JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

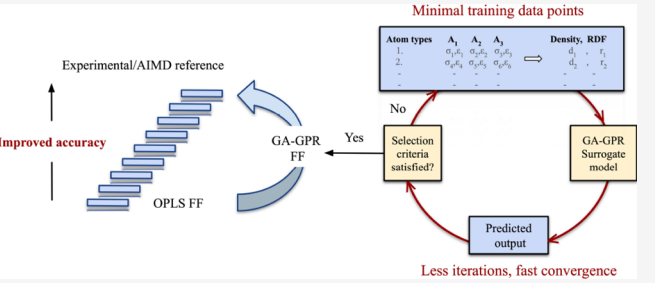
Active-Learning Assisted General Framework for Efficient Parameterization of Force-Fields

Yati, Yash Kokane, and Anirban Mondal*

Cite This: https://doi.org/10.1021/acs.jctc.5c00061 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: This work presents an efficient approach to optimizing force field parameters for sulfone molecules using a combination of genetic algorithms (GA) and Gaussian process regression (GPR). Sulfone-based electrolytes are of significant interest in energy storage applications, where accurate modeling of their structural and transport properties is essential. Traditional force field parametrization methods are often computationally expensive and require extensive manual intervention. By integrating GA and GPR, our active learning framework addresses these challenges by achieving optimized parameters in 12 iterations using only 300 data points, significantly outperforming previous attempts requiring thousands of iterations and parameters. We demonstrate the efficiency of our method through a comparison with state-of-the-art techniques, including Bayesian Optimization. The optimized GA-GPR force field was validated against experimental and reference data, including density, viscosity, diffusion coefficients, and surface tension. The results demonstrated excellent agreement between GA-GPR predictions and experimental values, outperforming the widely used OPLS force field. The GA-GPR model accurately captured both bulk and interfacial properties, effectively describing molecular mobility, caging effects, and interfacial arrangements. Furthermore, the transferability of the GA-GPR force field across different temperatures and sulfone structures underscores its robustness and versatility. Our study provides a reliable and transferable force field for sulfone molecules, significantly enhancing the accuracy and efficiency of molecular simulations. This work establishes a strong foundation for future machine learning-driven force field development, applicable to complex molecular systems.



1. INTRODUCTION

Lithium-ion batteries (LIBs) have become the cornerstone of modern energy storage technologies due to their high energy density and long cycle life.[1−3] However, traditional LIBs that rely on carbonate-based electrolytes encounter significant challenges, such as poor thermal stability, slow charge−discharge cycles, and volatility.[4,5] These limitations hinder their efficiency and long-term performance. In response to these issues, researchers have explored the addition of organic additives to high-concentration electrolytes (HCEs) as a promising strategy to enhance battery performance.[6−8] Among these additives, sulfones have emerged as particularly effective due to their unique properties.[7,9−11] Sulfolane, a type of sulfone, has been shown to improve the thermal stability and electrochemical performance of HCEs, making it a valuable candidate for next-generation LIBs.[6,9,12−17] The addition of sulfones facilitates lithium-ion transport through a hopping conduction mechanism,[9,14] which significantly enhances the conductivity of the electrolyte, thus addressing the challenges associated with traditional carbonate-based systems.

Among the sulfone family, additives such as sulfolane (SL), dimethyl sulfone (DMS), ethyl methyl sulfone (EMS), and 3-methyl sulfolane (MSL) are frequently used to enhance the performance of these electrolytes.[18−20] Recent studies have increasingly focused on HCEs, which may consist of binary, ternary, or even more complex mixtures, depending on the combination of different additives.[18−21] The challenge, however, lies in the sheer complexity of these mixtures. Including various additives at different concentrations introduces a vast number of potential combinations, with no established guidelines to determine the optimal choice of additives or their concentrations for a given lithium salt. This complexity makes it difficult to predict which combinations yield the most effective electrolytic properties. Given these challenges, there is a growing need for a computationally efficient approach to screen and identify the most promising electrolyte mixtures from the myriad possibilities. To achieve this, it is crucial to accurately simulate the structural and dynamical properties of each system, ensuring that predictions are both reliable and precise.

Received: January 14, 2025
Revised: February 18, 2025
Accepted: February 20, 2025

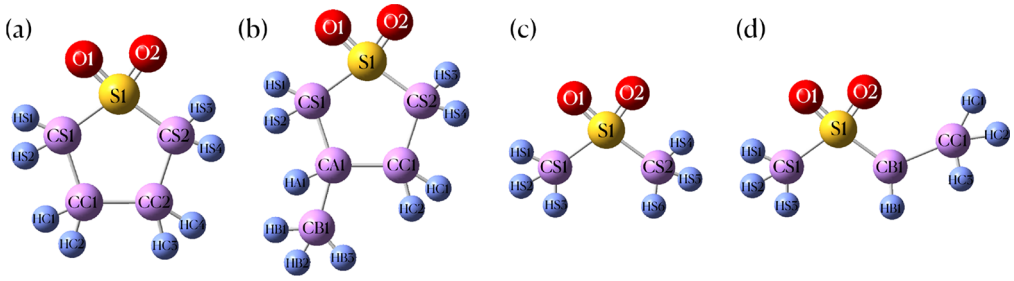


Figure 1. Molecular structures and atom types of the investigated sulfone molecules in this study: (a) sulfolane (SL), (b) 3-methyl sulfolane (MSL), (c) dimethyl sulfone (DMS), and (d) ethyl methyl sulfone (EMS).

The choice of simulation method plays a critical role in accurately depicting the bulk dynamics of electrolyte systems. While quantum mechanical (QM) simulations offer high accuracy, they are limited by the computational cost, which restricts the size of the system and the time scale that can be simulated. This limitation makes it difficult to capture the dynamic correlations over time, which are essential for understanding the behavior of complex electrolyte mixtures. On the other hand, classical molecular dynamics (MD) simulations can overcome these challenges by allowing larger system sizes and longer simulation times, provided that the atomic charges and parameters defining nonbonded and bonded interactions are appropriately chosen. However, while sulfones share similar chemical structures, they exhibit significant differences in their physicochemical properties, requiring distinct treatment in simulations.[22] This means that generalized force fields (FFs), such as OPLS[23,24] or GAFF,[25] are often inadequate for accurately modeling these molecules.

The primary issue with existing force fields like OPLS[23,24] is their inaccurate charge distribution, which can distort the bulk dynamics of the system. Additionally, the structural arrangement of these molecules, governed by van der Waals and electrostatic interactions, is sensitive to the parameters used in simulations. Except for sulfolane, computational studies on other sulfones, such as dimethyl sulfone, ethyl methyl sulfone, and 3-methyl sulfolane, are sparse or nonexistent.[14,26] This gap in the literature underscores the need to develop specific parameters for each sulfone before tackling more complex mixtures.

The traditional force field parametrization approach often involves manually tuning parameters based on QM calculations.[14,26−28] While this method can yield accurate results, it is time-consuming and inefficient, particularly when applied to a series of structurally similar molecules. For example, previous efforts have successfully parametrized sulfolane using manual techniques, but this approach lacks scalability for more extensive studies.[26] An alternative to manual parametrization is using semiautomated workflows, as discussed in the literature.[29−32] These methods have achieved accuracy comparable to ab initio references but still require multiple iterations of simulations, reducing overall efficiency.[30] Moreover, building workflows that converge quickly often requires the generation of large input data sets, which shifts the burden of manual effort to the initial stages of the process rather than eliminating it.[29]

In recent years, machine learning (ML) approaches have been explored to refine force field (FF) parameters, offering improved efficiency.[29−31,33−36] Notably, Bayesian Optimization (BO) has demonstrated promise in effectively navigating parameter spaces[37−40] However, these methods often depend heavily on large initial data sets to train the model, which can be expensive and time-consuming to gather. The need for extensive training data can diminish the practical advantages of ML-based parametrization, particularly when the goal is to minimize computational and manual effort. Given these challenges, we aimed to develop a parametrization method that balances efficiency with accuracy. By reducing reliance on manual intervention and optimizing the workflow, we sought to create a more streamlined and effective approach to FF development that leverages the strengths of machine learning without the drawbacks of excessive data dependence.

To address the limitations of traditional and existing ML-based force field (FF) development approaches, we devised a workflow that combines a genetic algorithm (GA) with a Gaussian process regression (GPR) surrogate model, which emulates the outcomes of molecular simulations to significantly reduce the computational cost of parameter refinement.[41] While GA and GPR-based surrogate models have been previously employed for FF optimization, existing methods often require large data sets[29,33,40,42−44] or numerous iterations[44,45] to converge. Our novel active learning framework efficiently balances the trade-off between data set size and optimization iterations, achieving convergence within only 12 iterations and 300 data points. The GA-GPR active learning framework offers a complementary approach to gradient-based methods, with potential advantages in global optimization and handling complex properties. We initiated the process with a small set of 200 training data points, carefully selected for their relevance to the target properties. 8−15 new parameters were added during each iteration based on a predefined fitness function, resulting in rapid and efficient parameter optimization. The GA-GPR workflow emphasizes system-specific parametrization for pure sulfone systems, accurately capturing both structural and transport properties across temperatures. Moreover, the final parameters, derived through this automated framework and validated via extended MD simulations, accurately reproduce atomic-level details and dynamic behaviors while demonstrating excellent transferability across cyclic and acyclic sulfones. Despite training the model at a single temperature, the derived parameters maintained their accuracy across different temperatures, demonstrating a transferability level that is rare and valuable in force field development. These advancements establish a robust and computationally efficient platform for FF development, significantly outperforming traditional force fields like OPLS in reproducing bulk and interfacial properties.

2. METHODS

In this study, we aimed to develop a reliable and accurate force field for four sulfone molecules utilizing the OPLS functional

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

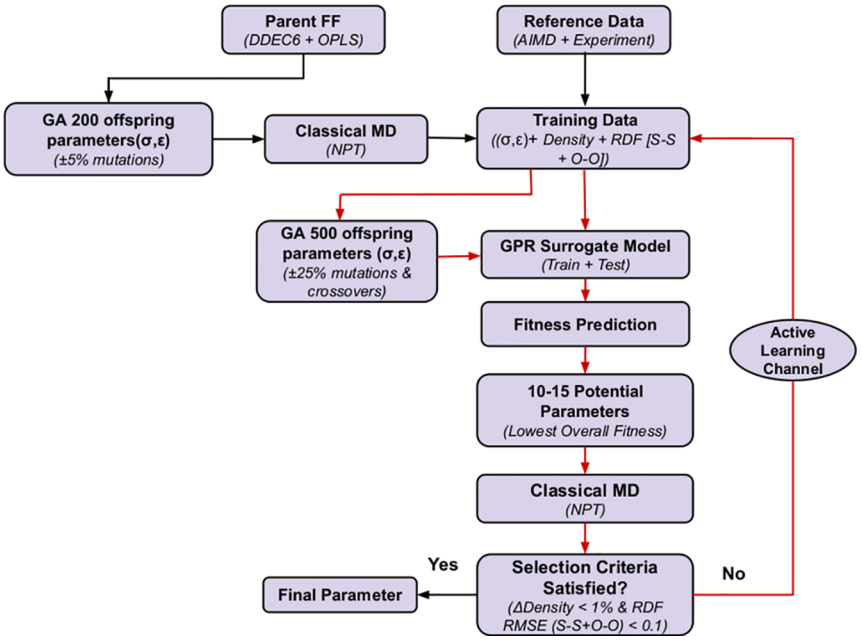


Figure 2. Workflow for the optimization process using Genetic Algorithm (GA) and Gaussian Process Regression (GPR). The GA generates new LJ parameters through mutation and crossover. A surrogate GPR model predicts the fitness of these parameters, and the top candidates are selected for classical molecular dynamics simulations. The results are used to retrain the GPR model in an active learning loop (red arrows), iteratively refining the force field parameters to optimize density and RDF predictions.

form. The molecular structures with atom labeling utilized are shown in Figure 1. The force field describes the interactions within and between molecules through a combination of bonded and nonbonded terms. The total potential energy U_{total} for a system of N molecules can be expressed as

$$U_{total} = \sum_{bonds} k_b(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \sum_{n=0}^{5} C_n \cos^n(\phi) + \sum_{i<j} \left[4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right]$$

where r and θ represent bond lengths and angles, respectively, with r_0 and θ_0 being their equilibrium values. k_b and k_θ are force constants for bonds and angles. ϕ is the dihedral angle, and C_n are the Ryckaert-Bellemans coefficients. The summation runs from n = 0 to 5 for each dihedral, allowing for a flexible representation of dihedral interactions in nonpolarizable force fields. ϵ_{ij} and σ_{ij} represent the Lennard−Jones (LJ) potential parameters for nonbonded interactions between atoms i and j. q_i and q_j are the partial charges on atoms i and j, respectively, with r_{ij} being the distance between them. In this framework, intramolecular bonded parameters, including bond lengths, angles, and dihedral potentials, were directly adopted from the standard OPLS force field.[23,24] For the partial charges, we employed the DDEC6[46] method, implemented in `Chargemol v09_26_2017`,[47] to compute accurate atomic charges for each molecule. First, isolated sulfone molecules were optimized in their ground state using density functional theory (DFT) at the M062X/aug-cc-pVDZ level of theory, as implemented in Gaussian 09.[48] The resulting wave function information from these DFT calculations served as the input for the DDEC6 method, which then calculated the partial charges by partitioning the electron density to match the electrostatic potential and chemical characteristics of the molecules.

We aimed to estimate the optimal values using reference data for the nonbonded interactions, particularly the LJ parameters σ and ϵ. The unlike interactions between atom types were calculated using the Lorentz−Berthelot combining rules. The optimization of the LJ parameters was guided by two primary reference quantities: the liquid phase density and site-specific radial distribution functions (RDFs), particularly the S−S and O−O RDFs. The liquid phase density was primarily obtained from experimental data.[22,49−56] In cases where experimental data were unavailable, we relied on values derived from ab initio molecular dynamics (AIMD) simulations. These AIMD simulations also provided the trajectories necessary to compute the RDFs, offering a detailed depiction of the molecular structure and interactions in the liquid phase. A summary of the steps involved in the optimization process is displayed in Figure 2.

2.1. Reference Database Generation. *2.1.1. Ab Initio Molecular Dynamics Simulations.* The reference values for fitness calculation were obtained through AIMD simulations conducted in the liquid state for each sulfone molecule. The optimized ground-state geometries of each molecule were used to construct initial configurations where 30 molecules were packed into a cubic box with a 20 Å side length using Packmol.[57] Before performing the AIMD run, the initial configuration was equilibrated using classical molecular dynamics in an NPT ensemble for 500 ps. The supercell was then geometry optimized within the density functional theory framework at the same level of theory as described below. The gradients on the wave functions and the nuclear positions were optimized with convergence criteria of 10^{-6} and 10^{-3} a.u., respectively. Using the Quickstep module, the quenched geometry was used to set up Born−Oppenheimer MD

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

simulations with the electronic structure code CP2K.[58] The AIMD simulations used density functional theory within the Gaussian and plane wave (GPW) framework, utilizing Perdew–Burke–Ernzerhof's (PBE) functional[59] to account for exchange-correlation effects. Dispersive interaction corrections were included using the empirical dispersion correction (D3) from Grimme,[60] with a cutoff of 40 Å. The wave functions were calculated using double ζ valence polarization (DZVP) basis sets with short-range terms for each atom type with an energy cutoff of 500 Ry. The norm-conserving Goedecker–Teter–Hutter (GTH) pseudopotentials[61,62] were applied to consider the effect of nuclei and core electrons. Equations of motion were integrated with a time step of 1 fs. The system was equilibrated for 56 ps in the isothermal–isobaric (NPT) ensemble at 1 bar using an isotropic unit cell according to the scheme of Martyna et al.,[63] with a time constant of 500 fs. A production simulation followed equilibration in the canonical NVT ensemble for 20 ps. We set the temperature at 303 K, except for DMS, where a higher temperature of 393 K was used due to its elevated melting point, controlled by one chain of six Nosé–Hoover thermostats with a time constant of 100 fs.[64] Three-dimensional periodic boundary conditions were employed in all simulations. The trajectory after every 10 fs time step was stored for postsimulation analysis. The RDFs for sulfur–sulfur (S–S) and oxygen–oxygen (O–O) pairs were computed by analyzing the trajectories. The final set of liquid phase densities and the S–S and O–O RDFs served as reference data for the machine learning model, providing a benchmark to optimize the force field parameters efficiently.

2.1.2. Classical Molecular Dynamics Simulations. : To develop the Gaussian process regression surrogate model, we first generated a training data set through a series of classical MD simulations. The initial force field was based on the OPLS functional form, but we replaced the atomic site charges with those computed using the DDEC6 method, resulting in a parent parameter set. To explore the parameter space, we employed a genetic algorithm that introduced mutations to this parent vector by altering the parameters by ±5%, thereby generating 200 offspring parameter sets.

We then performed classical MD simulations on 3000 sulfone molecules using these 200 mutated parameter sets using the GROMACS engine.[65] These simulations were conducted under the same thermodynamic conditions as those used in the AIMD simulations to achieve equilibrated configurations. The simulations were initialized from energetically minimized configurations and employed a time step of 1 fs. A cubic simulation box was used with periodic boundary conditions applied in all directions. The simulations were conducted in the NPT ensemble, utilizing a velocity-rescaling thermostat[66] to maintain temperature and a Berendsen barostat[67] to control pressure. Electrostatic interactions were calculated using the particle mesh Ewald (PME) method for real-space calculations and fast Fourier transform (FFT) for reciprocal lattice points. The cutoff for Coulombic and van der Waals interactions was 1.3 nm. These simulations were done for 500 ps.

For each of the 200 parameter sets, the liquid phase density and radial distribution functions for sulfur–sulfur (S–S) and oxygen–oxygen (O–O) pairs were calculated once volume equilibration was achieved. These computed densities and RDFs constituted the training and testing data for the GPR surrogate model, providing the necessary input to optimize the force field parameters. This approach ensured a comprehensive exploration of the parameter space, facilitating the development of an accurate and efficient force field tailored for sulfones.

2.2. Optimization Workflow. *2.2.1. Fitness Function for LJ Parameter Optimization.* The next step in developing optimal force field parameters involves defining a metric to evaluate how well the predicted force fields reproduce the reference liquid phase properties, specifically the density and radial distribution functions. These reference properties were obtained from AIMD simulations or available experimental measurements. The accuracy of the predicted force fields was quantified using a fitness function that measures the deviation between the predicted values and the reference data.

The overall fitness function, F, was designed to simultaneously minimize two key objectives: the error in density and the deviation in RDFs. This is expressed as a weighted sum of the fitness components for density (F_{density}) and RDFs (F_{RDF}).

$$F = w_{\text{density}}F_{\text{density}} + w_{\text{RDF}}F_{\text{RDF}} \tag{1}$$

Here, the weights w_{density} and w_{RDF} account for the different units and magnitudes of the two components. The fitness component F_{density} represents the absolute error between the predicted and reference densities, and it is defined as follows.

$$F_{\text{density}} = |D_{\text{FF}} - D_{\text{Ref.}}| \tag{2}$$

Where D_{FF} is the density obtained from classical MD simulations using the predicted force field, and $D_{\text{ref.}}$ is the density from experiment or AIMD. The RDF component of the fitness function, F_{RDF}, is unitless and calculates the root-mean-square error between the predicted and reference RDFs for both intermolecular sulfur–sulfur (S–S) and oxygen–oxygen (O–O) interactions.

$$F_{\text{RDF}} = \sqrt{\frac{\chi^2_{\text{SS}} + \chi^2_{\text{OO}}}{2}} \tag{3}$$

The terms χ^2_{SS} and χ^2_{OO} quantify the deviation between the predicted and reference RDFs and are computed as follows.

$$\chi^2_{\alpha\alpha} = \frac{\sum_r [g^{\text{AIMD}}_{\alpha\alpha}(r) - g^{\text{MD}}_{\alpha\alpha}(r)]^2}{\sum_r [g^{\text{AIMD}}_{\alpha\alpha}(r)]^2} \tag{4}$$

where $g^{\text{AIMD}}_{\alpha\alpha}(r)$ and $g^{\text{MD}}_{\alpha\alpha}(r)$ are the RDFs for atom pairs α–α (S–S or O–O), and r is the distance over which RDFs were calculated, ranging from 0.0 to 9.0 Å, with a spacing of 0.02 Å.

Thus, the fitness function (eq 1) reduces the task of determining the optimal force field parameters to an optimization problem. The goal is to minimize this function by adjusting the nonbonding parameters σ and ϵ, ensuring that both the structural (RDF) and thermodynamic (density) properties of the sulfone molecules are accurately reproduced.

2.2.2. Surrogate Gaussian Process. The Gaussian process regression model, implemented using the `Scikit-learn` library,[68] was employed to predict the fitness values corresponding to different LJ parameter sets. Initially trained on a data set of 200 parameter sets, the GPR model was integrated into the Genetic Algorithm framework to guide the search for optimized parameter sets (see Figure 2). The GPR model used a composite kernel function, which was defined as follows:

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

$$k(x, x') = C(1.0, (1e^{-3}, 1e^{3})) \times \mathrm{RBF}(1, (1e^{-2}, 1e^{2})) \quad (5)$$

In this expression, $C(1.0, (1e^{-3}, 1e^{3}))$ denotes the constant kernel, which applies a scaling factor to account for the global variance in the data. The constant kernel starts with an initial variance of 1.0, with bounds between 0.001 and 1000, allowing the model to estimate global variance flexibly.

The Radial Basis Function (RBF) kernel, with an initial length scale $l = 1$, ensures smoothness in the predicted fitness values. The length scale governs how quickly correlations between data points decay, and its bounds (0.01−100) provide the flexibility to capture both fine details and broader trends in the data. The GPR model's hyperparameters, such as the RBF kernel's length scale and the constant kernel's variance, were optimized through 10 restarts to ensure robust training. A standard scaler was applied to the data to enhance numerical stability during the optimization process.

It should be noted that the GPR model was initially trained (iteration zero) using 200 reference data points, as described in Section 2.1. Integrating the GPR model with the GA represents an active learning process (denoted by red arrows in Figure 2), where the GPR model is continuously improved by incorporating new data into the reference data set at each iteration. This active learning step, discussed in the next section, enhances the model's predictive accuracy as it iteratively refines its understanding of the search space.

2.2.3. Genetic Algorithm and Active Learning. The genetic algorithm was employed to systematically explore the LJ parameter space by generating new candidate sets of parameters. A custom GA was developed using the `DEAP library`,[69] which integrated the surrogate GPR model to estimate the fitness of newly generated parameters. The initial population for the GA was generated by applying mutations on the reference data set individuals, consisting of parameters for 200 MD simulations, as discussed in Section 2.1. These mutations involved scaling each parameter within a random range of 75−125% of its original value while being constrained within predefined bounds to ensure physically meaningful results. Following the mutation step, crossover operations were carried out with a crossover probability of 0.4. In each crossover, parameters (genes) from two-parent individuals were swapped based on a randomly generated mask, producing new offspring individuals. A set of 500 new candidate parameters was generated through this mutation and crossover process.

The initial parameter bounds for σ and ε were set within ±5% deviations from OPLS nonbonded parameters to ensure numerical stability and physically meaningful values in preliminary simulations. This narrow range prevented the optimizer from exploring highly unphysical regions before sufficient training data was available. While initial deviations in density and RDFs were noticeable, they followed expected trends, allowing us to refine the parameter space systematically. Based on the observed range of σ and ε values in the initial data set, we expanded the bounds to a broader range (0.1−2.5) for subsequent iterations. Additionally, extreme parameter values leading to simulation crashes were excluded from the training set, enabling adaptive learning while ensuring the physical plausibility of the optimized parameters. This iterative refinement strategy balanced exploration with stability, ensuring efficient and reliable optimization of force field parameters.

The fitness of these newly generated individuals was then predicted using the surrogate GPR model trained on predicting fitness using the reference data. This allowed for rapid assessment without the need for computationally expensive simulations (see Figure 2). The individuals were ranked based on the GPR-predicted fitness, and the top 8−15 candidates were selected. These selected candidates were subjected to classical MD simulations to compute their actual fitness, following the same protocol as discussed in Section 2.1. These MD simulation fitness values, along with the corresponding LJ parameter sets, were then added to the training data set. The GPR model was retrained on this expanded data set, incorporating the new simulation results to improve its accuracy.

The active learning loop—where the GA generated new candidates, the GPR model predicted their fitness, and the top candidates were refined through MD simulations—was repeated iteratively. Classical MD simulations were conducted during each iteration for these 8−15 parameters with the lowest predicted fitness values, following the same procedure as discussed above. These new data points were added to the data set, including their corresponding density and RDFs. This iterative workflow ensured that with each cycle, the GPR model gained a deeper understanding of the optimization landscape, gradually honing in on the regions of the parameter space most likely to yield optimal LJ parameters.

2.2.4. Convergence and Freeze. By the end of 10−12 iterations, the model successfully predicted parameter sets with low overall fitness, including appropriate individual fitness scores for both density and RDF. Active learning significantly reduced fitness values compared to the initial data set, leading to faster convergence and more extensive exploration of parameter space beyond the original 200 data points. At the end of the active learning loop, we gathered nearly 300 data points on average for each sulfone system. The parameter set with the lowest fitness value from the final iteration was selected as the optimal force field. After obtaining the final set of parameters from the surrogate model, minor manual adjustments were made to further optimize the liquid-phase density and RDFs. This fine-tuning involved systematically varying parameters within a very narrow range (typically within ±2%) to evaluate their impact. Adjustments were retained only if they led to a simultaneous improvement in both density and RDFs while preserving consistency with experimental and ab initio reference data. The surrogate model's parameters were adopted without modification if no improvement was observed. This step ensured that the optimized force field achieved the best possible agreement with reference data while maintaining physical plausibility and transferability. The converged force field will be termed the GA-GPR model throughout the text.

2.3. Production Simulation. In the production phase, the GA-GPR optimized force field was utilized to perform classical MD simulations on 5000 sulfone molecules using the GROMACS engine.[65] The simulation protocol mirrored the earlier procedures as stated in Section 2.1, starting with an initial energy minimization step to ensure a stable configuration. This was followed by equilibration in the NPT ensemble for 5 ns, allowing the system to achieve the correct temperature and pressure. After equilibration, the production run was conducted in the NVT ensemble for an additional 5 ns. During this phase, the system's trajectory was recorded at every 1 fs to capture detailed molecular motion and dynamics.

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

The remaining simulation parameters were the same as Section 2.1.

To validate the developed force fields, viscosity (η) was determined by integrating the pressure autocorrelation function, employing the Green−Kubo relation.[70−72]

$$\eta = \frac{V}{k_B T}\int_0^\infty \langle P_{\alpha\beta}(0)P_{\alpha\beta}(t)\rangle dt \quad (6)$$

Here, V is the simulation volume, $P_{\alpha\beta}$ represents the pressure tensor components (both diagonal and off-diagonal). For robust capture of the short-time behavior of the autocorrelation function, the pressure tensor was recorded at every simulation step (1 fs). The integral in eq 6 converged under 30 ps for all systems, except for MSL, which converged within 55 ps, as shown in Figure S1 of the Supporting Information (SI). Autocorrelation functions were computed for 50−75 ps intervals, with starting points spaced by ∼2 ps for GA-GPR final parameters and ∼5 ps for OPLS parameters. The final reported viscosities represent an average of the cumulative integral of the last 20 ps.

Surface tension was estimated from a two-phase simulation, where sulfone molecules were in equilibrium with their vapor (essentially a vacuum under the studied conditions). A pre-equilibrated configuration of the bulk liquid containing 5000 molecules was used as the starting point. The simulation box length along the z-axis was extended up to 28 nm to create two liquid−vapor interfaces. The surface tension (γ) was calculated based on the pressure tensor differences between the component normal to the interface (P_{zz}) and the components parallel to the interface (P_{xx} and P_{yy}), using the following equation.

$$\gamma = \frac{L_z}{4}(2P_{zz} - P_{xx} - P_{yy}) \quad (7)$$

Here, L_z represents the length of the simulation box along the z-axis, and $P_{\alpha\alpha}$ are the diagonal terms of the pressure tensor. The NVT simulation was performed for 8 ns, with pressure tensor values extracted every 100 fs. The surface tension was then calculated by averaging the results over the trajectory.

3. RESULTS AND DISCUSSION

3.1. Liquid Density Comparison: AIMD vs Experimental Results. We first compare the liquid densities obtained from ab initio molecular dynamics simulations with experimental measurements to validate the accuracy of the AIMD method. The computed densities from AIMD (ρ^{AIMD}) at 303 K and 1 bar show excellent agreement with the experimental values (ρ^{exp}), as summarized in Table 1. The deviations ($\Delta\rho$) between the simulated and experimental densities are minimal, ranging from 0.09 to 1.40%. This level of agreement builds confidence in the reliability of AIMD as a reference for training machine learning models, particularly in optimizing LJ potential parameters. For instance, in the case of SL, the AIMD density is 1280.6 kg m^{-3}, which deviates by only 1.40% from the experimental average of 1262.9 kg m^{-3}. Similarly, the agreement for MSL and DMS is remarkable, with deviations of 0.09 and 0.36%, respectively. These minor deviations can be attributed to the inherent approximations in AIMD simulations and experimental uncertainties. Nonetheless, the overall close match supports the reliability of the AIMD results in predicting accurate liquid densities.

Table 1. Comparison between AIMD Simulated (ρ^{AIMD}) Liquid Density (in kg m^{-3}) against Experimental Measurements (ρ^{exp}), and Their Deviation ($\Delta\rho$)[a]

sulfones	ρ^{exp}	ρ^{AIMD}	$\Delta\rho$ (%)
SL	1262.9,[49] 1262.0,[50] 1262.3,[51,52] 1261.9,[22] 1261.4,[53] 1260.4,[54,55] 1260.8[56]	1280.6	1.40
MSL	1183.3,[53] 1184.1[22]	1184.4	0.09
DMS	1152.5[22]	1156.7	0.36

[a]All values are at 303 K and 1 bar, except for DMS, where measured and simulated values are at 393 K.

The close agreement between AIMD and experimental densities gives us substantial confidence in the quality of other computed properties, particularly the radial distribution functions derived from AIMD trajectories. Since the AIMD densities align closely with experimental data, it strongly suggests that the structural details captured by the RDFs are also accurate. These RDFs, in turn, provide a reliable foundation for training machine learning models aimed at optimizing LJ potential parameters. Thus, the high fidelity of the AIMD density results serves as a key indicator that the microscopic details reflected in the RDFs can be trusted to refine force fields, ensuring that the final optimized models faithfully replicate both macroscopic and molecular-level behaviors.

3.2. GA-GPR Force Field Parameterization. To assess the prediction accuracy of the GPR model before force field parametrization, an initial test was conducted to validate its performance in predicting the densities of the four sulfone molecules. The data set of 200 MD simulations was split into an 80:20 ratio for training and testing, respectively. The GPR model was trained using 160 parameter sets and their corresponding densities and then validated on the remaining 40 testing parameter sets. The results showed an R^2 value of 0.96 or higher for all four molecules. This indicates that the GPR model can effectively predict density based on the parameter sets (see Figure 3a). A similar process was repeated to evaluate the prediction accuracy for the overall fitness, which was defined as errors in radial distribution function and density. Here, too, the GPR model achieved an R^2 value exceeding 0.96, confirming its reliability in predicting both density and fitness.

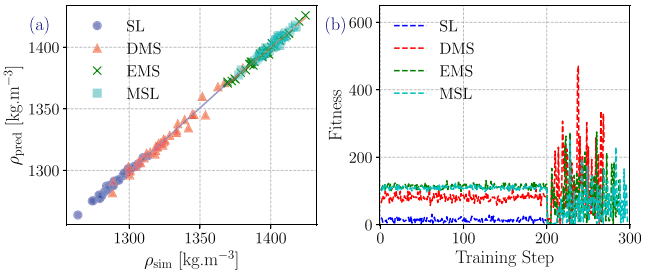


Figure 3. (a) Performance of the GPR surrogate model in predicting density: shows the correlation between actual simulated densities and GPR-predicted densities for the four sulfone systems. A strong agreement demonstrates the model's high prediction accuracy, with R^2 values exceeding 0.96 across all systems. (b) Evolution of the fitness function during GA-GPR training: illustrates the progressive improvement in the fitness function as new individuals are added using active learning. The rapid increase in fitness observed after 200 individuals reflects the exploration phase, followed by less variation as the algorithm fine-tunes the solution.

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

The evolution of the fitness function throughout the optimization process is illustrated in Figure 3b, which shows the fitness values of the first 300 individuals generated by the genetic algorithm for each of the four sulfone molecules. During the initial generations, significant variation in fitness values was observed among individuals, reflecting the broad exploration of the parameter space by the genetic algorithm. This variability was critical in ensuring the algorithm could effectively navigate the complex landscape of LJ parameter combinations.

The genetic algorithm leveraged the predictive power of the GPR model to guide the search toward promising regions of the parameter space, continually updating the fitness function to favor the best-performing parameter sets. Over subsequent iterations, the range of fitness values gradually narrowed, and convergence toward optimized parameter sets began to emerge. After approximately 12 iterations (except for SL, which took only a single iteration), the model successfully identified the optimal parameter set with the lowest fitness value for density and RDF. This indicated that the genetic algorithm, combined with the GPR model, could find LJ parameters that effectively reproduced the target properties of the sulfone molecules.

In the early stages of the optimization process, the model often predicted unrealistic Lennard-Jones parameters, as seen on verification with MD simulations. These deviations arose primarily from the model's exploration strategy, where it ventured into less relevant areas of the parameter search space. However, as the active learning framework iterated and continuously refined the GPR model using feedback from MD simulations, the model began to adapt and focus its exploration on the more meaningful and physically valid regions of the parameter space. This progressive refinement significantly reduced the occurrence of unrealistic predictions and guided the model toward generating LJ parameters that aligned well with expected physical properties.

The novelty of our active learning approach lies in its remarkable efficiency in converging to optimized force field parameters. In our approach, the active learning process required only 12 iterations (except for SL, which took only a single iteration), equivalent to approximately 300 individual parameter sets, to achieve convergence. This is significantly superior to earlier attempts reported in the literature, where a much larger number of training parameters, typically at least 1000 and up to 600 iterations, were required.[29,40,43,44] The efficiency of the combined GA-GPR workflow became evident as the genetic algorithm consistently converged to parameter sets with favorable fitness values in far fewer iterations. The gradual reduction in fitness variability and convergence to an optimal parameter set highlights the strength of this active learning-based optimization strategy in accurately modeling the target system properties with greater computational efficiency. Unlike our gradient-free GA-GPR framework, ref 73 employs a differentiation-based method, converging in 40−60 iterations. In comparison, our approach achieves optimization in as few as 1−10 iterations without requiring gradient evaluations, offering broader applicability to complex parameter landscapes.

The proposed GA-GPR framework is designed to efficiently optimize nonbonded parameters, specifically σ and ϵ, for each atom type. This method systematically explores the parameter space, ensuring robust performance even as the number of parameters increases. The iterative optimization process balances computational cost and accuracy by selectively expanding the training set with the most promising parameters in each iteration. This approach allows the model to explore a broader fitness landscape, leading to faster convergence while maintaining excellent agreement with target properties.

The tables below provide an overview of the optimized atomic site charges and Lennard-Jones parameters for the sulfone molecules investigated in this study. Specifically, Table 2 lists the DDEC computed atomic site charges for the four sulfone molecules, while Table 3 presents the optimized σ and ϵ parameters for the same set of molecules. For further details, individual data for each molecule are presented in Tables S1−S4.

Table 2. Atomic Site Charges for the Four Sulfone Molecules, Obtained from DDEC Calculations[a]

atom names	SL	DMS	EMS	MSL
S1	1.187726	1.596340	1.555625	1.493091
O1	−0.633322	−0.684529	−0.691636	−0.679080
O2	−0.633059	−0.684529	−0.691633	−0.665534
CS1	−0.318247	−0.560950	−0.555699	−0.400675
CS2	−0.317703	−0.560950		−0.361935
HS1	0.140548	0.149103	0.146207	0.126726
HS2	0.149467	0.149103	0.146206	0.120162
HS3	0.149774	0.149103	0.154753	0.118429
HS4	0.140364	0.149103		0.121963
HS5		0.149103		
HS6		0.149103		
CA1				0.081547
HA1				0.056889
CB1			−0.314231	−0.349532
HB1			0.101949	0.103750
HB2			0.101949	0.116082
HB3				0.103603
CC1	−0.124087		−0.303628	−0.156752
CC2	−0.124899			
HC1	0.099307		0.109445	0.088810
HC2	0.092293		0.120345	0.082456
HC3	0.092745		0.120348	
HC4	0.099093			

[a]Charges are listed according to the atom labeling scheme as shown in Figure 1.

3.3. GA-GPR Force Field Predictions. The following section discusses the predictions obtained using the GA-GPR force field and compares these results with available experimental data or AIMD simulations, as well as with predictions from other force fields. Table 4 provides a detailed comparison of simulated densities obtained from the OPLS and GA-GPR parametrizations, along with their percentage deviation from reference densities at 303 K (for SL, EMS, and MSL) and 393 K (for DMS). The liquid density of the sulfone molecules was calculated from molecular dynamics simulations performed using an NPT ensemble over 5 ns. Before data collection, an equilibration period was applied to ensure the system reached a stable state. The density values were averaged from the production phase of the simulation, which was divided into four blocks of 1 ns each to evaluate the uncertainty in the computed average.

The tabulated results (Table 4) show that the densities calculated using the OPLS parameters show significant deviations when compared to reference values, particularly for the cyclic sulfones. For SL, the OPLS-derived density

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 3. Atom Names and Their Corresponding σ and ϵ Parameters for the Four Sulfone Molecules[a]

	SL		DMS		EMS		MSL	
atom names	σ (nm)	ϵ (kJ/mol)	σ (nm)	ϵ (kJ/mol)	σ (nm)	ϵ (kJ/mol)	σ (nm)	ϵ (kJ/mol)
S	0.34443	1.13170	0.33885	1.15816	0.32019	1.04355	0.33140	1.71473
O	0.32411	0.41901	0.34916	0.52055	0.36028	0.58288	0.35990	0.35096
CS	0.35781	0.25916	0.35509	0.35276	0.17855	0.40939	0.32714	0.23181
HS	0.23111	0.13015	0.18692	0.14873	0.21544	0.25787	0.16952	0.17702
CA							0.36170	0.21388
HA							0.23590	0.15386
CB					0.32883	0.23501	0.36159	0.20522
HB					0.20275	0.13916	0.23136	0.15570
CC	0.34880	0.25952			0.19429	0.20673	0.37414	0.25074
HC	0.23821	0.13955			0.19234	0.12642	0.23279	0.17560

[a]The values are listed as per the atom labeling scheme as shown in Figure 1.

Table 4. Comparison of Simulated Densities (kg/m^3) Obtained Using OPLS (ρ^{OPLS}) and GA-GPR (ρ^{GA-GPR}) Parameters, along with Their Percentage Deviations ($\Delta\rho$) from Reference Densities ($\rho^{ref.}$) Taken from AIMD/Experimental Data at 303 K for SL, EMS, and MSL, and at 393 K for DMS

sulfones	$\rho^{ref.}$	ρ^{OPLS}	$\Delta\rho$ (%)	ρ^{GA-GPR}	$\Delta\rho$ (%)
SL	1262.90[49]	1304.00 ± 0.30	3.25	1259.45 ± 0.01	−0.27
EMS	1168.81	1158.55 ± 0.10	−0.88	1179.71 ± 0.16	0.93
MSL	1183.30[53]	1232.55 ± 0.23	4.16	1187.33 ± 0.19	0.34
DMS	1152.50[22]	1119.55 ± 0.22	−2.86	1155.85 ± 0.19	0.30

Table 5. Comparison of Simulated Densities (ρ^{GA-GPR}, kg/m^3) Using the GA-GPR Force Field with Experimental Densities ($\rho^{Exp.}$, kg/m^3) at Different Temperatures for Sulfone Molecules

	323 K		373 K	
sulfones	$\rho^{Exp.}$	ρ^{GA-GPR}	$\rho^{Exp.}$	ρ^{GA-GPR}
SL	1245.20,[49] 1244.70[22,53]	1239.05 ± 0.08	1200.90[22]	1188.80 ± 0.41
EMS		1148.08 ± 0.13		1066.82 ± 0.21
MSL	1167.20,[53] 1166.90[22]	1164.67 ± 0.19	1124.10[22]	1107.96 ± 0.13
	388 K		398 K	
sulfones	$\rho^{Exp.}$	ρ^{GA-GPR}	$\rho^{Exp.}$	ρ^{GA-GPR}
DMS	1157.40[22]	1161.62 ± 0.33	1147.70[22]	1150.34 ± 0.26

overestimated the reference by approximately 3.25%, while for MSL, an even more considerable overestimation of 4.16% was observed. In contrast, for the chain sulfones EMS and DMS, the densities were underestimated by 0.88 and 2.86%, respectively. This discrepancy can be attributed to the limitations of the OPLS force field in accurately capturing intermolecular interactions for different types of sulfones, suggesting that the intermolecular forces in cyclic sulfones are overestimated, leading to an overly dense structure. The GA-GPR force field, however, produced significantly improved results, with the computed densities for all the sulfone molecules showing deviations of less than 1% from the reference values. These slight deviations demonstrate the accuracy of the GA-GPR force field, particularly in capturing the intermolecular interactions necessary for reproducing the experimental or AIMD densities.

Continuing from the previous discussion on the accuracy of the GA-GPR force field in predicting liquid densities, we now focus on evaluating its performance across different temperatures and comparing these results with experimental measurements. As presented in Table 5, the GA-GPR force field parameters exhibit high accuracy in predicting the density of sulfones over a range of temperatures. For SL, the predicted densities at 323 and 373 K closely match the experimental values with deviations within 0.5%. This level of agreement indicates that the GA-GPR parameters effectively capture the temperature-dependent behavior of SL, reflecting accurate intermolecular interactions. The GA-GPR force field for MSL demonstrates strong predictive accuracy, with deviations consistently within 1%. This agreement further supports the transferability of the GA-GPR parameters across different temperatures, as they maintain high accuracy in representing bulk properties. Similarly, for DMS, the GA-GPR force field accurately predicts the density at 388 and 398 K, with deviations from experimental values within a narrow range.

Overall, the GA-GPR force field consistently provides accurate density predictions for sulfones across various temperatures (see Figure S2), with minimal deviations from experimental data. This demonstrates not only the reliability of the force field in predicting temperature-dependent properties but also its robust transferability. The previously reported SL density at room temperature showed a deviation of 2.71[14] and 2.33%.[26] At high temperatures, the maximum deviation reported using GA-GPR force field is below 1%, which is lower than the maximum deviation reported in ref 26 as 1.90%. The improved representation of intermolecular interactions achieved with GA-GPR parameters underscores their superiority over traditional force fields in simulating the liquid densities of sulfones.

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

3.3.1. Radial Distribution Functions. Figure 4 illustrates the radial distribution functions between sulfur−sulfur (S−S) and

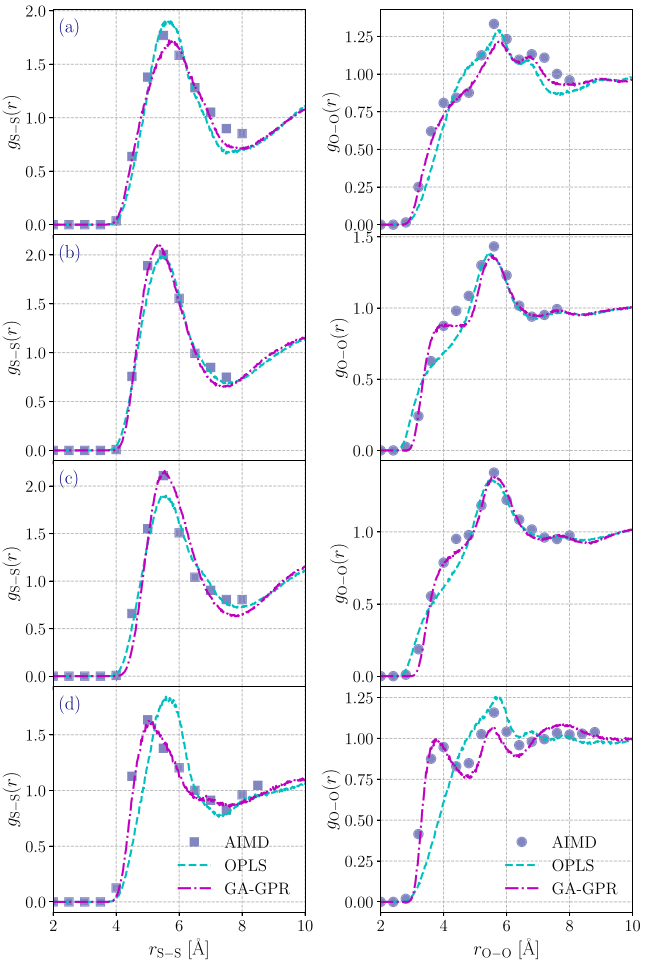


Figure 4. Radial distribution function for S−S and O−O atom pairs: (a) SL, (b) DMS, (c) EMS, and (d) MSL. Curves compare the atomic distributions obtained from AIMD (filled circles), OPLS (cyan), and the GA-GPR model (purple). The close agreement between the GA-GPR and AIMD curves demonstrates the model's accuracy in capturing the structural characteristics, while deviations in the OPLS curve highlight the differences in general force field-based predictions.

oxygen−oxygen (O−O) pairs in the sulfone molecules. These RDFs provide insight into the local structure and packing behavior of the molecules, and we compare the RDFs obtained using the GA-GPR force field with those from AIMD simulations and the OPLS force field to evaluate the accuracy of our approach. In this study, AIMD simulations were performed for 30 sulfone molecules over 20 ps under NVT conditions to generate reliable RDFs for structural validation of the optimized force field. While a 20 ps trajectory was chosen for better statistical accuracy, RDFs can often be obtained from shorter AIMD runs, particularly when focusing solely on structural properties rather than dynamical or transport properties, which require longer simulation times. Due to larger system sizes and molar ratio constraints, AIMD may be more computationally demanding for complex electrolyte systems. However, with carefully designed smaller-scale models or shorter runs, AIMD remains a practical and effective tool for deriving structural insights.

A close inspection of these RDFs reveals that the GA-GPR-derived RDFs are in excellent agreement with the AIMD results, indicating that the GA-GPR parameters successfully capture the local structure around sulfur and oxygen atoms. For the S−S RDF (left panels of Figure 4), the GA-GPR force field closely reproduces both the peak positions and intensities as observed in the AIMD simulations, indicating a well-defined representation of the coordination environment in the liquid phase. The first peak, representing the nearest neighbor S−S interactions, aligns precisely with the AIMD data, reflecting the accuracy of GA-GPR in predicting interatomic distances. In contrast, the OPLS force field shows noticeable deviations from the AIMD results. Specifically, the OPLS-derived S−S RDF displays a shift in the peak position (prominent in MSL, Figure 4d) or incorrect estimate of peak heights (evident from Figure 4a,c), indicating an unrealistic depiction of the local molecular environment. This suggests that OPLS parameters do not adequately represent the intermolecular interactions for sulfones, particularly in capturing the subtle features of S−S correlations.

Similarly, the GA-GPR model again demonstrates excellent predictive capability for the O−O RDFs, as shown in the right panels of Figure 4. The RDF obtained from GA-GPR matches the AIMD RDF in the location and magnitude of the primary and secondary peaks, indicating an accurate description of the short- and medium-range order. However, the OPLS force field fails to reproduce the AIMD reference accurately, with discrepancies observed in the peak positions and their relative intensities. Specifically, the OPLS-derived RDFs entirely miss the first shoulder peak present in O−O RDFs in all sulfone systems, as seen from the AIMD trajectory, with the maximum deviation seen in the case of MSL (Figure 4d).

Overall, the GA-GPR force field exhibits remarkable agreement with the AIMD results for both S−S and O−O RDFs, outperforming the OPLS force field in representing the local structural features of sulfone molecules. These observations highlight the effectiveness of the GA-GPR parametrization in capturing intermolecular interactions and suggest that the GA-GPR model offers a more reliable depiction of the molecular packing and coordination environment in the liquid phase.

3.3.2. Dipole Moment. The dipole moment is essential as it provides insight into the charge distribution within molecules, both in the isolated state and in the bulk liquid environment. Table 6 provides a summary of the dipole moments obtained from QM calculations (μ^{QM}), OPLS simulations (μ^{OPLS}), and GA-GPR simulations (μ^{GA-GPR}). The QM calculations were performed for isolated molecules using the M062X functional with the augmented correlation-consistent pVDZ basis set to account for the diffusion functions of the orbitals involved

Table 6. Comparison between Dipole Moment (Debye) Obtained from Quantum Mechanical Calculations (μ^{QM}) of Isolated Molecules with Classical Molecular Dynamics Simulations Employing OPLS (μ^{OPLS}) and GA-GPR (μ^{GA-GPR}) Potentials

sulfones	μ^{QM}	μ^{OPLS}	μ^{GA-GPR}
SL	5.68	7.44	6.17
EMS	4.81	4.75	4.74
MSL	5.72	7.71	5.05
DMS	4.97	4.55	4.57

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

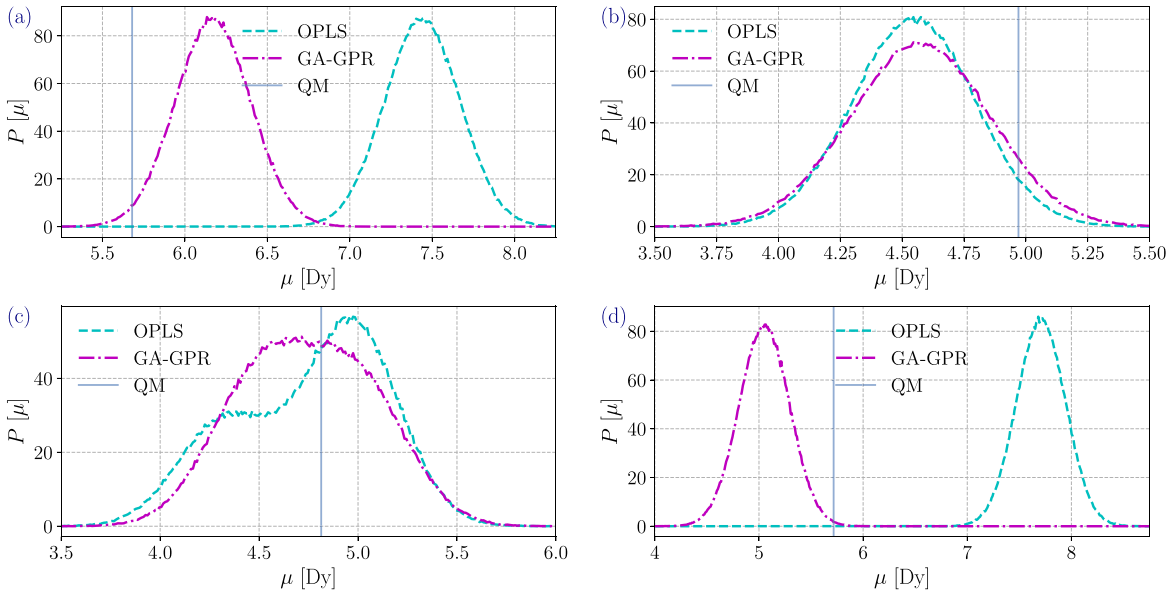


Figure 5. Distributions of dipole moments from MD simulated trajectories using GA-GPR and OPLS force fields: (a) SL, (b) DMS, (c) EMS, and (d) MSL. The solid vertical line indicates the dipole moment of the isolated molecule obtained from quantum mechanical calculations at the M062X/aug-cc-pVDZ level of theory, serving as a reference.

adequately. The dipole moments were extracted from a 5 ns NVT trajectory containing 5000 sulfone molecules for the MD simulations. The distribution of dipole moments over the molecular population is illustrated in Figure 5.

The results indicate that the GA-GPR force field better represents the dipole moment compared to OPLS, as evident from both the mean values and the overall distributions. For SL, the QM dipole moment is reported as 5.68 Dy, while the OPLS force field significantly overestimates it to 7.44 Dy. In contrast, the GA-GPR model predicts a value of 6.17 Dy, much closer to the QM reference, indicating improved accuracy in capturing the charge distribution in bulk. The dipole moment of SL reported in previous force field-based simulations were 6.178 Dy[14] and 5.922 Dy.[26] Similarly, for MSL, the OPLS force field produces a dipole moment of 7.71 Dy, which deviates considerably from the QM reference value of 5.72 Dy. On the other hand, the GA-GPR prediction of 5.05 Dy shows significantly improved agreement, highlighting the efficacy of the refined force field in accurately modeling dipole moments for such systems. In the case of EMS and DMS, the GA-GPR and OPLS force fields both show reasonable accuracy compared to the QM values, with GA-GPR performing slightly better overall. For EMS, the QM dipole moment is 4.81 Dy, and GA-GPR and OPLS predict similar values of 4.74 Dy and 4.75 Dy, respectively. For DMS, the QM dipole is 4.97 Dy, with GA-GPR predicting 4.57 Dy and OPLS predicting 4.55 Dy. It is important to note that within the GA-GPR framework, the atomic site charges were recalculated using the DDEC/c3 partitioning scheme to accurately capture the electronic distribution in sulfone molecules. The dipole moment was computed as the average molecular dipole across frames from the NVT trajectory, allowing us to assess the qualitative impact of charge modifications. While bonded parameters from the OPLS force field were retained, the recalibrated charges significantly improved the representation of intermolecular interactions, as reflected in the discrepancies between OPLS and GA-GPR force fields.

The dipole moment distributions in Figure 5 further support these observations. The GA-GPR-derived distributions show a clear peak that closely aligns with the QM reference, while the OPLS distributions are often broader and shifted away from the QM peaks, especially for SL and MSL. This suggests that GA-GPR provides a more realistic depiction of the electrostatic environment within the bulk, resulting in better agreement with the isolated molecular dipole moments. It is evident that the differences in dipole moment predictions between GA-GPR and OPLS align well with the trends observed for density predictions. Both force fields yield similar dipole moment values for linear sulfones such as DMS and EMS, reflecting their comparable performance in predicting liquid densities for these systems. However, in the case of cyclic sulfones like SL and MSL, significant discrepancies are observed between the dipole moments predicted by GA-GPR and OPLS, consistent with the differences in their density predictions. These findings highlight the GA-GPR force field's superior ability to accurately capture the electrostatic environment and bulk behavior, particularly for cyclic sulfones, where intermolecular interactions are more complex and require a refined parametrization for reliable modeling.

The dielectric constant (ε) is a critical property that reflects the collective polarization response of a material under an electric field and directly indicates the accuracy of molecular force fields in reproducing electrostatic interactions. To validate the GA-GPR potentials developed in this study, we computed the dielectric constants for sulfone molecules using MD simulations and compared the results with experimental measurements and the OPLS force field. These calculations were conducted using the `gmx_dielectric` module as implemented in GROMACS. As seen from Table 7, The GA-GPR potentials demonstrated a marked improvement over OPLS in reproducing experimental dielectric constants. For instance, in the case of SL, the GA-GPR dielectric constant (34.0) is significantly closer to the experimental range (43.26−43.33) than OPLS (29.8). Similarly, for MSL, GA-GPR (33.2) aligns better with experimental values (29.08−29.2) compared

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 7. Comparison between Dielectric Constant (ε) Obtained from Experiments (ε^{exp}) with Classical MD Simulations Employing OPLS (ε^{OPLS}) and GA-GPR (ε^{GA-GPR}) Potentials

sulfones	ε^{exp}	ε^{OPLS}	ε^{GA-GPR}
SL	43.26,[22] 43.33,[52] 43.3[53]	29.8	34.0
EMS		33.0	32.0
MSL	29.2,[53] 29.08[22]	35.4	33.2
DMS	45.78[22]	32.7	31.8

to OPLS (35.4). A previous study on GROMOS force field parametrization for DMS reported a dielectric constant of approximately 27[74] In contrast, the GA-GPR force field yielded a higher dielectric constant of 31.8, closer to the experimentally reported value of 45.78. These results highlight the capability of the GA-GPR framework to capture the subtle balance of short- and long-range electrostatic interactions, which are often inadequately described by traditional force fields. The systematic larger deviation of dielectric constants by OPLS, as observed for SL, MSL, and DMS, suggests limitations in its parametrization for accurately describing polarizable systems like sulfones. In contrast, the GA-GPR approach effectively incorporates polarization effects through ML-based parameter optimization, leading to superior agreement with experimental data.

3.3.3. Viscosity. Viscosity is a crucial transport property that reflects the ease molecules move past each other in a liquid. It is essential in the context of sulfone molecules used as additives in electrolytes, as their viscosity affects the ion transport efficiency and, consequently, the overall performance of battery systems. The viscosity values in this study were computed from classical MD simulations using the Green−Kubo formalism. As discussed earlier, the viscosity was estimated from the time integral of the pressure tensor autocorrelation function (see eq 6).

Table 8 compares the viscosity values obtained using OPLS (η^{OPLS}) and GA-GPR (η^{GA-GPR}) force fields against the experimental viscosities ($\eta^{Exp.}$) for different temperatures. The high-temperature viscosity obtained using OPLS parameters is tabulated in Table S5. The results highlight significant differences between the two force fields, particularly their ability to predict viscosity across various sulfones and temperatures accurately. The GA-GPR force field closely matches the experimental viscosities across the entire temperature range for SL. At 303 K, the GA-GPR predicted viscosity of 10.29 cP is in excellent agreement with the experimental values, which range from 10.05 to 10.40 cP. In contrast, the OPLS force field significantly overpredicts the viscosity at this temperature, with a value of 20.47 cP, almost twice the experimental value. GA-GPR continues to perform well at higher temperatures and aligns well with the experimental reference. The previously reported viscosity at 303, 323, and 373 K had a maximum deviation of 10.02% at room temperature and 20.23% at high temperature.[26]

No experimental reference is available for EMS, but a comparison between OPLS and GA-GPR reveals that GA-GPR consistently predicts significantly lower viscosities. At 303 K, GA-GPR predicts a viscosity of 2.08 cP, while OPLS overestimates it with a value of 5.80 cP. This trend continues at 323 and 373 K, demonstrating GA-GPR parameters' ability to represent molecular transport in the bulk phase accurately. The GA-GPR force field also shows excellent agreement with experimental values for MSL. At 303 K, the predicted viscosity is 10.09 cP, which matches well with the experimental data ranging from 10.09 to 10.14 cP. However, OPLS significantly overpredicts the viscosity at this temperature, with a value of 21.12 cP, similar to the case of SL. At higher temperatures, GA-GPR predicted viscosities agree with the experimental reference, whereas OPLS parameters continue to overestimate. For DMS, GA-GPR predictions are once again closer to the experimental measurements than OPLS. At 388 K, the GA-GPR predicted viscosity is 1.25 cP, which closely matches the experimental value of 1.285 cP. Similarly, at 393 and 398 K, the GA-GPR value aligns well with the experimental value, whereas OPLS underpredicts it throughout the temperature range.

Overall, the GA-GPR force field outperforms the OPLS force field in accurately predicting the viscosities of sulfone molecules at different temperatures (see Figure S2). The GA-GPR model provides viscosity estimates that are in excellent agreement with experimental data, with deviations consistently within acceptable limits. In contrast, OPLS consistently overpredicts the viscosity values, particularly for cyclic sulfones such as SL and MSL. This overestimation can be attributed to the overestimated electrostatic interactions and incorrect representation of Lennard-Jones parameters in the OPLS force field. In comparison, the GA-GPR model uses more accurate DDEC atomic charges and optimized LJ parameters, resulting in a better representation of intermolecular interactions and the dynamic behavior of the system.

3.3.4. Velocity Autocorrelation Function. To further understand the transport properties of sulfone molecules, we

Table 8. Comparison of Viscosity (cP) Obtained from Classical Molecular Dynamics Simulations Employing OPLS (η^{OPLS}) and GA-GPR (η^{GA-GPR}) Force Fields, with Experimental ($\eta^{Exp.}$) Viscosities at Different Temperatures

	303 K			323 K		373 K	
sulfones	$\eta^{Exp.}$	η^{OPLS}	η^{GA-GPR}	$\eta^{Exp.}$	η^{GA-GPR}	$\eta^{Exp.}$	η^{GA-GPR}
SL	10.284,[22] 10.29[52] 10.228,[54] 10.05[51] 10.3,[50] 10.074[56] 10.401[49]	20.47 ± 0.04[26]	10.29 ± 0.01	6.312[22,53]	6.35 ± 0.04	2.57[22]	2.77 ± 0.03
EMS		5.80 ± 0.55	2.08 ± 0.002		1.56 ± 0.002		0.72 ± 0.001
MSL	10.09[22] 10.14[53]	21.12 ± 0.27	10.09 ± 0.03	5.882,[22] 5.88[53]	5.87 ± 0.03	2.289[22]	2.35 ± 0.02

	388 K		393 K			398 K	
sulfones	$\eta^{Exp.}$	η^{GA-GPR}	$\eta^{Exp.}$	η^{OPLS}	η^{GA-GPR}	$\eta^{Exp.}$	η^{GA-GPR}
DMS	1.285[22]	1.25 ± 0.002	1.208[22]	1.1 ± 0.001	1.23 ± 0.002	1.14[22]	1.10 ± 0.00

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

examine the velocity autocorrelation function (VACF) obtained from the GA-GPR simulated MD trajectories. The VACF provides insight into how a molecule's velocity changes over time, thereby shedding light on the dynamics of sulfone molecules in the liquid phase. Figure 6 illustrates the VACF for

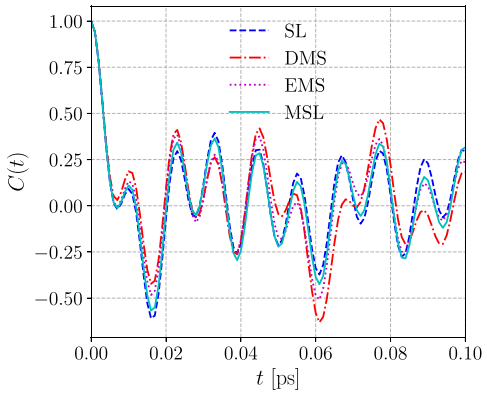


Figure 6. Normalized center of mass velocity autocorrelation functions in sulfolane molecules obtained from simulations employing the optimized force field (GA-GPR) parameters.

different sulfones. As time progresses, the decay in the VACF indicates the loss of correlation due to molecular collisions that alter the velocity. Negative correlations in the VACF, as shown in Figure 6, indicate a reversal in velocity direction following collisions. The well depth of the first minimum in the VACF is particularly significant, as it quantifies the extent of the caging effect experienced by the molecules. In dense liquids, neighboring molecules create a cage-like environment that effectively traps a given molecule, reducing its mobility. This caging effect leads to decreased diffusivity and increased viscosity.

The trends in the VACF are consistent with the viscosity results discussed previously. SL exhibits the deepest well in the VACF, reflecting the most substantial caging effect among the studied sulfones, corresponding to the highest observed viscosity. MSL also shows a prominent caging effect, although to a lesser extent than SL, aligning with its relatively lower viscosity. For EMS and DMS, the shallower wells in the VACF reflect weaker caging effects, which translate into lower viscosities. The observed order of the caging effect, SL > MSL > EMS > DMS, matches the viscosity order, SL > MSL > EMS > DMS, indicating that the extent of molecular trapping directly impacts the dynamics and viscosity of the system.

3.3.5. Self-Diffusion Coefficient. Table 9 presents the self-diffusion coefficients for the four sulfone molecules, calculated from the diffusive regime observed in the mean square displacement (MSD) versus the time plot extracted from *NVT* trajectories. The diffusion coefficients were determined by analyzing the linear portion of the log−log MSD vs time plot, representing the diffusive regime (see Figure S3). The values reported were averaged from five independent trajectories of 1000 ps each, ensuring the reliability of the computed diffusion coefficients. The extent of diffusion is influenced by the caging effect, which was examined using the velocity autocorrelation function. Molecules that are trapped more by their neighbors are expected to show a reduced diffusion coefficient due to restricted mobility.

Table 9. Comparison of Self-Diffusion Coefficients (in Units of 10^{-10} m^2/s) for Sulfone Molecules, Calculated from the Diffusive Regime Observed in the MSD vs Time Plots Obtained from *NVT* Trajectories Using OPLS and GA-GPR Force Fields

sulfones	OPLS	GA-GPR
SL	0.13 ± 0.002	0.41 ± 0.02
EMS	2.03 ± 0.05	5.56 ± 0.16
MSL	0.06 ± 0.003	1.41 ± 0.02
DMS	15.15 ± 0.32	12.75 ± 0.20

Upon comparing the diffusion coefficients obtained from GA-GPR and OPLS simulations, it is evident that the GA-GPR model consistently predicts higher diffusion coefficients than OPLS, except for DMS. For SL, the diffusion coefficient obtained using OPLS is 0.13×10^{-10} m^2/s, whereas GA-GPR predicts a significantly higher value of 0.41×10^{-10} m^2/s. This notable difference suggests that the OPLS force field overestimates the strength of intermolecular interactions, resulting in excessive caging and restricted mobility. In contrast, the GA-GPR force field more accurately represents the intermolecular forces, allowing for a more realistic depiction of molecular mobility in the bulk. A similar trend is observed for EMS and MSL—the diffusion coefficient obtained from OPLS is significantly lower than the GA-GPR computed values. On the contrary, for DMS, the trend deviates slightly, with the GA-GPR model predicting a diffusion coefficient of 12.75×10^{-10} m^2/s, which is lower than the OPLS value of 15.15×10^{-10} m^2/s. This difference may arise from differences in how the two force fields represent the interactions specific to the linear structure of DMS. It is important to note that, in the case of DMS, the viscosity predicted by the OPLS force field was smaller compared to GA-GPR, aligning well with the trend seen in the diffusion coefficient.

The observed order of diffusion coefficients, SL < MSL < EMS < DMS, aligns well with the order of the caging effect, SL > MSL > EMS > DMS, as determined from VACF calculations. This consistency demonstrates that the GA-GPR model not only provides more accurate diffusion coefficients compared to OPLS but also effectively captures the underlying dynamics that govern molecular mobility. The OPLS force field, by overestimating intermolecular interactions, leads to reduced diffusion coefficients, particularly for cyclic sulfones. The GA-GPR model's ability to accurately represent diffusion properties highlights its suitability for modeling the transport behavior of sulfone-based electrolyte systems, reinforcing its transferability and reliability across different sulfone structures.

3.3.6. Surface Tension. Surface tension is a crucial property in understanding the interfacial behavior of sulfone molecules, particularly in the context of their use in electrolyte systems, where interactions at the liquid−vapor or liquid−solid interface can significantly influence performance. As discussed earlier, surface tension was computed using classical MD simulations with an 8 ns NVT run. The calculation was based on the difference between the pressure tensor components parallel and perpendicular to the interface, providing insight into the cohesive forces at the molecular level. The running average surface tension profiles in the sulfones are plotted in Figures S4 and S5. These profiles clearly show that the computed surface tension is well converged within the 1 ns.

Table 10 presents the surface tension values obtained using the OPLS and GA-GPR force fields alongside the experimental

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 10. Surface Tension (mN/m) Calculated Using Classical Molecular Dynamics Run with OPLS (γ^{OPLS}) and GA-GPR (γ^{GA-GPR}) Force Fields[a]

sulfones	$\gamma^{Exp.}$	γ^{OPLS}	γ^{GA-GPR}
SL	47.95[51]	55.25	45.60
EMS		39.83	25.51
MSL		45.27	35.64
DMS		30.54	34.31

[a]Experimental value ($\gamma^{Exp.}$) for SL is provided for validation, and GA-GPR predictions for other sulfones are considered reliable based on this validation.

reference for SL. For SL, the GA-GPR predicted surface tension is 45.60 mN/m, which is in close agreement with the experimental value of 47.95 mN/m. This excellent agreement validates the accuracy of the GA-GPR model in predicting surface tension. In contrast, the OPLS force field overestimates the surface tension, yielding a value of 55.25 mN/m, which indicates stronger intermolecular cohesion than experimentally observed. Previously reported surface tension using classical MD simulation was 53.43[14] and 34.62 mN/m.[26] No experimental reference values are available for the other sulfones, EMS, MSL, and DMS. However, based on the strong validation of GA-GPR for SL, it is reasonable to infer that the GA-GPR predictions for these molecules are reliable. The GA-GPR model consistently predicts lower surface tension values than OPLS. The consistent overestimation by OPLS across all sulfones can be attributed to the overestimation of the intermolecular forces, leading to artificially high surface tension values. In contrast, the GA-GPR force field, with its more refined parametrization, provides a better match with experimental data and a more realistic depiction of the interfacial behavior of sulfone molecules.

To further understand the interfacial properties of these sulfone-based systems, we investigated the liquid−vapor interface structure by analyzing the density profiles of different atom types. As shown in Figure 7, these density profiles provide insight into the atomic arrangement at the interface and bulk regions. As expected, the density profiles are flat in the bulk liquid regions, indicating that the two-phase morphology is well-equilibrated. The placement of heteroatoms, including sulfur (S), oxygen (O), and carbon (C), was further analyzed to gain more insights into the interface structure. The observed density profiles reveal a distinct arrangement of atoms in sulfone molecules at the liquid−vapor interface, which can be attributed to differences in polarity and molecular interactions. Carbon atoms, being less electronegative than sulfur and oxygen, prefer to occupy the interface, as their lower polarity makes them more surface-active. In contrast, sulfur atoms are more electronegative and polarizable, making them interact more strongly with neighboring molecules in the bulk, resulting in their greater presence away from the interface. Oxygen atoms, with a higher electronegativity than carbon but smaller size and polarizability compared to sulfur, occupy an intermediate position between the interface and the bulk, balancing their interactions. In the case of DMS, the symmetric molecular structure leads to an equal placement of carbon and oxygen atoms at the interface, while sulfur atoms remain in the bulk. This arrangement is due to the similar polarity of carbon and oxygen, which results in both atoms being surface-active. At the same time, sulfur prefers the bulk to maximize its interactions with neighboring molecules. The overall distribution of atoms reflects the

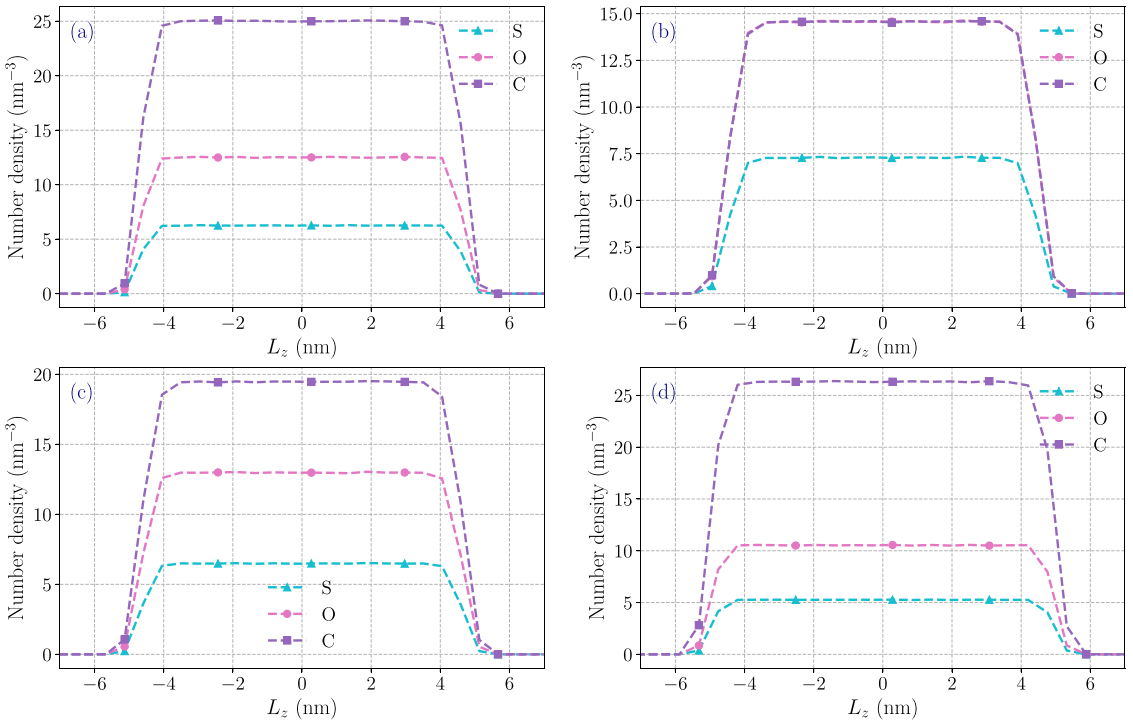


Figure 7. Density profiles of atom types in sulfolane molecules obtained from simulations employing the optimized force field (GA-GPR) parameters: (a) SL, (b) DMS, (c) EMS, and (d) MSL. Similar profiles obtained from simulations employing the OPLS force field are provided in Figure S6.

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

interplay between molecular polarity, electronegativity, and the need to minimize the system's energy.

3.3.7. Comparison of GA-GPR with Bayesian Optimization. To further establish the novelty and robustness of the GA-GPR framework, we performed a direct comparison with Bayesian Optimization (BO), a widely used method for force field optimization.[39] Both methods were applied to the ethyl methyl sulfone (EMS) system, starting from the same initial parameters derived from the OPLS force field. The fitness function was defined as the cumulative error in reproducing density and radial distribution functions. For Bayesian Optimization, the q-Expected Improvement (q-EI) acquisition function was employed to select multiple sampling points per iteration. Significant hyperparameter tuning was performed for BO to ensure a fair comparison. As shown in Figure S7a, the GA-GPR framework demonstrated significantly faster convergence compared to Bayesian Optimization. Specifically, out of the first 70 LJ parameter sets generated, 38 from GA-GPR achieved a fitness below 200, whereas only 8 from Bayesian Optimization met the same criterion. This result highlights the stability and efficiency of GA-GPR in balancing exploration and exploitation of the parameter space.

Despite extensive tuning, Bayesian Optimization often explored regions of high uncertainty, frequently generating unphysical LJ parameter sets that resulted in failed MD simulations or unreasonably high fitness values. Strict bounds (±25% deviation from the initial OPLS parameters) were imposed on the LJ parameters to mitigate this. However, despite these constraints, BO prioritized exploration near boundary values, leading to parameter sets with unacceptable fitness values. In contrast, GA-GPR exhibited a more consistent progression toward optimal parameter sets, effectively avoiding unphysical regions while maintaining efficient parameter space exploration.

To further validate the robustness of GA-GPR, we compared the RDFs obtained from the best LJ parameter sets identified by both methods (Figures 4c and S7b,c). While both methods achieved low RDF errors, Bayesian Optimization produced unwanted peaks in the RDF plots (Figure S7b,c), indicative of unphysical parametrization. These artifacts were absent in the RDFs generated by GA-GPR, underscoring its ability to produce physically meaningful force field parameters. The GA-GPR framework also requires minimal manual tuning of hyperparameters compared to Bayesian Optimization, which necessitated significant effort to optimize the acquisition function and impose bounds to prevent unphysical sampling. These results collectively demonstrate the novelty of GA-GPR in achieving faster convergence, greater stability, and physically meaningful parameter sets compared to state-of-the-art methods.

4. CONCLUSIONS

This work presented an efficient and accurate approach to developing optimized force field parameters for sulfone molecules using a combined genetic algorithm and Gaussian process regression (GA-GPR) model. Our active learning-based optimization strategy significantly outperformed previous methods regarding convergence speed and required parameter sets, achieving optimized parameters within 12 iterations (except for SL, which took only a single iteration) and using only 300 data points. This efficiency demonstrates the strength of the GA-GPR workflow in capturing the essential intermolecular interactions of complex systems with minimal computational cost. In addition, our method can handle properties challenging for automatic differentiation-based approaches, such as phase equilibration or long-time dynamics. While density and evaporation enthalpy are not examples of such properties, our framework is designed to be extensible to more complex scenarios.

The GA-GPR-derived force field was thoroughly validated against various experimental and reference data, including density, dipole moment, viscosity, diffusion coefficients, and surface tension. The GA-GPR force field exhibited superior accuracy in all cases over the OPLS force field, particularly in describing bulk and interfacial properties. The predicted viscosity showed excellent agreement with experimental measurements, with the GA-GPR force field effectively capturing the trends in molecular mobility and caging effects. The surface tension results further demonstrated the transferability of the GA-GPR parameters, while the accurate representation of density profiles highlighted its ability to describe the structural arrangement of atoms at the interface and in bulk.

Our approach provides a reliable and transferable force field for sulfone molecules, with broad applicability to studies of electrolytes and other complex systems. The combination of genetic algorithms and machine learning offers a robust framework for force field optimization, reducing manual intervention and providing accurate predictions for both bulk and interfacial properties. This study establishes a foundation for further exploration of machine learning-driven force field development, which can be extended to other classes of molecules to enhance the accuracy and efficiency of molecular simulations.

■ ASSOCIATED CONTENT

Data Availability Statement

Codes used to perform the GA-GPR training, relevant simulations, and all generated parameter sets are available at https://github.com/cocokane/LJ_paramopt_framework.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.5c00061.

Running integrals, comparison of density and viscosity against the experiment, MSD, density profiles, and tables of optimized nonbonded parameters (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Anirban Mondal − *Department of Chemistry, Indian Institute of Technology Gandhinagar, Gandhinagar, Gujarat 382355, India;* orcid.org/0000-0003-3029-8840; Email: amondal@iitgn.ac.in

Authors

Yati − *Department of Chemistry, Indian Institute of Technology Gandhinagar, Gandhinagar, Gujarat 382355, India*

Yash Kokane − *Department of Materials Engineering, Indian Institute of Technology Gandhinagar, Gandhinagar, Gujarat 382355, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.5c00061

Author Contributions

A.M. conceived the problem. Yati conducted all the simulations. Y.K. contributed to developing the surrogate model. Yati and A.M. analyzed the results and prepared the draft.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge the Indian Institute of Technology Gandhinagar, India, for providing research facilities and financial support. Yati and A.M. thank PARAM Ananta for computational resources.

■ REFERENCES

(1) Yao, N.; Chen, X.; Fu, Z.-H.; Zhang, Q. Applying classical, ab initio, and machine-learning molecular dynamics simulations to the liquid electrolyte for rechargeable batteries. *Chem. Rev.* **2022**, *122*, 10970−11021.
(2) Manthiram, A.; Yu, X.; Wang, S. Lithium battery chemistries enabled by solid-state electrolytes. *Nat. Rev. Mater.* **2017**, *2*, 16103.
(3) Chu, S.; Cui, Y.; Liu, N. The path towards sustainable energy. *Nat. Mater.* **2017**, *16*, 16−22.
(4) Xu, K. Nonaqueous liquid electrolytes for lithium-based rechargeable batteries. *Chem. Rev.* **2004**, *104*, 4303−4418.
(5) Zhu, G.; Wen, K.; Lv, W.; Zhou, X.; Liang, Y.; Yang, F.; Chen, Z.; Zou, M.; Li, J.; Zhang, Y.; et al. Materials insights into low-temperature performances of lithium-ion batteries. *J. Power Sources* **2015**, *300*, 29−40.
(6) Xu, K.; Angell, C. A. Sulfone-based electrolytes for lithium-ion batteries. *J. Electrochem. Soc.* **2002**, *149*, A920.
(7) Sun, X.; Angell, C. A. Doped sulfone electrolytes for high voltage Li-ion cell applications. *Electrochem. Commun.* **2009**, *11*, 1418−1421.
(8) Xue, L.; Ueno, K.; Lee, S.-Y.; Angell, C. A. Enhanced performance of sulfone-based electrolytes at lithium ion battery electrodes, including the $LiNi_{0.5}Mn_{1.5}O_4$ high voltage cathode. *J. Power Sources* **2014**, *262*, 123−128.
(9) Abouimrane, A.; Belharouak, I.; Amine, K. Sulfone-based electrolytes for high-voltage Li-ion batteries. *Electrochem. Commun.* **2009**, *11*, 1073−1076.
(10) Ren, X.; Chen, S.; Lee, H.; Mei, D.; Engelhard, M. H.; Burton, S. D.; Zhao, W.; Zheng, J.; Li, Q.; Ding, M. S.; et al. Localized high-concentration sulfone electrolytes for high-efficiency lithium-metal batteries. *Chem.* **2018**, *4*, 1877−1892.
(11) Ugata, Y.; Tatara, R.; Mandai, T.; Ueno, K.; Watanabe, M.; Dokko, K. Understanding the reductive decomposition of highly concentrated Li Salt/Sulfolane Electrolytes during Li Deposition and Dissolution. *ACS Appl. Energy Mater.* **2021**, *4*, 1851−1859.
(12) Alvarado, J.; Schroeder, M. A.; Zhang, M.; Borodin, O.; Gobrogge, E.; Olguin, M.; Ding, M. S.; Gobet, M.; Greenbaum, S.; Meng, Y. S.; et al. A carbonate-free, sulfone-based electrolyte for high-voltage Li-ion batteries. *Mater. Today* **2018**, *21*, 341−353.
(13) Ko, S.; Yamada, Y.; Yamada, A. An overlooked issue for high-voltage Li-ion batteries: Suppressing the intercalation of anions into conductive carbon. *Joule* **2021**, *5*, 998−1009.
(14) Dokko, K.; Watanabe, D.; Ugata, Y.; Thomas, M. L.; Tsuzuki, S.; Shinoda, W.; Hashimoto, K.; Ueno, K.; Umebayashi, Y.; Watanabe, M. Direct evidence for Li ion hopping conduction in highly concentrated sulfolane-based liquid electrolytes. *J. Phys. Chem. B* **2018**, *122*, 10736−10745.
(15) Kolosnitsyn, V.; Sheina, L.; Mochalov, S. E. Physicochemical and electrochemical properties of sulfolane solutions of lithium salts. *Russ. J. Electrochem.* **2008**, *44*, 575−578.
(16) Flamme, B.; Haddad, M.; Phansavath, P.; Ratovelomanana-Vidal, V.; Chagnes, A. Anodic stability of new sulfone-based electrolytes for lithium-ion batteries. *ChemElectroChem.* **2018**, *5*, 2279−2287.
(17) Hirata, K.; Morita, Y.; Kawase, T.; Sumida, Y. Electrochemical performance of an ethylene carbonate-free electrolyte based on lithium bis (fluorosulfonyl) imide and sulfolane. *J. Power Sources* **2018**, *395*, 163−170.
(18) Xu, K.; Angell, C. High anodic stability of a new electrolyte solvent: Unsymmetric noncyclic aliphatic sulfone. *J. Electrochem. Soc.* **1998**, *145*, L70.
(19) Xu, W.; Shusterman, A. J.; Videa, M.; Velikov, V.; Marzke, R.; Angell, C. A. Structures of orthoborate anions and physical properties of their lithium salt nonaqueous solutions. *J. Electrochem. Soc.* **2003**, *150*, E74.
(20) Lee, S.-Y.; Ueno, K.; Angell, C. A. Lithium salt solutions in mixed sulfone and sulfone-carbonate solvents: A Walden plot analysis of the maximally conductive compositions. *J. Phys. Chem. C* **2012**, *116*, 23915−23920.
(21) Ugata, Y.; Chen, Y.; Sasagawa, S.; Ueno, K.; Watanabe, M.; Mita, H.; Shimura, J.; Nagamine, M.; Dokko, K. Eutectic Electrolytes Composed of $LiN(SO_2F)_2$ and Sulfones for Li-Ion Batteries. *J. Phys. Chem. C* **2022**, *126*, 10024−10034.
(22) Casteel, J. F.; Sears, P. G. Dielectric constants, viscosities, and related physical properties of 10 liquid sulfoxides and sulfones at several temperatures. *J. Chem. Eng. Data* **1974**, *19*, 196−200.
(23) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(24) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.
(25) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(26) Mukherji, S.; Avula, N. V.; Balasubramanian, S. Refined force field for liquid sulfolane with particular emphasis to its transport characteristics. *ACS omega* **2020**, *5*, 28285−28295.
(27) Mondal, A.; Balasubramanian, S. Quantitative Prediction of Physical Properties of Imidazolium Based Room Temperature Ionic Liquids through Determination of Condensed Phase Site Charges: A Refined Force Field. *J. Phys. Chem. B* **2014**, *118*, 3409−3422.
(28) Mondal, A.; Balasubramanian, S. A Refined All-Atom Potential for Imidazolium-Based Room Temperature Ionic Liquids: Acetate, Dicyanamide, and Thiocyanate Anions. *J. Phys. Chem. B* **2015**, *119*, 11041−11051.
(29) Wang, N.; Carlozo, M. N.; Marin-Rimoldi, E.; Befort, B. J.; Dowling, A. W.; Maginn, E. J. Machine learning-enabled development of accurate force fields for refrigerants. *J. Chem. Theory Comput.* **2023**, *19*, 4546−4558.
(30) Mondal, A.; Young, J. M.; Barckholtz, T. A.; Kiss, G.; Koziol, L.; Panagiotopoulos, A. Z. Genetic algorithm driven force field parameterization for molten alkali-metal carbonate and hydroxide salts. *J. Chem. Theory Comput.* **2020**, *16*, 5736−5746.
(31) Madin, O. C.; Shirts, M. R. Using physical property surrogate models to perform accelerated multi-fidelity optimization of force field parameters. *Digit. Discovery* **2023**, *2*, 828−847.
(32) Yu, Y.; Kramer, A.; Venable, R. M.; Simmonett, A. C.; MacKerell, A. D., Jr; Klauda, J. B.; Pastor, R. W.; Brooks, B. R. Semi-automated optimization of the CHARMM36 lipid force field to include explicit treatment of long-range dispersion. *J. Chem. Theory Comput.* **2021**, *17*, 1562−1580.
(33) Befort, B. J.; DeFever, R. S.; Tow, G. M.; Dowling, A. W.; Maginn, E. J. Machine learning directed optimization of classical molecular modeling force fields. *J. Chem. Inf. Model.* **2021**, *61*, 4400−4414.
(34) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; et al. Escaping atom types in force fields using direct chemical perception. *J. Chem. Theory Comput.* **2018**, *14*, 6076−6092.

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(35) Boothroyd, S.; Wang, L.-P.; Mobley, D. L.; Chodera, J. D.; Shirts, M. R. Open force field evaluator: An automated, efficient, and scalable framework for the estimation of physical properties from molecular simulation. *J. Chem. Theory Comput.* **2022**, *18*, 3566−3576.
(36) Horton, J. T.; Boothroyd, S.; Wagner, J.; Mitchell, J. A.; Gokey, T.; Dotson, D. L.; Behara, P. K.; Ramaswamy, V. K.; Mackey, M.; Chodera, J. D.; et al. Open force field BespokeFit: automating bespoke torsion parametrization at scale. *J. Chem. Inf. Model.* **2022**, *62*, 5622−5633.
(37) Chen, Y.; Huang, Q.; Liu, T.-H.; Yang, R.; Qian, X. Modeling solvation dynamics of transition metal redox ion through on-the-fly multi-objective Bayesian-optimized force field. *J. Chem. Phys.* **2024**, *161*, 124111.
(38) Befort, B. J.; DeFever, R. S.; Maginn, E. J.; Dowling, A. W.*Computer Aided Chemical Engineering*; Elsevier: 2022; Vol. *49*, pp 1249−1254.
(39) Liu, H.; Li, Y.; Fu, Z.; Li, K.; Bauchy, M. Exploring the landscape of Buckingham potentials for silica by machine learning: Soft vs hard interatomic forcefields. *J. Chem. Phys.* **2020**, *152*, No. 051101.
(40) Liu, H.; Fu, Z.; Li, Y.; Sabri, N. F. A.; Bauchy, M. Machine learning forcefield for silicate glasses. *arXiv preprint arXiv* **2019**, 1−18.
(41) Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G. Gaussian process regression for materials and molecules. *Chem. Rev.* **2021**, *121*, 10073−10141.
(42) Li, Y.; Li, H.; Pickard, F. C., IV; Narayanan, B.; Sen, F. G.; Chan, M. K.; Sankaranarayanan, S. K.; Brooks, B. R.; Roux, B. Machine learning force field parameters from ab initio data. *J. Chem. Theory Comput.* **2017**, *13*, 4492−4503.
(43) Liu, H.; Fu, Z.; Li, Y.; Sabri, N. F. A.; Bauchy, M. Balance between accuracy and simplicity in empirical forcefields for glass modeling: insights from machine learning. *J. Non-Cryst. Solids* **2019**, *515*, 133−142.
(44) Liu, H.; Fu, Z.; Li, Y.; Sabri, N. F. A.; Bauchy, M. Parameterization of empirical forcefields for glassy silica using machine learning. *MRS Commun.* **2019**, *9*, 593−599.
(45) Ge, Y.; Wang, X.; Zhu, Q.; Yang, Y.; Dong, H.; Ma, J. Machine Learning-Guided Adaptive Parametrization for Coupling Terms in a Mixed United-Atom/Coarse-Grained Model for Diphenylalanine Self-Assembly in Aqueous Ionic Liquids. *J. Chem. Theory Comput.* **2023**, *19*, 6718−6732.
(46) Limas, N. G.; Manz, T. A. Introducing DDEC6 atomic population analysis: part 4. Efficient parallel computation of net atomic charges, atomic spin moments, bond orders, and more. *RSC Adv.* **2018**, *8*, 2678−2707.
(47) Manz, T. A. Introducing DDEC6 atomic population analysis: part 3. Comprehensive method to compute bond orders. *RSC Adv.* **2017**, *7*, 45552−45581.
(48) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09 Revision E.01*; Gaussian Inc.: Wallingford CT, 2009.
(49) Mesquita, F. M. R.; Feitosa, F. X.; Aznar, M.; de Sant'Ana, H. B.; Santiago-Aguiar, R. S. Density, viscosities, and excess properties for binary mixtures of sulfolane+ alcohols and sulfolane+ glycols at different temperatures. *J. Chem. Eng. Data* **2014**, *59*, 2196−2206.
(50) Jalili, A. H.; Shokouhi, M.; Samani, F.; Hosseini-Jenab, M. Measuring the solubility of CO_2 and H_2S in sulfolane and the density and viscosity of saturated liquid binary mixtures of (sulfolane + CO_2) and (sulfolane + H_2S). *J. Chem. Thermodyn.* **2015**, *85*, 13−25.
(51) Kelayeh, S. A.; Jalili, A. H.; Ghotbi, C.; Hosseini-Jenab, M.; Taghikhani, V. Densities, viscosities, and surface tensions of aqueous mixtures of sulfolane+ triethanolamine and sulfolane+ diisopropanolamine. *J. Chem. Eng. Data* **2011**, *56*, 4317−4324.
(52) Fernandez-Prini, R.; Prue, J. Conductance measurements on solutions of salts in sulpholane and their interpretation. *Trans. Faraday Soc.* **1966**, *62*, 1257−1264.
(53) Vaughn, J. W.; Hawkins, C. F. Physical Properties of Tetrahydrothiophene-1, 1-Dioxide and 3-Methyltetrahydrothiophene-1, 1-Dioxide. *J. Chem. Eng. Data* **1964**, *9*, 140−142.
(54) Águila-Hernández, J.; Trejo, A.; García-Flores, B. E.; Molnar, R. Viscometric and volumetric behaviour of binary mixtures of sulfolane and N-methylpyrrolidone with monoethanolamine and diethanolamine in the range 303−373 K. *Fluid Phase Equilib.* **2008**, *267*, 172−180.
(55) Murrieta-Guevara, F.; Rebolledo-Libreros, E.; Trejo, A. Gas solubility of carbon dioxide and hydrogen sulfide in mixtures of sulfolane with monoethanolamine. *Fluid Phase Equilib.* **1993**, *86*, 225−231.
(56) Yang, C.; Yu, W.; Ma, P. Densities and viscosities of binary mixtures of ethylbenzene+ n-methyl-2-pyrrolidone, ethylbenzene+ sulfolane, and styrene+ octane from (303.15 to 353.15) K and atmospheric pressure. *J. Chem. Eng. Data* **2005**, *50*, 1197−1203.
(57) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157−2164.
(58) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. cp2k: atomistic simulations of condensed matter systems. *WIREs Comput. Mol. Sci.* **2014**, *4*, 15−25.
(59) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(60) 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. *J. Chem. Phys.* **2010**, *132*, 154104.
(61) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703.
(62) Hartwigsen, C.; Gœdecker, S.; Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641.
(63) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177−4189.
(64) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé−Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97*, 2635−2643.
(65) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(66) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
(67) Berendsen, H. J.; Postma, J. v.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(68) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(69) Fortin, F.-A.; De Rainville, F.-M.; Gardner, M.-A.; Parizeau, M.; Gagné, C. DEAP: Evolutionary Algorithms Made Easy. *J. Mach. Learn. Res.* **2012**, *13*, 2171−2175.

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(70) Harada, M.; Yamanaka, A.; Tanigaki, M.; Tada, Y. Mass and size effects on the transport properties of molten salts. *J. Chem. Phys.* **1982**, *76*, 1550−1556.
(71) Daivis, P. J.; Evans, D. J. Comparison of constant pressure and constant volume nonequilibrium simulations of sheared model decane. *J. Chem. Phys.* **1994**, *100*, 541−547.
(72) Hansen, J.-P.; McDonald, I. R. *Theory of simple liquids: with applications to soft matter*; Academic Press: 2013.
(73) Wang, X.; Li, J.; Yang, L.; Chen, F.; Wang, Y.; Chang, J.; Chen, J.; Feng, W.; Zhang, L.; Yu, K. Dmff: an open-source automatic differentiable platform for molecular force field development and molecular dynamics simulation. *J. Chem. Theory Comput.* **2023**, *19*, 5897−5909.
(74) Hansen, N.; Kraus, P.; Saßmannshausen, H.; Timmerscheidt, T.; van Gunsteren, W. F. An effective force field for molecular dynamics simulations of dimethyl sulfone. *Mol. Phys.* **2011**, *109*, 2593−2605.

https://doi.org/10.1021/acs.jctc.5c00061
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX