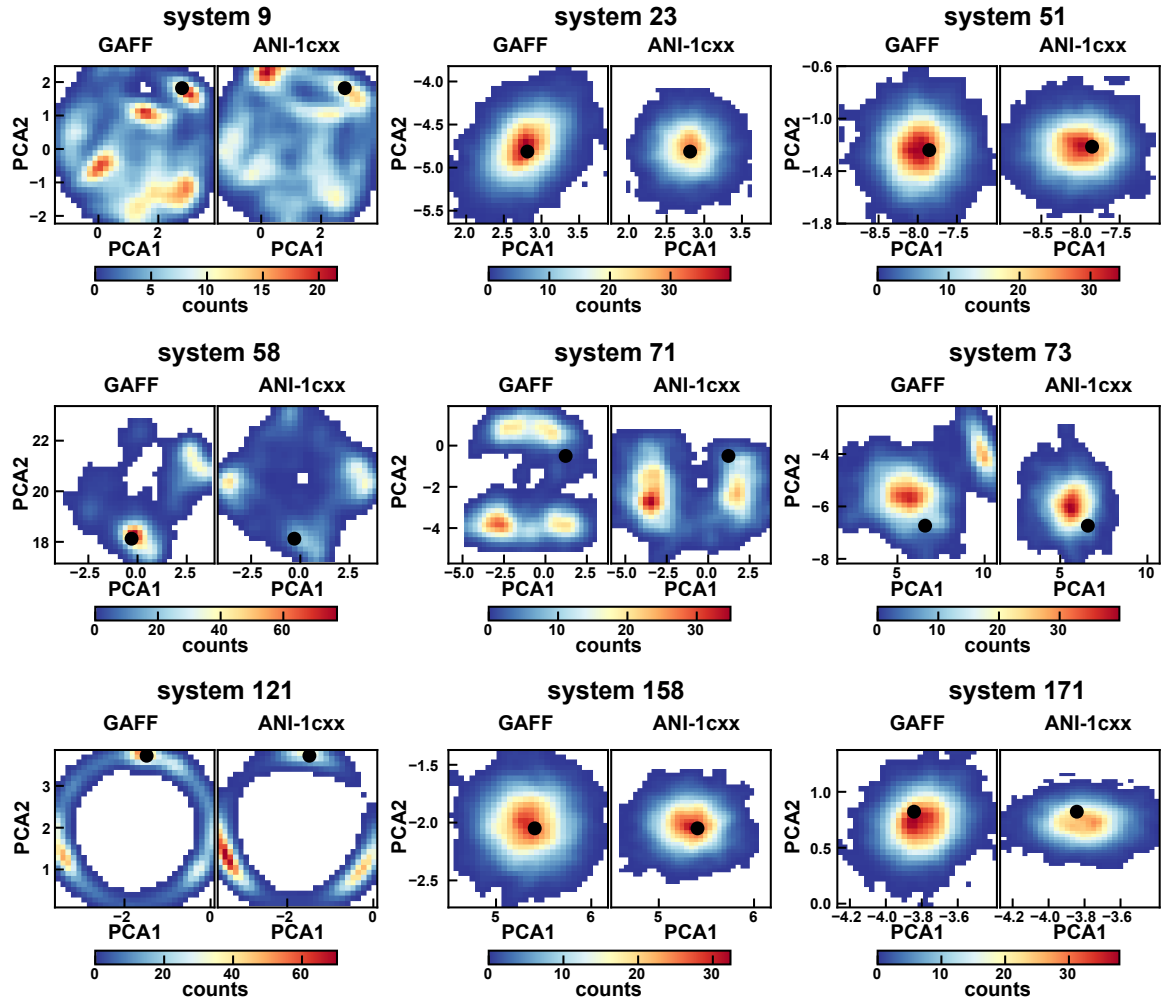


Figure 1. The plots illustrate the PCA projections along the PCA1 and PCA2 directions for nine solutes based on 80 ps of NVT simulations. For every system presented, the left panel shows results from GAFF, while the right panel reflects those from ANI-1ccx. Solid black circles in the plots represent projections of the DFT-optimized solute structures at B3YLP-D3/6-31G level within a 4 Å solvent shell.

observed in the NVT simulations, we employ principal component analysis (PCA)—a widely used tool for dimensionality reduction to depict the solute motions in both GAFF and ANI-1ccx NVT simulations.

As illustrated in Figure 1, systems with rigid structures, such as system 23, system 51, system 158, and system 171, display a single minimum in their PCA plots. These minima from the MD simulations also align well with the DFT-optimized outcomes. Conversely, systems with more flexible functional groups—like system 9, system 58, system 71, system 73, and system 171 exhibit multiple local minima in their PCA plots. For systems 9, 58, and 121, the minima produced by GAFF are in closer proximity to the DFT-optimized results. Meanwhile, for systems 71 and 73, the minima from both GAFF and ANI-1ccx diverge from the DFT-optimized findings. Therefore, the motions of solutes under GAFF and ANI-1ccx can differ depending on the system. Within limited conventional-simulation time, GAFF and ANI-1ccx display different preference of the conformations in systems with flexible solutes. When benchmarked against QM geometry optimizations using B3LYP, GAFF tends to show a marginally better agreement compared to ANI-1ccx. The observed outcome might be attributed to the initial geometry used in the QM optimization, which was derived from the final equilibrated structure of GAFF-NPT simulations.

2. Solvation Shell Structure Comparison

Obtaining reliable solvation shell structures is a major goal of solution-phase MD simulations and the starting point for subsequent quantum chemistry studies. Traditionally, solvation shell structures are often characterized by radial distribution functions (RDFs). However, RDF is not ideal for depicting the detailed distribution of solvents around non-spherical solutes. With solutes in various sizes and shapes, the minimum-distance distribution function (MDDF) measures the shortest distances between the solute and solvents, and the

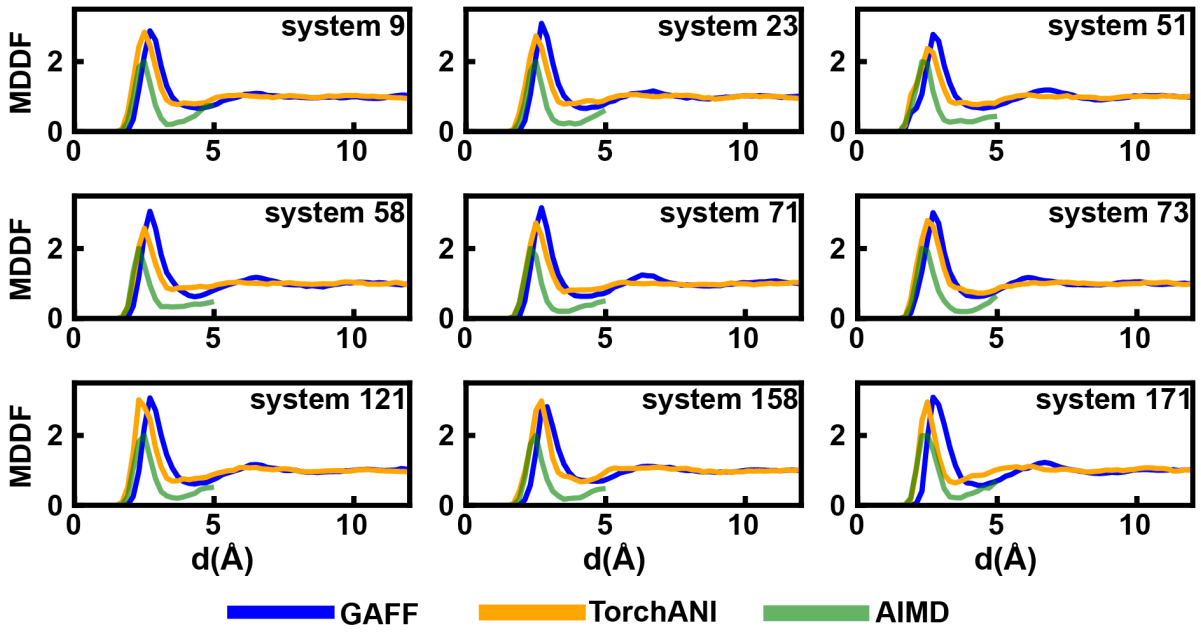


Figure 2. MDDFs of ANI-1ccx GAFF (blue line), MLFF (orange line) and AIMD (green line) NVT simulations at 300 K. The up-right number of each panel is the serial number from scheme 1.

results can be more easily compared across different systems[69, 70] including protein[71] and chemical compounds.[72] Here, to compare the solvation shell structures generated by ANI-1ccx and GAFF simulations, we exploited the ComplexMixtures[70] software to generate the solute-acetonitrile MDDFs from the 80 ps of NVT simulation trajectories and compared them against the reference AIMD trajectories (5.1 ps long).

As shown in Figure 2, the location of the first-solvent shell, corresponding to the peak of the MDDF curve, varies distinctly among the different simulation methods. The MDDF curves from the ANI-1ccx simulations shift slightly to the left compared to those from the GAFF simulations, indicating closer proximity between the solvent and the solute in the ANI-1ccx simulations. Furthermore, the ANI-1ccx MDDF curves are more aligned with the MDDF

curves derived from AIMD simulations. To quantify the differences in the MDDFs, we estimate the first solvation shell's cutoff distance as the MDDF curve's first local minimum after the first peak. For each solvated system, different simulation methods yield distinct cutoffs for the first solvation shell (Table 1). The ANI-1ccx and AIMD estimated cutoff distances agree reasonably well, with a difference of less than 0.6 Å for each of the nine systems studied. In contrast, the cutoff distances estimated by GAFF simulations are systematically larger than the ANI-1ccx results by 0.6-1.0 Å, exceeding 4 Å for all nine systems.

Table 1. The estimated cutoffs of the first-solvent shells according to the MDDF curves

system	GAFF(Å)	ANI-1ccx(Å)	AIMD(Å)
9	4.5	3.7	3.3
23	4.3	3.5	3.5
51	4.5	3.5	3.3
58	4.3	3.5	3.1
71	4.3	3.5	3.3
73	4.3	3.6	3.9
121	4.3	3.5	3.7
158	4.7	4.1	3.5
171	4.5	3.5	3.7

3. Hydrogen Bond Dynamics Between Phenol and Acetonitrile in system 51

Hydrogen bonds between solvents and solutes have been shown to play crucial roles in various chemical processes, from causing solvatochromism[73-76] to tuning enzyme activity[77, 78] and selectivity.[79] Since implicit-solvent models cannot describe hydrogen bonds,

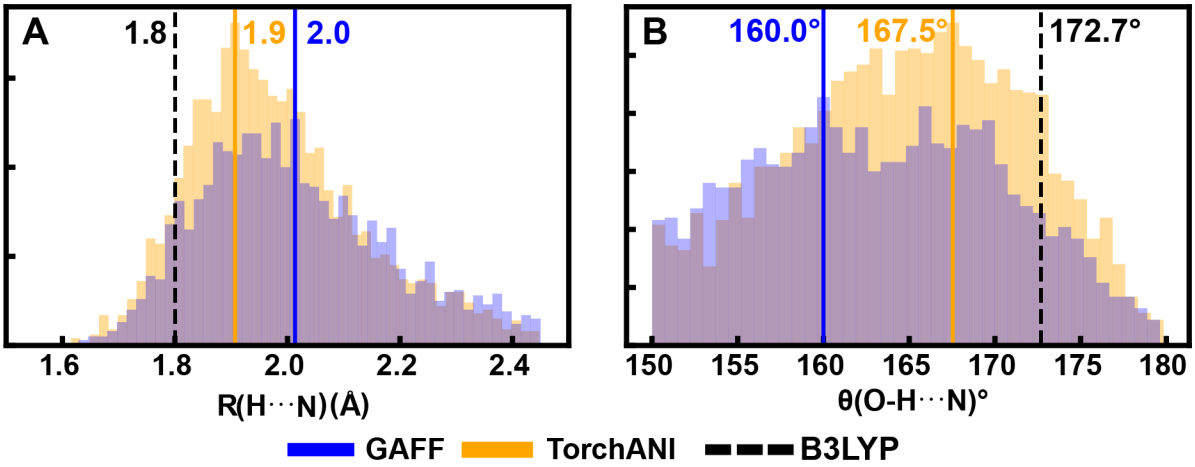


Figure 3. The histograms of A) R(H···N) and B) θ(O-H···N) in hydrogen bonds formed within 80-ps NVT simulations using ANI-1ccx (orange columns) and GAFF (blue columns) The orange and blue solid lines indicate the values with the highest probability in ANI-1ccx and GAFF simulations respectively. The black dashed lines denote the R(H···N) and θ(O-H···N) in QM calculation at B3LYP-D3 level.

computational studies of solutions involving hydrogen bonds often require explicit-solvent simulations, whose accuracy is reflected in the hydrogen-bonding structure and dynamics. As a polar aprotic solvent, acetonitrile can form O-H···N-type hydrogen bonds with many solutes. However, few studies have examined the accuracy of MLFF in describing the hydrogen-bonding interaction between acetonitrile and the solute. We investigate the geometric parameters and dynamics of the hydrogen bond formed between the phenol (solute) and acetonitrile (solvent) in system 51 to evaluate how GAFF and ANI-1ccx describe hydrogen bonding interactions. As shown in Scheme 3, the proton donor is the hydroxyl group of the phenol, and the acceptor is the nitrogen atom in acetonitrile.

The hydrogen bond lengths and angles between the phenol and solvents exhibit different distributions in the ANI-1ccx and GAFF force fields (Figure 3). The hydrogen bonds simulated by ANI-1ccx have shorter bond lengths and larger angles than those by GAFF. The most probable hydrogen bond length simulated by ANI-1ccx (the peak of the orange line) is 1.9 Å, shorter than that of GAFF (the peak of the blue line) by 0.1 Å (Figure 3A). Additionally, the most probable hydrogen bond angle simulated by ANI-1ccx (the peak of the orange line) is larger than that of GAFF by 7.5° (Figure 3B). The hydrogen bond length optimized by B3LYP-D3/6-31G* is 1.8 Å, shorter than those from both ANI-1ccx and GAFF MD simulations. The hydrogen bond angle obtained from B3LYP-D3 geometry optimization is also the largest (172.7°). Overall, the hydrogen bonds between the phenol solute and acetonitrile solvents predicted by B3LYP-D3 geometry optimization are stronger than those predicted by ANI-1ccx and GAFF MD simulations. Compared to the GAFF simulation results, the hydrogen bond geometric parameters simulated by ANI-1ccx are closer to the QM results.

The hydrogen bond dynamics of phenol-acetonitrile pairs obtained by GAFF and ANI-1ccx are summarized in Figure 4 and Table 2. In the GAFF simulations, 11 pairs of phenol-acetonitrile participate in hydrogen bonding, whereas the ANI-1ccx simulations only feature 3 such pairs. As depicted in Figure 4, hydrogen bonds in the GAFF simulations both form and break more frequently than in the ANI-1ccx simulations. Additionally, the average lifetime, $\overline{\tau^{\mathrm{HB}}}$, of hydrogen bonds in GAFF is notably shorter than in ANI-1ccx. When the GAFF force field is employed, acetonitrile molecules that bonded with phenol display a higher dissociation rate from the hydroxyl groups of the solute. Furthermore, these molecules also show an increased tendency to interchange with other acetonitrile molecules in the solvents.

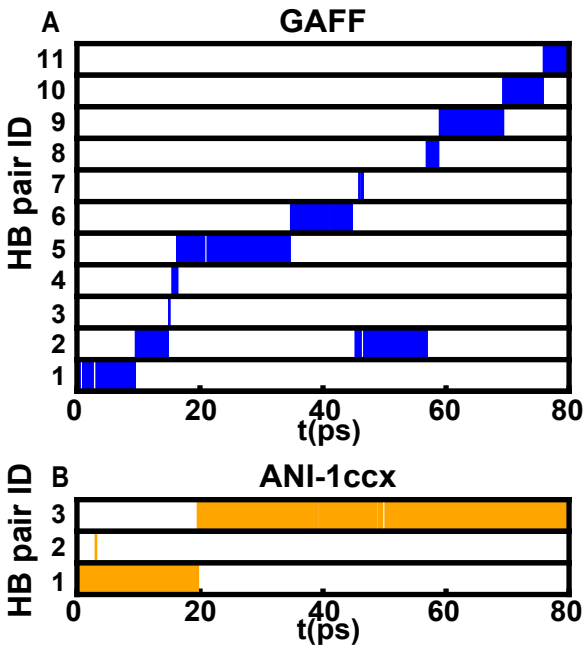


Figure 4. The hydrogen bonds formed by different acetonitrile-phenol pairs in system 51 from 80-ps NVT (A)GAFF and (B)ANI-1ccx simulations. The HB pair ID on the left indicate the phenol-acetonitrile pairs involved in hydrogen bonding, listed in order of their appearance from early to late in the simulations.

Table 2. The lifetime $\overline{\tau^{\mathrm{HB}}}$ (Eq. (9)) of each hydrogen bond formed by different acetonitrile-phenol pairs in system 51 from 80-ps NVT GAFF (left) and ANI-1ccx simulations.

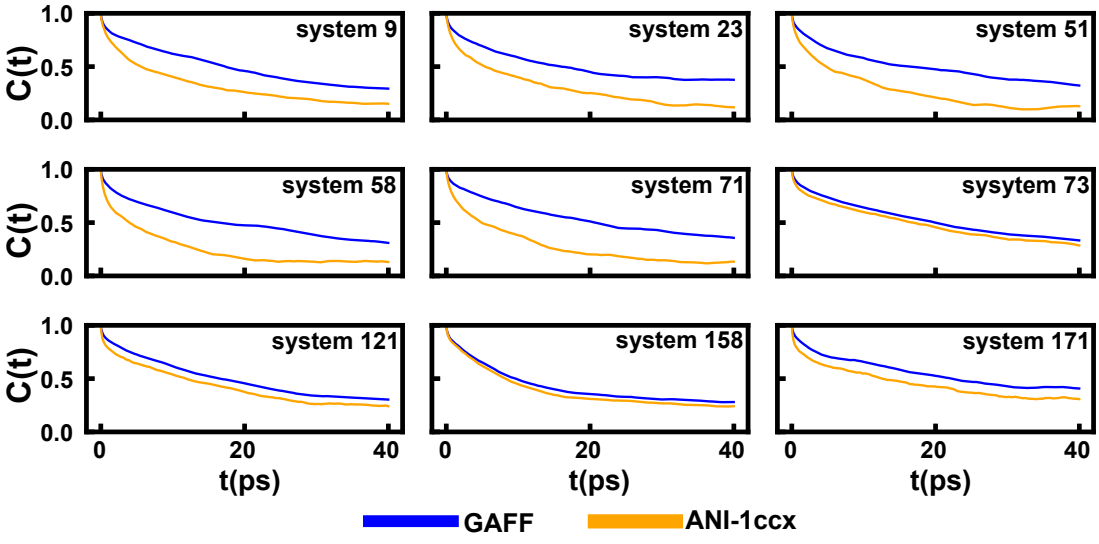
GAFF		ANI-1ccx	
[a]HB pair ID	τ_{HB} (ps)	[a]HB pair ID	τ_{HB} (ps)
1	5.13	1	16.71
2	11.73	2	0.01
3	0.01	3	50.64
4	0.31	$\overline{\tau^{\mathrm{HB}}}$	**22.45**
5	12.95		
6	6.57		
7	0.05		
8	1.45		
9	8.39		
10	4.51		
11	3.02		
$\overline{\tau^{\mathrm{HB}}}$	**4.92**		

[a] The solvent molecules that participated in hydrogen bonding with the solute were identified and numbered sequentially based on the time of their engagement during the simulation, from the earliest to the latest occurrence

In summary, when simulated by ANI-1ccx, hydrogen bonds are more prone to forming at shorter distances and larger angles, resulting in more confined and robust hydrogen bonds. Our observation of stronger O-H···N hydrogen bonds in ANI-1ccx compared to GAFF is similar to Babin *et al.*'s results where the O-H···O hydrogen bonds of water solvents in torchani.ANI2x were compared with that obtained from SPC[80] and TIP3P[81] water models.[35]

4. The solvents' dynamics in the first solvent shell

Finally, we analyzed the retention of acetonitrile solvents in the first solvation shell for all nine solvated systems in this study. In Figure 5, the autocorrelation functions of solvents within the first solvation shell simulated by GAFF consistently lie above those by ANI-1ccx. This suggests that solvents remain within the first-solvent shell for a longer duration in GAFF simulations compared to ANI-1ccx. Furthermore, as shown in Table 3, when integrated over a 40-ps simulation duration, the lifetime τ in GAFF exceeds that of ANI-1ccx by 1.6 to 10.5 ps, depending on the specific system under consideration. The assessment of the solvents' lifetime within the first-solvent shell suggests a more frequent solvent exchange in ANI-1ccx. This could imply that the potential well, which traps the solvents in GAFF's first-solvent shell, might be deeper compared to that of ANI-1ccx. However, this conclusion does not contradict our observation of stronger hydrogen bonds in ANI-1ccx. For the nine solvation systems



.

Figure 5. Autocorrelation functions for solvents within the first-solvent shell surrounding the solute during 40-ps NVT simulations as a function of simulation time. Comparisons are made between GAFF (blue line) and ANI-1ccx (orange line) simulations.

investigated in this study, the first solvation shell typically contains 16 to 22 solvent molecules. For system 51, at most one acetonitrile-phenol hydrogen bond presents at a specific time point during the simulation (see Figure 4). Although the hydrogen-bonded acetonitrile interacts more strongly with the solute in ANI-1ccx than in GAFF, the other acetonitrile in the first solvation may be less trapped in ANI-1ccx. This can result in the overall more frequent solvent exchange in ANI-ccx.

Table 3. The lifetime τ in Eq. (11) of solvents in the first-solvent shell integrated over 40 ps.

system	GAFF (ps)	ANI-1ccx (ps)
9	19.8	12.7
23	20.1	12.0
51	20.0	10.9
58	20.1	10.0
71	21.7	11.2
73	21.2	19.5
121	19.8	16.8
158	16.9	15.4
171	22.2	18.1

IV. CONCLUSION

In this research, we compared the ANI-1ccx machine learning force field and the GAFF classical force field in simulating the solvation of organic molecules in acetonitrile from various aspects. To ensure fair comparisons, we performed production MD simulations with ANI-1ccx and GAFF from the same starting structure and velocities equilibrated by GAFF. We constructed PCA landscapes for nine organic solutes from their ANI-1ccx and GAFF simulation trajectories in acetonitrile solvents. For solutes with rigid structures, both ANI-1ccx and GAFF display similar minima, consistent with the DFT-optimized results. For solutes with more flexibility, the dynamics of solutes can vary significantly between GAFF and ANI-1ccx.

We then use MDDF to evaluate the closeness between the solute and acetonitrile solvents. The solvents are closer to the solute in ANI-1ccx simulations than in GAFF simulations. The ANI-1ccx's MDDF also agrees better with AIMD's MDDF than GAFF's. Additionally, we compared ANI-1ccx and GAFF in describing the O-H···N hydrogen bond in system 51 between the phenol solute and the acetonitrile solvents. Hydrogen bonds in ANI-1ccx tend to have shorter bond lengths, more expanded bond angles, and longer lifetimes, leading to more localized and stronger hydrogen bonds. In contrast, acetonitrile solvents bonded to phenol demonstrate a faster rate of dissociation from the solute's hydroxyl groups in GAFF. Finally, we investigated the solvent dynamics in the first solvation shell for all nine systems. The autocorrelation functions suggest a more rapid solvent exchange in ANI-1ccx than GAFF. This might indicate a deeper potential well for non-hydrogen-bonded solvents in GAFF than in ANI-1ccx, responsible for retaining solvents in its primary solvent shell.

In summary, we compared ANI-1ccx MLFF and GAFF classical FF in the simulations of solute-acetonitrile interactions from different aspects. Overall, ANI-1 ccx has more agreements with QM level descriptions (geometry optimization or AIMD) on the solvation shell structure and hydrogen bonds. However, GAFF has marginally better agreement with QM geometry optimization on the solute minimum-energy structure. These potential benefits of using MLFF to simulate solvated systems motivate their applications in automated solvation configuration generation workflows,[43] which may further improve the accuracy of computationally generated chemical property data sets with the help of machine-learning-based corrections.[2, 82]

V. ACKNOWLEDGEMENT

Financial support for this publication results from Scialog grant #28755 from Research Corporation for Science Advancement. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231.

VI. SUPPORTING INFORMATION AVAILABLE

See the supplementary material for the initial solute coordinates; production trajectories including GAFF-NVT, ANI-1ccx-NVT and AIMD simulations; geometry optimized pdb coordinates for all 4 Å-thickness-microsolvated clusters (ZIP). The SI (ZIP) is available at https://figshare.com/s/64333aa869a47dd65b24

CITATIONS

1. Mabesoone, M. F. J.; Palmans, A. R. A.; Meijer, E. W., Solute–Solvent Interactions in Modern Physical Organic Chemistry: Supramolecular Polymers as a Muse. *Journal of the American Chemical Society* **2020**, 142, 19781-19798.

2. Hruska, E.; Gale, A.; Liu, F., Bridging the Experiment-Calculation Divide: Machine Learning Corrections to Redox Potential Calculations in Implicit and Explicit Solvent Models. *Journal of Chemical Theory and Computation* **2022**, 18, 1096-1108.

3. Busch, M.; Ahlberg, E.; Ahlberg, E.; Laasonen, K., How to Predict the pKa of Any Compound in Any Solvent. *ACS Omega* **2022**, 7, 17369-17383.

4. Chapman, C. F.; Maroncelli, M., Fluorescence studies of solvation and solvation dynamics in ionic solutions. *The Journal of Physical Chemistry* **1991**, 95, 9095-9114.

5. Sun, Q.; Wang, S.; Aguila, B.; Meng, X.; Ma, S.; Xiao, F.-S., Creating solvation environments in heterogeneous catalysts for efficient biomass conversion. *Nature Communications* **2018**, 9, 3236.

6. Rawat, N.; Biswas, P., Hydrogen Bond Dynamics in Intrinsically Disordered Proteins. *The Journal of Physical Chemistry B* **2014**, 118, 3018-3025.

7. Kurnikova, M. G.; Balabai, N.; Waldeck, D. H.; Coalson, R. D., Rotational Relaxation in Polar Solvents. Molecular Dynamics Study of Solute−Solvent Interaction. *Journal of the American Chemical Society* **1998**, 120, 6121-6130.

8. Simm, G. N.; Türtscher, P. L.; Reiher, M., Systematic microsolvation approach with a cluster-continuum scheme and conformational sampling. *Journal of Computational Chemistry* **2020**, 41, 1144-1155.

9. Zhang, J.; Dolg, M., ABCluster: the artificial bee colony algorithm for cluster global optimization. *Physical Chemistry Chemical Physics* **2015**, 17, 24173-24181.
10. Zhang, J.; Dolg, M., Global optimization of clusters of rigid molecules using the artificial bee colony algorithm. *Physical Chemistry Chemical Physics* **2016**, 18, 3003-3010.
11. Spicher, S.; Plett, C.; Pracht, P.; Hansen, A.; Grimme, S., Automated Molecular Cluster Growing for Explicit Solvation by Efficient Force Field and Tight Binding Methods. *Journal of Chemical Theory and Computation* **2022**, 18, 3174-3189.
12. Pracht, P.; Bohle, F.; Grimme, S., Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Physical Chemistry Chemical Physics* **2020**, 22, 7169-7192.
13. Talmazan, R. A.; Podewitz, M., PyConSolv: A Python Package for Conformer Generation of (Metal-Containing) Systems in Explicit Solvent. *Journal of Chemical Information and Modeling* **2023**, 63, 5400-5407.
14. Ghosh, T.; García, A. E.; Garde, S., Molecular Dynamics Simulations of Pressure Effects on Hydrophobic Interactions. *Journal of the American Chemical Society* **2001**, 123, 10997-11003.
15. Kitchen, D. B.; Reed, L. H.; Levy, R. M., Molecular dynamics simulation of solvated protein at high pressure. *Biochemistry* **1992**, 31, 10083-10093.
16. Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R., Machine Learning Force Fields. *Chemical Reviews* **2021**, 121, 10142-10186.
17. Deringer, V. L.; Caro, M. A.; Csányi, G., Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. *Advanced Materials* **2019**, 31, 1902765.
18. Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R., Machine Learning Force Fields: Construction, Validation, and Outlook. *The Journal of Physical Chemistry B* **2017**, 121, 511-522.
19. Behler, J., First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Journal of the American Chemical Society* **2017**, 56, 12828-12840.
20. Hagg, A.; Kirschner, K. N., Open-Source Machine Learning in Computational Chemistry. *Journal of Chemical Information and Modeling* **2023**, 63, 4505-4532.
21. Gokcan, H.; Isayev, O., Learning molecular potentials with neural networks. *WIREs Computational Molecular Science* **2022**, 12, e1564.
22. Magdău, I.-B.; Arismendi-Arrieta, D. J.; Smith, H. E.; Grey, C. P.; Hermansson, K.; Csányi, G., Machine learning force fields for molecular liquids: Ethylene Carbonate/Ethyl Methyl Carbonate binary solvent. *npj Computational Materials* **2023**, 9, 146.
23. Chmiela, S.; Sauceda, H. E.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A., sGDML: Constructing accurate and data efficient molecular force fields using machine learning. *Computer Physics Communications* **2019**, 240, 38-45.
24. Sauceda, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A., Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *The Journal of Chemical Physics* **2019**, 150, 114102.
25. Sauceda, H. E.; Gastegger, M.; Chmiela, S.; Müller, K.-R.; Tkatchenko, A., Molecular force fields with gradient-domain machine learning (GDML): Comparison and synergies with classical force fields. *The Journal of Chemical Physics* **2020**, 153, 124109.
26. Deringer, V. L.; Csányi, G., Machine learning based interatomic potential for amorphous carbon. *Physical Review B* **2017**, 95, 094203.

27. Deringer, V. L.; Bernstein, N.; Bartók, A. P.; Cliffe, M. J.; Kerber, R. N.; Marbella, L. E.; Grey, C. P.; Elliott, S. R.; Csányi, G., Realistic Atomistic Structure of Amorphous Silicon from Machine-Learning-Driven Molecular Dynamics. *The Journal of Physical Chemistry Letters* **2018**, 9, 2879-2885.
28. Hellström, M.; Ceriotti, M.; Behler, J., Nuclear Quantum Effects in Sodium Hydroxide Solutions from Neural Network Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **2018**, 122, 10158-10171.
29. Dajnowicz, S.; Agarwal, G.; Stevenson, J. M.; Jacobson, L. D.; Ramezanghorbani, F.; Leswing, K.; Friesner, R. A.; Halls, M. D.; Abel, R., High-Dimensional Neural Network Potential for Liquid Electrolyte Simulations. *The Journal of Physical Chemistry B* **2022**, 126, 6271-6280.
30. Gao, X.; Ramezanghorbani, F.; Isayev, O.; Smith, J. S.; Roitberg, A. E., TorchANI: A Free and Open Source PyTorch-Based Deep Learning Implementation of the ANI Neural Network Potentials. *Journal of Chemical Information and Modeling* **2020**, 60, 3408-3415.
31. Kjeldal, F. Ø.; Eriksen, J. J., Decomposing Chemical Space: Applications to the Machine Learning of Atomic Energies. *Journal of Chemical Theory and Computation* **2023**, 19, 2029-2038.
32. Smith, J. S.; Isayev, O.; Roitberg, A. E., ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical Science* **2017**, 8, 3192-3203.
33. Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E., Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nature Communications* **2019**, 10, 2903.
34. Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E., Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *Journal of Chemical Theory and Computation* **2020**, 16, 4192-4202.
35. Rosenberger, D.; Smith, J. S.; Garcia, A. E., Modeling of Peptides with Classical and Novel Machine Learning Force Fields: A Comparison. *The Journal of Physical Chemistry B* **2021**, 125, 3598-3612.
36. Lahey, S.-L. J.; Thien Phuc, T. N.; Rowley, C. N., Benchmarking Force Field and the ANI Neural Network Potentials for the Torsional Potential Energy Surface of Biaryl Drug Fragments. *Journal of Chemical Information and Modeling* **2020**, 60, 6258-6268.
37. Tu, N. T. P.; Rezajooei, N.; Johnson, E. R.; Rowley, C. N., A neural network potential with rigorous treatment of long-range dispersion. *Digital Discovery* **2023**, 2, 718-727.
38. Reimers, J. R.; Hall, L. E., The Solvation of Acetonitrile. *Journal of the American Chemical Society* **1999**, 121, 3730-3744.
39. D.A. Case, H. M. A., K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, N. Forouzesh, G. Giambaşu, T. Giese, M.K. Gilson, H. Gohlke, A.W. Goetz, J. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, X. Wu, Y. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, Q. Zhu, and P.A. Kollman Amber 2023. *University of California, San Francisc.*, **2023**.
40. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004**, 25, 1157-1174.

41. Mobley, D. L.; Dumont, É.; Chodera, J. D.; Dill, K. A., Comparison of Charge Models for Fixed-Charge Force Fields: Small-Molecule Hydration Free Energies in Explicit Solvent. *The Journal of Physical Chemistry B* **2007**, 111, 2242-2254.
42. He, X.; Walker, B.; Man, V. H.; Ren, P.; Wang, J., Recent progress in general force fields of small molecules. *Current Opinion in Structural Biology* **2022**, 72, 187-193.
43. Hruska, E.; Gale, A.; Huang, X.; Liu, F., AutoSolvate: A toolkit for automating quantum chemistry design and discovery of solvated molecules. *The Journal of Chemical Physics* **2022**, 156, 124801.
44. He, X.; Man, V. H.; Yang, W.; Lee, T. S.; Wang, J., A fast and high-quality charge model for the next generation general AMBER force field. *J Chem Phys* **2020**, 153, 114502.
45. Grabuleda, X.; Jaime, C.; Kollman, P. A., Molecular dynamics simulation studies of liquid acetonitrile: New six-site model. *Journal of Computational Chemistry* **2000**, 21, 901-908.
46. Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A., Machine Learning for Molecular and Materials Science. *Nature* **2018**, 559, 547.
47. Smith, J. S.; Roitberg, A. E.; Isayev, O., Transforming Computational Drug Discovery with Machine Learning and AI. *ACS Medicinal Chemistry Letters* **2018**, 9, 1065-1069.
48. Hobza, P.; Šponer, J., Toward True DNA Base-Stacking Energies: MP2, CCSD(T), and Complete Basis Set Calculations. *Journal of the American Chemical Society* **2002**, 124, 11802-11808.
49. Feller, D.; Peterson, K. A.; Crawford, T. D., Sources of error in electronic structure calculations on small chemical systems. *The Journal of Chemical Physics* **2006**, 124, 054107.
50. Neugebauer, H.; Bohle, F.; Bursch, M.; Hansen, A.; Grimme, S., Benchmark Study of Electrochemical Redox Potentials Calculated with Semiempirical and DFT Methods. *The Journal of Physical Chemistry A* **2020**, 124, 7166-7176.
51. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry* **2002**, 23, 1623-1641.
52. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977**, 23, 327-341.
53. Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Bjerre Jensen, P.; Kermode, J.; Kitchin, J. R.; Leonhard Kolsbjerg, E.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Bergmann Maronsson, J.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W., The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **2017**, 29, 273002.
54. Seritan, S.; Bannwarth, C.; Fales, B. S.; Hohenstein, E. G.; Isborn, C. M.; Kokkila-Schumacher, S. I. L.; Li, X.; Liu, F.; Luehr, N.; Snyder Jr., J. W.; Song, C.; Titov, A. V.; Ufimtsev, I. S.; Wang, L.-P.; Martínez, T. J., TeraChem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics. *WIREs Computational Molecular Science* **2021**, 11, e1494.
55. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, 132.
56. Raynaud, C.; Poteau, R.; Maron, L.; Jolibois, F., Ab initio molecular dynamics simulation of the UV absorption spectrum of β-ionone. *Journal of Molecular Structure: THEOCHEM* **2006**, 771, 43-50.

57. Schlegel, H. B.; Millam, J. M.; Iyengar, S. S.; Voth, G. A.; Daniels, A. D.; Scuseria, G. E.; Frisch, M. J., Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. *The Journal of Chemical Physics* **2001**, 114, 9758-9763.
58. Schlegel, H. B.; Iyengar, S. S.; Li, X.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frisch, M. J., Ab initio molecular dynamics: Propagating the density matrix with Gaussian orbitals. III. Comparison with Born–Oppenheimer dynamics. *The Journal of Chemical Physics* **2002**, 117, 8694-8704.
59. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **1988**, 37, 785-789.
60. Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F., Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]. *The Journal of Chemical Physics* **2018**, 148.
61. Liu, F.; Luehr, N.; Kulik, H. J.; Martínez, T. J., Quantum Chemistry for Solvated Molecules on Graphical Processing Units Using Polarizable Continuum Models. *Journal of Chemical Theory and Computation* **2015**, 11, 3131-3144.
62. Liu, F.; Sanchez, D. M.; Kulik, H. J.; Martínez, T. J., Exploiting graphical processing units to enable quantum chemistry calculation of large solvated molecules with conductor-like polarizable continuum models. *International Journal of Quantum Chemistry* **2019**, 119, e25760.
63. Klamt, A.; Schuurmann, G., COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2* **1993**, 799-805.
64. York, D. M.; Karplus, M., A Smooth Solvation Potential Based on the Conductor-Like Screening Model. *The Journal of Physical Chemistry A* **1999**, 103, 11060-11079.
65. Lange, A. W.; Herbert, J. M., A smooth, nonsingular, and faithful discretization scheme for polarizable continuum models: The switching/Gaussian approach. *The Journal of Chemical Physics* **2010**, 133, -.
66. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chemical Reviews* **2005**, 105, 2999-3094.
67. Chowdhuri, S.; Chandra, A., Dynamics of Halide Ion−Water Hydrogen Bonds in Aqueous Solutions: Dependence on Ion Size and Temperature. *The Journal of Physical Chemistry B* **2006**, 110, 9674-9680.
68. Gowers, R. J.; Carbone, P., A multiscale approach to model hydrogen bonding: The case of polyamide. *The Journal of Chemical Physics* **2015**, 142, 224907.
69. Martínez, L.; Shimizu, S., Molecular Interpretation of Preferential Interactions in Protein Solvation: A Solvent-Shell Perspective by Means of Minimum-Distance Distribution Functions. *Journal of Chemical Theory and Computation* **2017**, 13, 6358-6372.
70. Martínez, L., ComplexMixtures.jl: Investigating the structure of solutions of complex-shaped molecules from a solvent-shell perspective. *Journal of Molecular Liquids* **2022**, 347, 117945.
71. Shmool, T. A.; Martin, L. K.; Matthews, R. P.; Hallett, J. P., Ionic Liquid-Based Strategy for Predicting Protein Aggregation Propensity and Thermodynamic Stability. *JACS Au* **2022**, 2, 2068-2080.
72. Tanoury, G. J.; Iyemperumal, S. K.; Lee, E. C., Toward a Combined Molecular Dynamics and Quantum Mechanical Approach to Understanding Solvent Effects on Chemical Processes in the Pharmaceutical Industry: The Case of a Lewis Acid-Mediated SNAr Reaction. *Organic Process Research & Development* **2023**, 27, 742-754.

73. El-Ayaan, U.; Murata, F.; Fukuda, Y., Thermochromism and Solvatochromism in Solution. *Monatshefte für Chemie / Chemical Monthly* **2001**, 132, 1279-1294.
74. Sıdır, İ.; Sıdır, Y. G., Solvatochromism and intramolecular hydrogen-bonding assisted dipole moment of phenyl 1-hydroxy-2-naphthoate in the ground and excited states. *Journal of Molecular Liquids* **2016**, 221, 972-985.
75. De Silva, N.; Willow, S. Y.; Gordon, M. S., Solvent Induced Shifts in the UV Spectrum of Amides. *The Journal of Physical Chemistry A* **2013**, 117, 11847-11855.
76. DeFusco, A.; Minezawa, N.; Slipchenko, L. V.; Zahariev, F.; Gordon, M. S., Modeling Solvent Effects on Electronic Excited States. *The Journal of Physical Chemistry Letters* **2011**, 2, 2184-2192.
77. Simón, L.; Goodman, J. M., Enzyme Catalysis by Hydrogen Bonds: The Balance between Transition State Binding and Substrate Binding in Oxyanion Holes. *The Journal of Organic Chemistry* **2010**, 75, 1831-1840.
78. Zaks, A.; Klibanov, A. M., The effect of water on enzyme action in organic media. *Journal of Biological Chemistry* **1988**, 263, 8017-8021.
79. Fitzpatrick, P. A.; Klibanov, A. M., How can the solvent affect enzyme enantioselectivity? *Journal of the American Chemical Society* **1991**, 113, 3166-3171.
80. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The missing term in effective pair potentials. *The Journal of Physical Chemistry* **1987**, 91, 6269-6271.
81. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, 79, 926-935.
82. Chen, X.; Li, P.; Hruska, E.; Liu, F., Δ-Machine learning for quantum chemistry prediction of solution-phase molecular properties at the ground and excited states. *Physical Chemistry Chemical Physics* **2023**, 25, 13417-13428.