THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA Article

Improving Machine Learned Force Fields for Complex Fluids through Enhanced Sampling: A Liquid Crystal Case Study

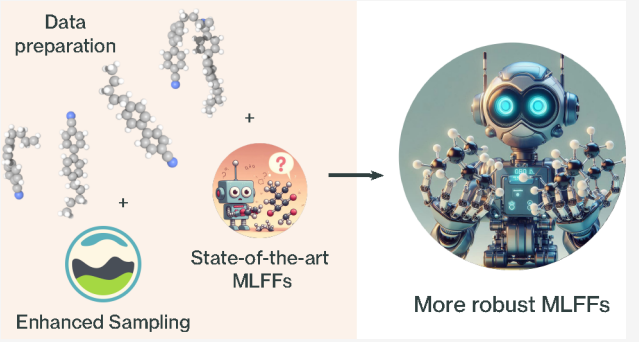
Published as part of The Journal of Physical Chemistry A virtual special issue "Recent Advances in Simulation Software and Force Fields".

Yezhi Jin, Gustavo R. Perez-Lemus, Pablo F. Zubieta Rico,* and Juan J. de Pablo*

Cite This: https://doi.org/10.1021/acs.jpca.4c01546 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Machine learned force fields offer the potential for faster execution times while retaining the accuracy of traditional DFT calculations, making them promising candidates for molecular simulations in cases where reliable classical force fields are not available. Some of the challenges associated with machine learned force fields include simulation stability over extended periods of time and ensuring that the statistical and dynamical properties of the underlying simulated systems are correctly captured. In this work, we propose a systematic training pipeline for such force fields that leads to improved model quality, compared to that achieved by traditional data generation and training approaches. That pipeline relies on the use of enhanced sampling techniques, and it is demonstrated here in the context of a liquid crystal, which exemplifies many of the challenges that are encountered in fluids and materials with complex free energy landscapes. Our results indicate that, whereas the majority of traditional machine learned force field training approaches lead to molecular dynamics simulations that are only stable over hundred-picosecond trajectories, our approach allows for stable simulations over tens of nanoseconds for organic molecular systems comprising thousands of atoms.



INTRODUCTION

Molecular simulations play a central role in a wide range of fields, ranging from materials science to molecular biology. It is widely appreciated that a simulation can generate a realistic representation of a given physical process phenomenon only if the underlying force field is sufficiently accurate and if the corresponding simulation algorithm is capable of sampling all relevant regions of phase space. Classical pairwise additive force fields offer computational efficiency, but parameters are not always available for a system of interest.[1,2] Quantum mechanical calculations offer an alternative in such cases,[3,4] but they are computationally demanding, hindering their applicability to complex systems over long length and time scales. The challenge that arises then is that of accurately describing the physical and chemical interactions in such systems, while achieving a reasonable computational performance.[5−8] This trade-off motivates ongoing efforts to devise force fields that provide a balance between tractability and accuracy.

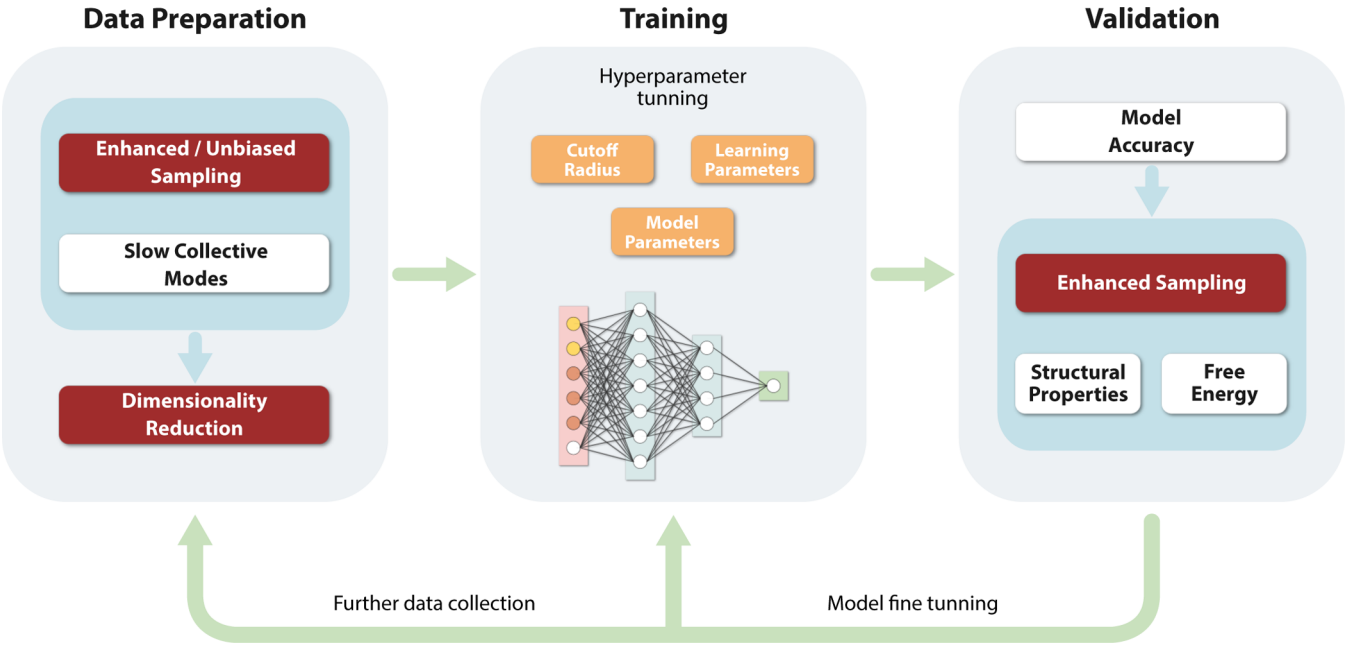
One promising approach to address this challenge is provided by machine learned force fields (MLFFs).[7,9−13] These force fields leverage the advantages of accurate quantum mechanical calculations to capture molecular interactions, while also offering the potential to conduct long time and length scale simulations and yielding new insights into complex physicochemical phenomena.[14−17]

Several studies have sought to validate MLFFs through a variety of approaches.[18,19] The most common validation strategy, inherited from the machine learning field, consists of assessing their predictive accuracy for energies and forces.[10,20] Other assessment approaches try to also evaluate their ability to compute radial and angular distribution functions,[11,21] as well as other physicochemical properties such as the stability of crystalline structures[15] or how well they can reproduce melting and stress−strain behavior.[22] More demanding calculations of thermodynamic properties such as free energies have not been performed within the context of standard MLFF training approaches.

For all of the aforementioned studies, the scaling of simulations with number of atoms has been explored in the

Received: March 8, 2024
Revised: July 23, 2024
Accepted: July 26, 2024

Scheme 1. Overview of Our Machine Learning Force Field Learning Pipeline[a]



[a]The first and last steps involve performing enhanced sampling simulations, which represents one of the key strategies proposed to improve the overall robustness of MLFFs.

context of inorganic materials; for organic molecules, benchmarks have largely focused on the MLFFs' performance for data sets consisting of single molecules.[7,23−27] Moreover, the stability of available MLFFs has generally been addressed for relatively short molecular dynamics simulations,[20] with most MLFF works to date reporting simulations of hundreds of picoseconds. For the particular case of organic systems, an exception is provided by a recent example which reported MD simulations on the order of a nanosecond.[28]

The lack of long time simulations using MLFFs in the field can be attributed to multiple reasons, one being insufficient sampling of high energy states in the training data or, in some cases, the high cost associated with long ab initio MD simulations. For a case-by-case analysis, we refer readers to recent work on the underlying issues.[20]

In this work we introduce a consolidated MLFF training pipeline that incorporates existing ML validation practices along with data generation strategies based on enhanced sampling techniques such as steered MD.[29] Furthermore, we also integrate into our pipeline available methods for the fast computation of free energy surfaces, such as spectral adaptive biasing force method (spectral ABF),[30] thereby enabling a more comprehensive and systematic evaluation of the quality of MLFFs for free energy calculations and ultimately leading to more robust force fields.

We demonstrate the efficacy of this methodology by training a MLFF for 4-cyano-4′-pentylbiphenyl (5CB), a widely studied liquid crystal.[31−36] Our simulations show stability in different simulations of condensed phases of 5CB with thousands of atoms over tens of nanoseconds. Importantly, to facilitate future research efforts, we have made our MLFF training pipeline available as open-source code.

RESULTS AND DISCUSSION

Workflow Outline. We propose a comprehensive training and verification pipeline that incorporates four stages: data preparation, hyperparameter tuning, MLFF training, and quality evaluation. One delineates the workflow associated with these stages. We describe the methodology for each stage in the following subsections using our training process for 5CB MLFF as an example.

Note that, while we have chosen 5CB, a prototypical example of a liquid crystalline material, to showcase our proposed pipeline (see Scheme 1), the overall strategy presented here can be applied to any other system of interest with the appropriate choice of slow order parameter or collective variable. Determination of slow modes in a system is an ongoing effort[37−40] on its own, and it is outside the scope of the present work.

Data Preparation. Most MLFF models rely on descriptors based on atomic coordinates. These descriptors capture the local atomic environment of each atom and, ideally, remain unchanged (are invariant) or transform in the same way (are equivariant) under rotations or translations of the reference frame and permutations of identical atoms.[7,8,11,27,41,42] This characteristic enables an accurate prediction of forces and potential energies associated with every atom and its environment. It is therefore crucial to prepare a representative and comprehensive training data set for the training and refinement of MLFFs. In what follows, we show that direct strategies for generating data may not adequately capture configurations that correspond to regions of high free energy. To overcome this challenge, we propose a data generation strategy based on dimensionality reduction and enhanced sampling.

Unbiased Sampling. We begin our workflow by running unbiased classical MD simulations of the 5CB system. Following existing works,[5,9,11] we prepared seven 5CB systems with varying box sizes to mimic the density variations observed at different temperatures (293 K, 298 K, 300 K, 303 K, 308 K, 327 K, 367 K), consisting of 684 atoms from 18 5CB molecules. These span the nematic and isotropic phases of 5CB. We then perform a canonical ensemble (NVT) simulation on each

https://doi.org/10.1021/acs.jpca.4c01546
J. Phys. Chem. A XXXX, XXX, XXX−XXX

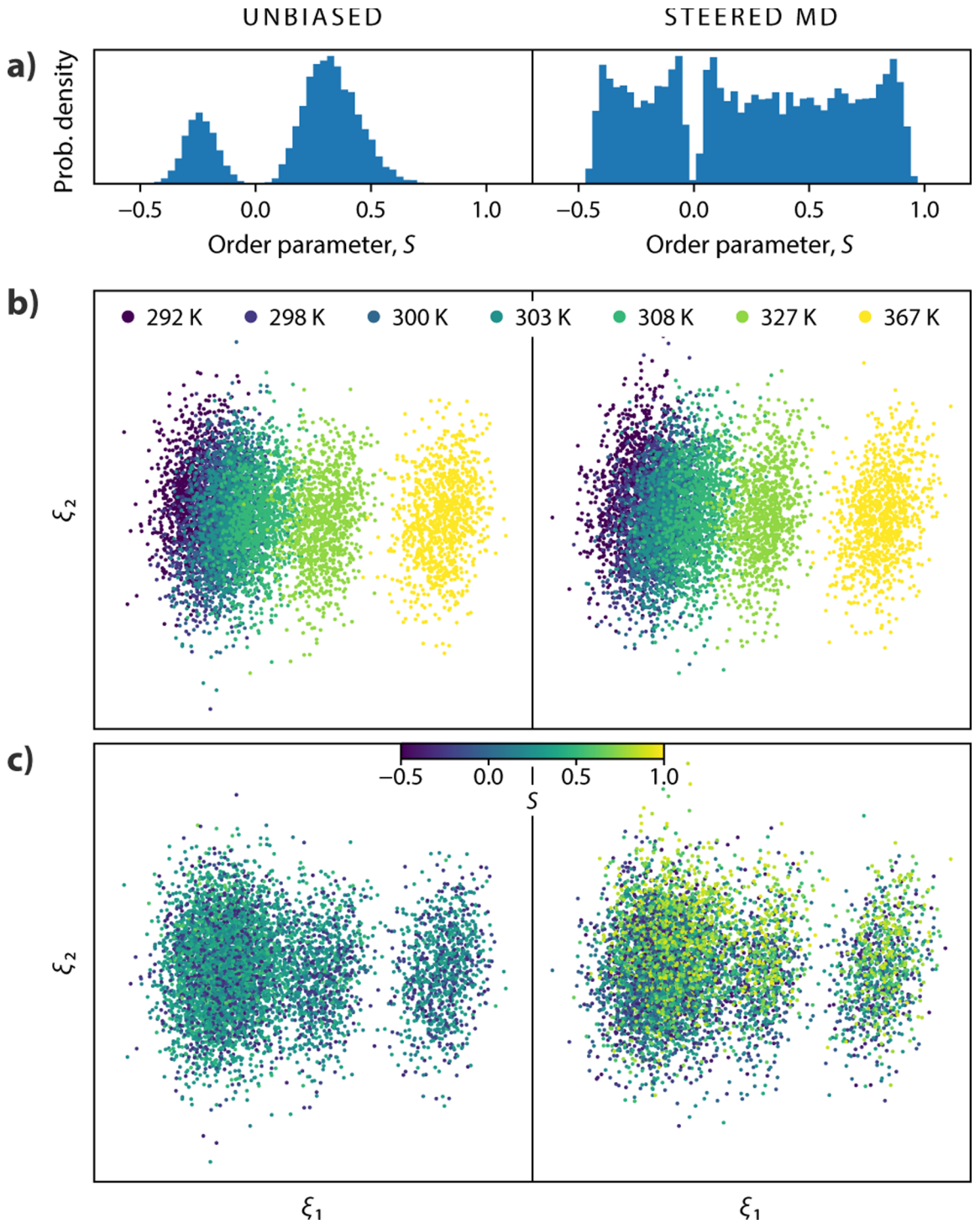


Figure 1. Histogram of the order parameter S collected from snapshots generated by (a, left) unbiased simulations and (a, right) enhanced sampling. Poor sampling of the extreme values of the nematic order parameter, $S \in (-0.5, 0.0, 1.0)$, is a consequence of the finite size of the system (which has only 18 molecules). Visualization of the principal components of the data embedding, colored by (b) the temperature at which the data is sampled and (c) the order parameter. Plots (b, left) and (c, left) are derived from unbiased sampling data, while (b, right) and (c, right) originate from enhanced sampling data.

system and collect 1000 snapshots from each trajectory. For each of the selected snapshots, DFT calculations are then performed to obtain the "ground truth" forces, energies, and stresses.

To evaluate the *representativeness* of our data set, we examine the order parameter, S, of the resulting snapshots. The order parameter S is a particularly important collective variable that describes the relative degree of orientational order in a liquid crystal.[43] Formally, it is defined as the largest eigenvalue of the tensor Q[44]

$$Q = \frac{3}{2}\frac{1}{N}\sum_{i=1}^{N}\left(\hat{u}_i\hat{u}_i - \frac{1}{3}\mathbb{I}\right) \tag{1}$$

https://doi.org/10.1021/acs.jpca.4c01546
J. Phys. Chem. A XXXX, XXX, XXX−XXX

where N is the number of molecules, $\hat{u}_i$ is the direction of the long axis of each molecule, and $\mathbb{I}$ is the identity matrix. The order parameter S can range from −0.5 to 1, with −0.5 corresponding to perfectly anti-nematic arrangements, 0 for perfectly isotropic phases, and 1 for a perfect nematic phase. However, generating configurations with S values close to these extremes can be particularly difficult in small systems, such as those accessible to DFT calculations.

As reported in past studies, MLFFs may not perform well on unseen configurations.[45,46] This necessitates that the training data span as wide a range of configurations as possible. One would therefore expect a proper training data set to include a uniform distribution of S values. Figure 1(a, left) plots the distribution of S values on the data set collected by unbiased simulations. As evidenced by Figure 1(a, left), the data sampled with a conventional strategy—which is one of the most common approaches to train MLFFs—roughly corresponds to a bimodal distribution of gaussians centered around −0.25 and 0.35 and values mostly ranging from −0.4 to 0.65, leaving large regions of the projected state space either heavily undersampled or not sampled at all. These gaps in the order parameter space negatively impact the performance of MLFFs trained on such data.

Data Dimensionality Reduction and Enhanced Sampling. As shown above, unbiased sampling alone may not adequately capture the full range of relevant configurations for training MLFFs. One approach to overcome this could be the use of an active learning procedure to increase the training data set quality. Some authors have suggested that, in the active learning for interatomic potentials, the goal of active learning should be maximizing the entropy of the training data set distribution with respect to some descriptors, so the model can be accurate and transferable.[47,48] It is known that, without constraints, the distribution that maximizes the entropy is the uniform distribution; therefore, an active learning procedure eventually will generate a uniform distribution over the descriptors used. This uniform distribution over a descriptor can be obtained more efficiently by the use of enhanced sampling techniques, since some of them were designed for this goal.[49] For this reason, we turn to enhanced sampling techniques, which lead to a more efficient exploration of phase space by directly forcing the system to overcome the high free energy barriers that separate metastable states. This idea has recently been adopted[50−52] to improve sampling for training MLFFs.

In our study, we rely on steered molecular dynamics simulations to incrementally adjust the order parameter, S, from −0.5 to 1 for a set of initial configurations of the 5CB liquid crystal. This allows us to systematically drive the system across a range of nonequilibrium states important for training,[53] thereby covering the full range of values of the nematic order parameter. See the Computational Details section for how we implemented steered MD simulation. Figure 1(a, right) depicts the distribution of S values collected by steered MD. We observe that the distribution of S is mostly uniform except for $S \in (-0.5, 0.0, \text{and } 1.0)$, which is a consequence of the finite size of the system.

Beyond focusing on the order parameter, we also aim to capture the behavior of 5CB at different densities. Therefore, we conducted steered MD simulations on the seven systems of different densities (293, 298, 300, 303, 308, 327, and 367 K) to achieve thorough coverage of 5CB's behavior across the nematic and isotropic phases.

Given the high computational demands of DFT calculations and the training and inference of MLFFs, it is helpful to undertake a preliminary analysis of the data to ensure model quality and training efficiency. This initial step involves a detailed a priori examination of the feature space, which is inherently high-dimensional due to the aggregation of local environmental information from all atoms in the data set snapshots. We then apply dimensionality reduction (DR) techniques to identify the underlying patterns and characteristics within the data. DR has long been adopted to analyze the feature space of the neural network and MLFF.[22,54−58] Through both linear and nonlinear transformations, DR methods offer a succinct yet comprehensive view of the data set's structure, capturing the relationships between collected frames.

Here we implement the following steps. Initially, the local environment of each atom within a data set is transformed into input features of the neural network. Subsequently, features derived from atoms in individual frames are aggregated in a composite embedding for each frame. The aggregation is conducted by summing all of the atom embeddings within a single frame, ensuring that the data embedding maintains permutational invariance. This preprocessing step is necessary for two primary reasons: first, extracting information directly from the snapshots, such as atomic positions and forces, is problematic due to the lack of translational and rotational invariance and equivariance. Such data are significantly influenced by periodic boundary conditions and minor disturbances within the system. Second, it allows for an analysis of the neural network's potential perspective during training, offering insights into the input space the neural network will encounter. This alignment ensures a coherent understanding of the neural network's operational environment. Following this preprocessing step, principal component analysis (PCA) is applied to these frame embeddings. This process effectively reduces the dimensionality of the data, mapping each frame onto a 2-dimensional embedding space conducive to direct visualization.

The resulting visual representation is illustrated in Figure 1(b, c). Figure 1(b) shows the strong correlation between phases and the data embedding: the nematic phase (293, 298, 300, 303, 308 K) and the isotropic phase (327, 367 K) are clearly separated on the plot. Figure 1(c) shows that the DR visualizations can be misleading in some cases. While the two data sets have a similar appearance, the data set curated through enhanced sampling exhibits a more uniform distribution in S compared to that of its traditional counterpart.

Importantly, exploring the behavior of 5CB across various phases from nematic to isotropic helps broaden the data set and the sampling space, thereby improving the quality and robustness of the model.

Hyperparameter Tuning. Like other neural networks, neural-network-based MLFFs also involve a large number of hyperparameters, underscoring their shared complexity. The specific hyperparameters for MLFFs depend on the exact architecture of the neural network and can be basically separated into three major categories:

1. *Cutoff Radius*: This parameter defines the size of the local environment for each atom.
2. *Network Structure Parameters*: These parameters determine the size and shape of each layer of neurons.

https://doi.org/10.1021/acs.jpca.4c01546
J. Phys. Chem. A XXXX, XXX, XXX−XXX

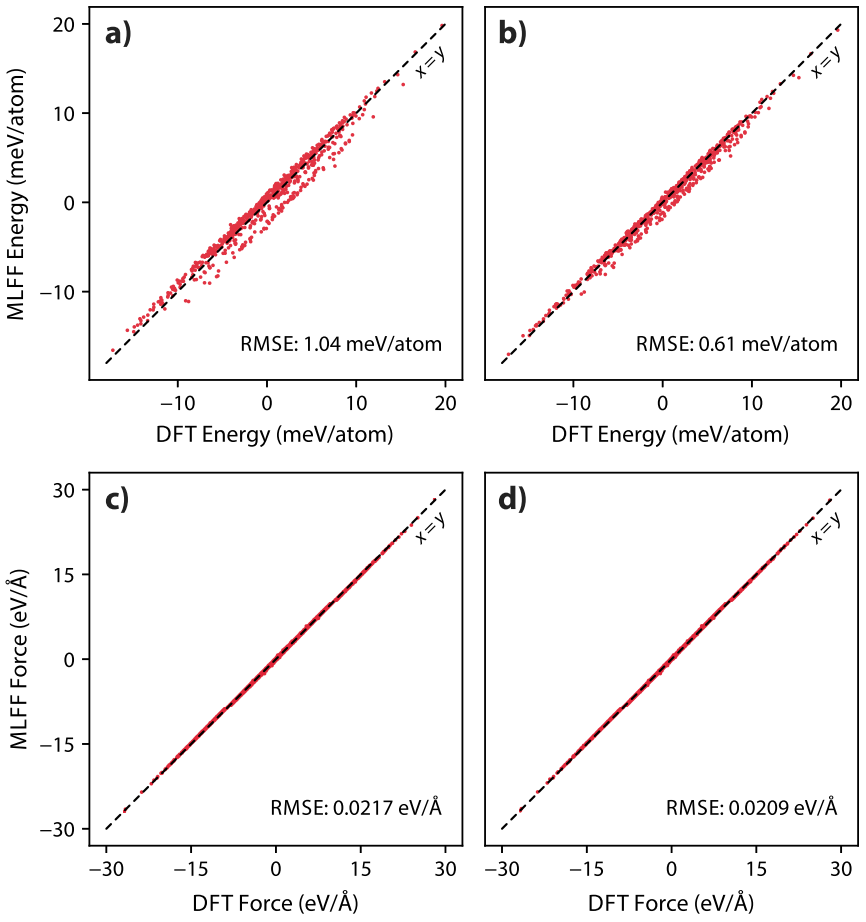


Figure 2. Scatter plot of the energy of the validation set calculated by DFT versus calculated by (a) MLFF-S and (b) MLFF-L. The energy shown is standardized by subtracting the mean of DFT energy from both MLFF energy and DFT energy. (c) Scatter plot of force on all atoms in the validation set in all directions, generated by DFT and MLFF-S. (d) Scatter plot of force generated by DFT and MLFF-L.

3. *Learning-Related Parameters*: This category includes the choice of optimizer for minimizing the loss, the coefficient of the loss function, and the learning rate.

Each of these categories presents unique challenges and trade-offs, necessitating a thoughtful approach to hyperparameter tuning. Striking a balance between model fidelity and practical constraints (such as computational resources and training time) is important. To explore the vast landscape of possible hyperparameter configurations, techniques like grid search and random search are commonly employed. In our software suite, we offer scripts that scan user-provided hyperparameter spaces and automatically perform grid searches or random searches within the configuration space.

In this study, we refrain from comparing learning-related parameters, as our experiments suggest that the learning rate has a minimal effect on the final performance of the force field. Instead, we focused on analyzing the effects of the cutoff radius and network size. For the sake of simplicity, we consider two distinct cases.

1. MLFF-S: A small network with a cutoff radius of 4 Å (R = 4 Å).
2. MLFF-L: A large network with a cutoff radius of 5 Å (R = 5 Å) (refer to the Computational Details section for additional details).

To better understand the impact of enhanced sampling, we train the neural network on both the enhanced sampling data set and the unbiased sampling data set for each case. Our primary focus is on enhanced sampling. Therefore, in the main text, we analyze only the behavior of the model trained on the enhanced sampling data sets. Additional analysis of the models trained on the unbiased sampling data set is provided in the Supporting Information.

Model Training. Once the data set has been created, we proceed to train multiple MLFFs with different neural network architectures and hyperparameters. We chose Allegro,[7] which is based on NequIP,[11] as our neural network architecture for training. MLFFs are trained with the DFT potential energy and force information from the prepared training data. Following previous works,[59] we use 90% of the data for the training process and use the other 10% of the data for validation purposes. As the DFT data are computationally expensive to obtain, we seek to maximize our training data as much as possible. In addition, our validation set is randomly sampled from the variety of configurations that we collected, and 10% is assumed to be sufficiently representative to reliably estimate the model's performance.

Quality Evaluation. Before MLFFs are incorporated into any molecular dynamics simulation, a rigorous evaluation of their quality is essential. Our evaluation framework follows a two-tiered approach to ensure the fidelity of the MLFF models.

1. *Energies and Forces Validation*: We begin by examining the MLFF's predictive accuracy using the collected training and validation data sets. This step ensures that the model accurately

https://doi.org/10.1021/acs.jpca.4c01546
J. Phys. Chem. A XXXX, XXX, XXX−XXX

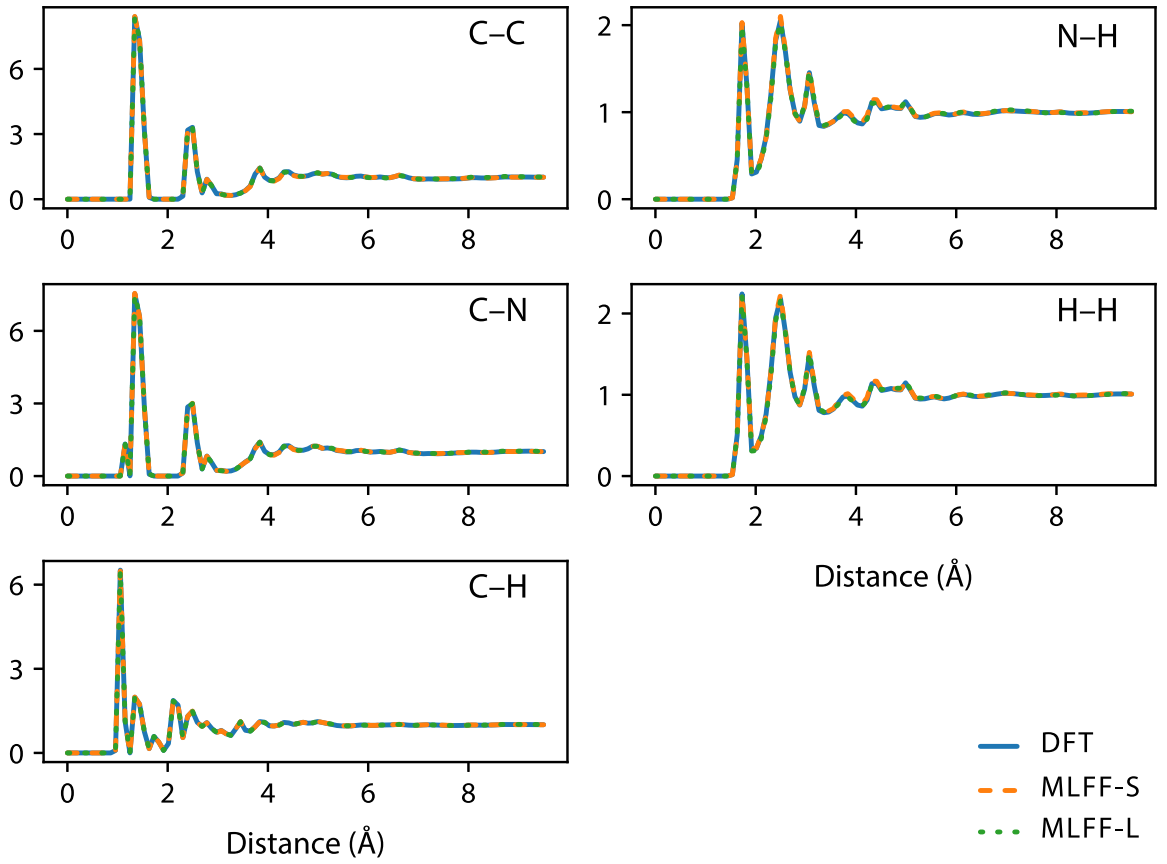


Figure 3. Radial distribution function between any pair of elements in the 5CB bulk system canonical ensemble (NVT) simulation.

reproduces the known outcomes. By comparing predictions against ground truth data, we verify the model's ability to replicate essential properties.

2. *Detailed Quality Assessment*: Beyond the preliminary assessment, we conducted a more thorough evaluation. Our focus shifts to preserving critical physical and chemical properties during MD simulations. Key properties include the radial distribution function, bond stability, dihedral distribution, and dimer formation. To explore the model's predictive performance across diverse conditions, we employ enhanced sampling techniques. This comprehensive evaluation ensures that the MLFF models meet the necessary criteria for reliable MD simulations.

Energies and Forces Validation. As with any machine learning model, MLFFs are trained to minimize a loss function defined on the reference set. Subsequently, the performance of MLFFs can be assessed with how well the MLFFs are able to minimize the error in predicting the energies and forces.

The evaluation process begins by feeding validation data into the MLFF to observe how accurately it predicts relevant properties. Specifically, the evaluation includes a comparison of the root mean squared error (RMSE) across two key metrics: force prediction on a per-atom basis and energy prediction on a per-frame basis. These comparisons yield validation errors associated with the MLFF, setting an initial quality threshold. However, it is essential to recognize that, while correct force and energy predictions are necessary, they alone do not guarantee overall model quality.

Figure 2 illustrates the performance of models with selected hyperparameters trained on the respective data set. From the plot, we can see that for the validation set both models are able to achieve a high level of accuracy on both tasks. Specifically, in Figure 2(a, c), with a small model the MLFF is able to achieve a RMSE of 1.04 meV/atom on energy prediction and 0.0217 eV/Å on force prediction, far beyond the chemical accuracy (1 kcal/mol ≈ 43 meV/atom). With a larger model, in Figure 2(b, d), the performance is even better, with 0.61 meV/atom RMSE on energy prediction and 0.0209 eV/Å on force prediction.

Detailed Quality Assessment. Accurate prediction of the forces and energy is only one aspect of a successful MLFF,[20] especially for soft materials such as liquid crystals. The MLFF must reproduce the important behaviors of the liquid crystal. Though some of these properties can be estimated with short simulations, many typically require long simulations. As noted earlier, MLFFs are in general much faster than DFT calculations, but they are not as fast as classical force fields, which presents challenges for long simulations.[14,20] To mitigate this problem, we adopted enhanced sampling techniques to measure properties efficiently. Our observations indicate that the MLFF proposed here exhibits remarkable robustness throughout our production evaluations. Notably, in our enhanced sampling experiments, the duration of simulations is extended up to 15 ns—a time frame that significantly surpasses those reported in prior studies.[14,20] This observation underscores the reliability of our MLFF pipeline and represents a new benchmark for duration and stability in such simulations.

We begin by studying the radial distribution function (RDF) of 5CB systems. Accurately replicating RDF is a critical goal in the development of new force fields[60] as it reflects the preferred distances for particle arrangements. We run a canonical ensemble simulation on an 18 molecule system with the MLFF for 1 ns and then measure the RDF between any pair of

https://doi.org/10.1021/acs.jpca.4c01546
J. Phys. Chem. A XXXX, XXX, XXX−XXX

elements in the molecule. To estimate the ground truth, we performed the same experiment with ab initio MD simulation for 100 ps. Our simulation with MLFF encompasses 1 order of magnitude longer simulations than those accessible with ab initio simulation. Both yield comparable results, serving to highlight the ability of MLFF to carry out extended simulations without compromising accuracy.

Figure 3 shows the RDF between each pair of elements estimated by each of the MLFFs obtained here as well as the ground truth estimated by DFT. Both the large model and the small model achieve a high level of agreement (RMSE < 0.025) on all pairs of elements. The level of agreement is one magnitude higher than that of the classical force fields (Supporting Information Table 1).

We also conducted an analysis of the potential energy for key dihedral angles within the 5CB molecular structure. This analysis focuses on two critical angles: the inter-ring dihedral, which is the angle between the molecule's two benzene rings, and the tail dihedral, which pertains to the molecule's aliphatic chain (defined by the last four carbons, see the Supporting Information).

For each dihedral angle, we rotate part of a molecule to sweep through a set of angles and calculate the potential energy of each configuration using the MLFF. Parallel DFT calculations are provided as baseline comparisons.

Figure 4 displays the comparison of the potential energy on the two major dihedral angles. From the plot we can see that our

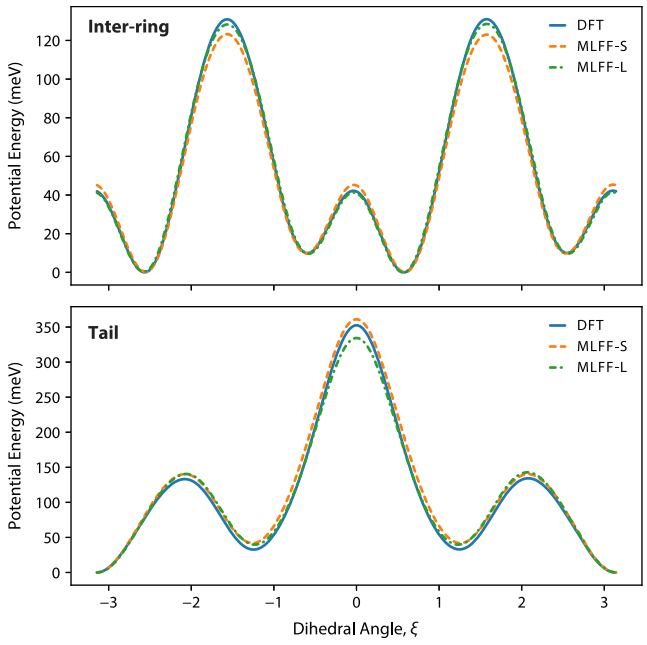


Figure 4. Potential energy obtained by trained MLFFs and calculated from DFT on the (a) inter-ring dihedral and (b) the tail dihedral of a 5CB molecule.

MLFF is able to achieve high accuracy on learning the dihedral angle's potential energy. With the small model, the MLFF can achieve an RMSE of 3.9661 meV on the inter-ring dihedral and 9.4882 meV on the tail dihedral. With the large model, MLFF can achieve an RMSE of 1.4201 meV on the inter-ring dihedral and 7.5756 meV on the tail dihedral.

Numerous applications of liquid crystals stem from their ability to undergo reorientation when exposed to different surfaces or interfaces. This characteristic positions liquid crystals as effective sensors for monitoring interfacial events, an area that has attracted significant interest.[61−66] Building on this feature, we now examine how the 5CB−air interface influences liquid crystal behavior using MLFFs. A 5CB interface system was created with 18 molecules, by expanding the vacuum on both sides of the 5CB bulk in order to mimic the 5CB−air interface, as reported in previous work.[61] We run a canonical ensemble simulation with the two MLFFs over 250 ps.

When MLFF-S is employed, the simulation struggles to accurately form the 5CB−vacuum interface. In this scenario, 5CB molecules tend to disperse throughout the simulation box, failing to exhibit the cohesive behavior characteristic of liquid crystal phases. On the contrary, for MLFF-L, Figure 5(a, b) provides a visual representation of the stable interface under different system sizes: 18 molecules (684 atoms) and 64 molecules (2432 atoms).

We also calculate the free energy between the distance of the center of mass of two 5CB molecules using the spectral adaptive biasing force method (spectral ABF)[30] in Figure 5(c). Utilizing MLFF-L facilitates the formation of the 5CB−vacuum interface. This is reflected in the free energy profile in (c), which notably features two local minima at approximately 4 and 7 Å. These minima correspond to the stable configurations that the system can maintain. The existence of these local minima in the free energy profile illustrates the model's ability to capture the complex intermolecular interactions that govern the equilibrium state of 5CB molecules. We also provided the free energy profiles generated by classical MD simulations and unbiased MLFF models in Figure 2 in the Supporting Information. It is noted the free energy estimated by the MLFF-unbiased model (Supporting Information Figure 2(b, c)) remains constant or decreasing when the distance between the two molecules exceeds 8 Å, unlike the behavior observed in the MLFF model. This indicates that the MLFF-unbiased model has failed to accurately capture the cohesive energy responsible for maintaining the alignment of the dimer.

The alignment of dimers plays a crucial role in determining not only the surface orientation ("anchoring") behavior but also the formation of a nematic phase, which in turn significantly impacts their optical and electro-optical properties, essential for applications from displays to sensors.[61−66]

To further study the alignment of dimers, we evaluate the free energy between a 5CB dimer on a 2D plane to evaluate the dimer alignment, using spectral ABF. Specifically, we use two collective variables; the distance projected between two carbon 6 atoms in two different 5CB molecules, where the axis of projection is defined as the direction from the carbon atom with index 6 to the nitrogen atom with index 1 of the first molecule (see the Supporting Information), whereas the second collective variable is defined as the perpendicular component of distance with respect to the same axis. These simulations provide a comprehensive examination of the energy landscape of dimer configurations and serve as a robust measure of the MLFF's performance for simulations of the complex behavior of liquid crystals.

Figure 6(a, b) shows the free energy of the dimer alignment generated by the two MLFFs. The 2D free energy profile of MLFF-S in (a) has a monotonically decreasing trend, encouraging dispersion. On the other hand, MLFF-L in (b) exhibits three local minima. Figure 6(c) shows the corresponding dimer configurations of each local minimum. These dimer configurations at local minima are in good agreement with the most stable dimer alignments in previous work[67] using DFT.

https://doi.org/10.1021/acs.jpca.4c01546
J. Phys. Chem. A XXXX, XXX, XXX−XXX

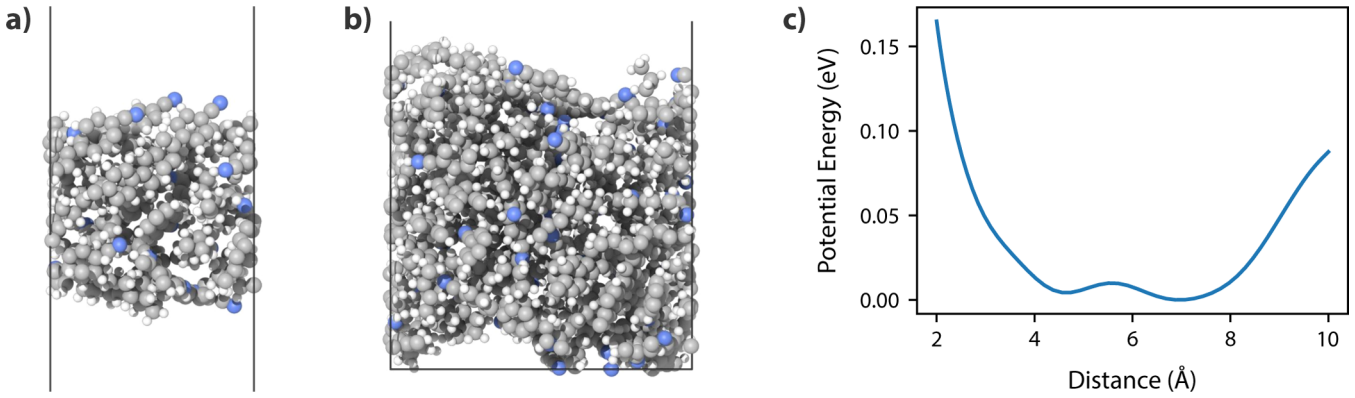


Figure 5. Snapshot of interface systems after NVT simulations of 250 ps with the model MLFF-L in (a) 18 molecules and (b) 64 molecules. (c) Free energy surface of the intermolecular center of mass distance estimated with MLFF-L.

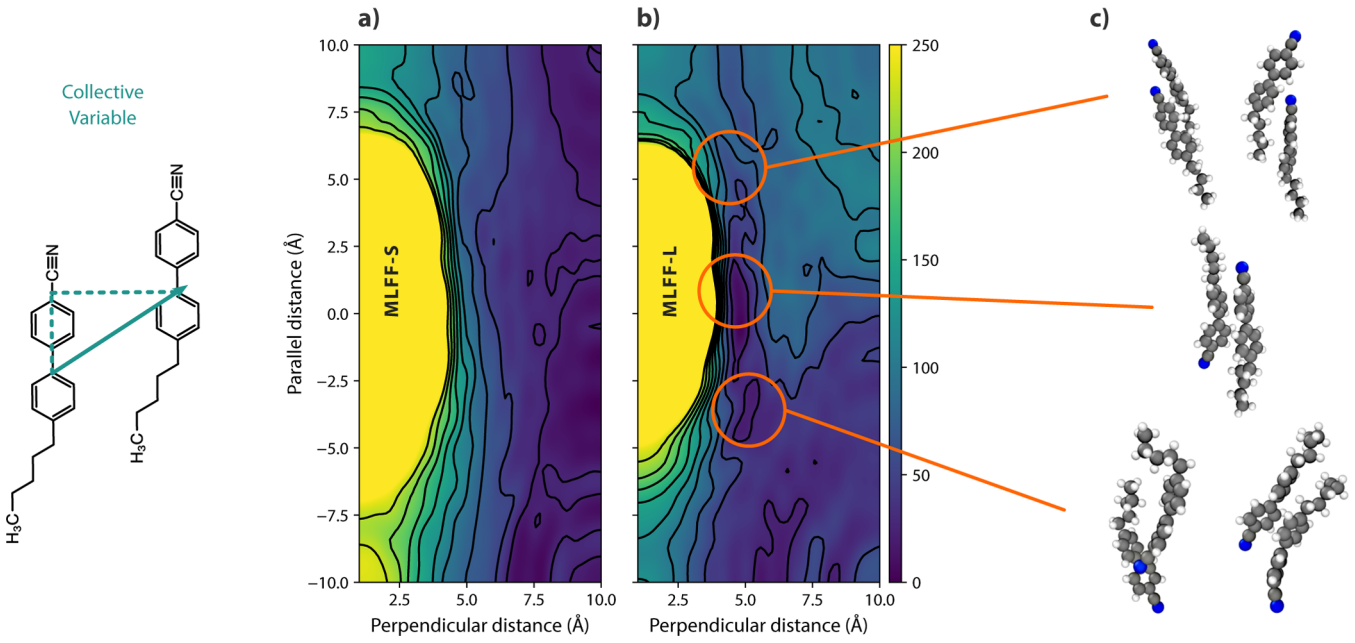


Figure 6. Free energy surface of the 5CB dimer on the 2D coordinate system, produced by (a) MLFF-S and (b) MLFF-L. The energy unit is meV. Local minima are marked with red circles. (c) Snapshots of the corresponding dimer configuration of local minima of MLFF-L from the simulation trajectory.

From top to bottom, the first local minimum corresponds to the side-to-side configuration in the parallel arrangement. The steepest local minimum, which is the second from the top, corresponds to the side-to-side configuration in the antiparallel arrangement, which has been reported to have the lowest potential energy.[67] Additionally, dimer conformations at the local minima have an average length of 25.33 Å, matching the correlation length of 5CB molecules in experimental X-ray measurements at interfaces[61] with an error of 2.69%. The third local minimum corresponds to a metastable region, where both parallel configurations and cross configurations exist. With the proper hyperparameter setting, the MLFF is able to replicate the correct dimer alignment of the 5CB molecules. As a comparison, Supporting Information Figure 3 shows that the free energy surface generated from the classical FF exhibits only one local minimum. Similarly, Supporting Information Figure 4(a, b) illustrates that the free energy surface generated from the MLFF-unbiased model also fails to capture all significant local minima, which are crucial for representing the stable dimer alignment.

Finally, we conduct a comparative analysis of the computational performance of the selected MLFFs with varying cutoff radii and classical force fields across multiple 5CB system sizes. As expected for Allegro, a reduction in the cutoff radius significantly enhances the simulation speed. Specifically, the performance of an MLFF with a 4 Å cutoff radius is observed to be twice as fast as that of an MLFF with a 6 Å cutoff radius. Additionally, our analysis reveals that the computational cost depends linearly on the inverse of the system size. Nevertheless, MLFFs are still 3 orders of magnitude slower than their classical pairwise additive counterparts, limiting their usage as general purpose MD force fields. As shown here, we cannot reduce the cutoff without losing the ability to properly capture the free energy of the system. Other MLFF models might fair better in this regard,[27] but we expect the orders of magnitude of the gap in Figure 7 to remain with other existing ML models.

COMPUTATIONAL DETAILS

We performed all density functional theory (DFT) calculations utilizing CP2K, an open-source software package for electronic

https://doi.org/10.1021/acs.jpca.4c01546
J. Phys. Chem. A XXXX, XXX, XXX−XXX

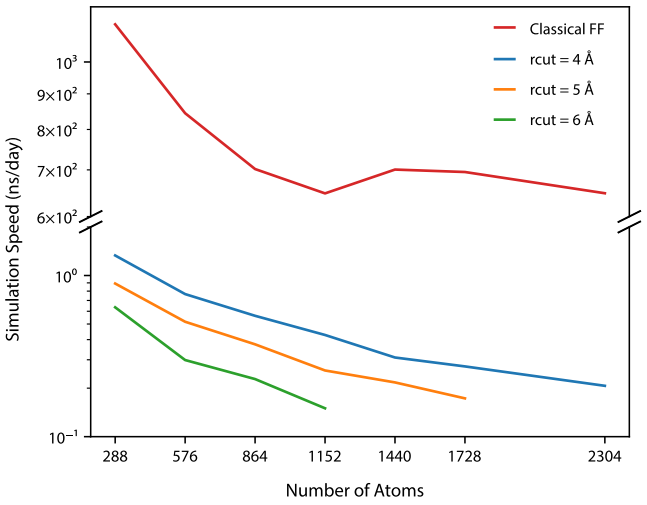


Figure 7. Simulation speed of MLFF with different cutoff radii and classical FF over different system sizes.

structure and molecular dynamics analysis.[68] These calculations were conducted by employing the Gaussian plane wave (GPW) method. Specifically, the GTH-PBE pseudopotential and the DZVP-GTH-PBE basis set were selected for our analysis. A plane-wave cutoff energy of 550 Ry was determined from the convergence tests. Furthermore, we used the Perdew−Burke−Ernzerhof (PBE) model[69] as the exchange-correlation functional. Additionally, to account for dispersion correction, the DFT-D3 method[70] was integrated into the calculations. It has been shown to capture the interfacial interaction of 5CB liquid crystal.[71]

Classical molecular dynamics simulations were conducted employing the general AMBER force field (GAFF2)[72] within the OpenMM software package.[73] The study involves a periodically repeating system comprised of 18 5CB molecules. We initialized the system by randomly packing the 5CB molecules using Packmol at fixed densities. We prepared seven 5CB systems with varying box sizes to mimic the experimental density variations observed at different temperatures (293, 298, 300, 303, 308, 327, and 367 K), covering the nematic and isotropic phases of 5CB. These systems were then equilibrated at 1000 K for a duration of 12 ns using the Langevin thermostat with a friction coefficient of 1 ps^{-1}. Nonbonded interactions were calculated up to 0.8 nm. The particle mesh Ewald (PME) method was employed for the long-range electrostatic interactions. The production simulations for all the systems were run for 2 ns in the NVT ensemble, and we sample 1000 frames at the interval of 2 ps from each of the seven systems for both unbiased and enhanced sampling.

For the enhanced sampling simulations, we implement steered MD using a time dependent harmonic bias in PySAGES.[74] Formally, for any collective variable of interest ξ and time t, we amend the Hamiltonian $\mathcal{H}$ with a term $\mathcal{H}_{HB}(\xi, t)$

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_{HB}(\xi, t) \tag{2}$$

in which

$$\mathcal{H}_{HB}(\xi, t) = \frac{k}{2}(\xi_0 + vt - \xi)^2 \tag{3}$$

where k is the spring constant, ξ_0 is the initial desired CV value, and v is the velocity of the desired CV value w.r.t. time t. In this case, we use $\xi = S$, k = 10,000 kJ/mol, $\xi_0 = -0.5$, and v = 0.75 ns^{-1}. Before the production run, an additional steered MD simulation of 1 ns is performed to ensure that the system's order parameter is sufficiently close to ξ_0.

The training of MLFF for 5CB was performed using NequIP[11] v0.6.0 and the Allegro package.[7] For all models mentioned in this paper, we utilized an initial learning rate of 0.002 and a batch size of 2 and trained the model for 48 h. MLFF-S adopts an environment embedding multiplicity of 64, an edge latent dimension of 128, a two-body latent multilayer perceptron (MLP) dimension of [128, 256, 512], and a latent MLP latent dimension of [512, 512, 512]. MLFF-L adopts an environment embedding multiplicity of 64, an edge latent dimension of 128, a two-body latent MLP dimension of [128, 256, 512, 1024], and a latent MLP latent dimension of [1024, 1024, 1024].

During the inference phase, MD simulations were performed using the atomic simulation environment (ASE).[75] The system was coupled to a Langevin thermostat with a friction of 0.001 fs^{-1}.

Enhanced sampling during data preparation and validation was conducted using the PySAGES package,[74] to facilitate a thorough exploration of configuration space. Data preparation employed the steered MD method. The collective variable chosen during training data preparation was the S order parameter of 5CB molecules, with a range spanning from −0.5 to 1, to accurately capture the systems' configurational diversity.

For model validation, we used the spectral ABF technique[30] to obtain free energy surfaces through enhanced sampling.

In performing spectral ABF simulations, we consistently utilized a grid size of 64 in each dimension. In the case of 1D dimer free energy calculations, the intermolecular distance was adjusted from 2 to 10 Å. Furthermore, for the 2D dimer free energy landscape, we employed a grid spanning from −10 to 10 Å along the axis parallel to the molecular orientation and from 1 to 10 Å along the perpendicular axis. The analysis of potential energy for dihedral angles was modulated across a range from $-\pi$ to π, partitioned into 36 equal intervals.

CONCLUSIONS

We have introduced a thoroughly validated machine learning force field training pipeline. The pipeline was developed and tested in the context of the liquid crystal 5CB, which exhibits considerable variability in the orientational degrees of freedom and presents considerable challenges for traditional two-body classical force fields. During the data generation stage, we distorted our simulations to ensure that the slow modes of the system are properly sampled. By employing dimensionality reduction techniques on our training data, we provide a comprehensive visualization of the data set across various states, ensuring both representativeness and robustness. Our methodology integrates MD simulations and free energy calculations via enhanced sampling to validate the quality of the trained MLFFs. Our findings illustrate that, through careful data set preparation and validation, some MLFFs can effectively model soft materials, such as liquid crystals, and potentially other complex molecular systems. We demonstrated this not only by running long and stable simulations, which have been elusive in the past, but also through validation of several physicochemical properties predicted in these simulations. This suggests that our approach could be applicable across a diverse range of systems, including both soft and solid-state materials.

We conclude with a few cautionary remarks about MLFFs. Despite a cutoff radius of 4 or 5 Å for the network architectures

https://doi.org/10.1021/acs.jpca.4c01546
J. Phys. Chem. A XXXX, XXX, XXX−XXX

examined here, computational efficiency remains inferior to that of traditional force fields. The MLFF's ability to capture rare events and long-term behaviors could still be somewhat limited, even after taking advantage of the strategy presented here. Nonetheless, a systematic approach to data set curation, combined with enhanced sampling and hyperparameter tuning, is shown to yield a more robust and effective MLFF, thereby suggesting promising directions for future research.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.4c01546.

Additional schemes for identification of the dihedral angles (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Pablo F. Zubieta Rico − *Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637-1476, United States;* orcid.org/0000-0002-1211-4836; Email: pzubieta@uchicago.edu

Juan J. de Pablo − *Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637-1476, United States;* orcid.org/0000-0002-3526-516X; Email: depablo@uchicago.edu

Authors

Yezhi Jin − *Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637-1476, United States*

Gustavo R. Perez-Lemus − *Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637-1476, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.4c01546

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is supported by the Department of Energy, Basic Energy Sciences, Materials Science and Engineering Division, through the Midwest Integrated Center for Computational Materials (MICCoM). Yezhi Jin is grateful for a training award from the National Science Foundation under NRT Grant No. 2022023. The authors also acknowledge the Research Computing Center of the University of Chicago for computational resources. The authors are grateful to Prof. Dan Mendels and Dr. Yinan Xu for fruitful discussions.

■ REFERENCES

(1) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.

(2) 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.

(3) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review* **1965**, *140*, A1133.

(4) Car, R.; Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Physical review letters* **1985**, *55*, 2471.

(5) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine learning force fields. *Chem. Rev.* **2021**, *121*, 10142−10186.

(6) Mo, P.; Li, C.; Zhao, D.; Zhang, Y.; Shi, M.; Li, J.; Liu, J. Accurate and efficient molecular dynamics based on machine learning and non von Neumann architecture. *npj Computational Materials* **2022**, *8*, 107.

(7) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* **2023**, *14*, 579.

(8) Jia, W.; Wang, H.; Chen, M.; Lu, D.; Lin, L.; Car, R.; E, W.; Zhang, L. Pushing the Limit of Molecular Dynamics with Ab Initio Accuracy to 100 Million Atoms with Machine Learning. SC20: International Conference for High Performance Computing, Networking, Storage and Analysis. 2020; pp 1−14.

(9) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Physical review letters* **2018**, *120*, 143001.

(10) Zhang, L.; Han, J.; Wang, H.; Saidi, W. A.; Car, R. End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. Advances in Neural Information Processing Systems. 2018.

(11) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13*, 2453.

(12) Gasteiger, J.; Becker, F.; Günnemann, S. G. Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems* **2021**, *34*, 6790−6802.

(13) Schütt, K.; Kindermans, P.-J.; Sauceda Felix, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems* **2017**, *30*.

(14) Liu, L.; Tian, Y.; Yang, X.; Liu, C. Mechanistic insights into water autoionization through metadynamics simulation enhanced by machine learning. *Phys. Rev. Lett.* **2023**, *131*, 158001.

(15) Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling deep learning for materials discovery. *Nature* **2023**, *624*, 80−85.

(16) Schaaf, L. L.; Fako, E.; De, S.; Schäfer, A.; Csányi, G. Accurate energy barriers for catalytic reaction pathways: an automatic training protocol for machine learning force fields. *npj Computational Materials* **2023**, *9*, 180.

(17) Goeminne, R.; Vanduyfhuys, L.; Van Speybroeck, V.; Verstraelen, T. DFT-Quality adsorption simulations in metal−organic frameworks enabled by machine learning Potentials. *J. Chem. Theory Comput.* **2023**, *19*, 6313−6325.

(18) Zhang, Y.; Wang, H.; Chen, W.; Zeng, J.; Zhang, L.; Wang, H.; E, W. A concurrent learning platform for the generation of reliable deep learning based potential energy models. *Comput. Phys. Commun.* **2020**, *253*, 107206.

(19) Morrow, J. D.; Gardner, J. L.; Deringer, V. L. How to validate machine-learned interatomic potentials. *J. Chem. Phys.* **2023**, *158*, 121501.

(20) Fu, X.; Wu, Z.; Wang, W.; Xie, T.; Keten, S.; Gomez-Bombarelli, R.; Jaakkola, T. Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations. 2022.

(21) Park, C. W.; Kornbluth, M.; Vandermause, J.; Wolverton, C.; Kozinsky, B.; Mailoa, J. P. Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture. *npj Computational Materials* **2021**, *7*, 73.

(22) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine learning force fields: construction, validation, and outlook. *J. Phys. Chem. C* **2017**, *121*, 511−522.

(23) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, 3887.

(24) Chmiela, S.; Vassilev-Galindo, V.; Unke, O. T.; Kabylda, A.; Sauceda, H. E.; Tkatchenko, A.; Müller, K.-R. Accurate global machine learning force fields for molecules with hundreds of atoms. *Science Advances* **2023**, *9*, No. eadf0873.
(25) von Rudorff, G. F.; von Lilienfeld, O. A. Simplifying inverse materials design problems for fixed lattices with alchemical chirality. *Science Advances* **2021**, *7*, No. eabf1173.
(26) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Science advances* **2019**, *5*, No. eaav6490.
(27) Batatia, I.; Kovacs, D. P.; Simm, G. N. C.; Ortner, C.; Csanyi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. Advances in Neural Information Processing Systems. **2022**.
(28) Shayestehpour, O.; Zahn, S. Efficient molecular dynamics simulations of deep eutectic solvents with first-principles accuracy using machine learning interatomic potentials. *J. Chem. Theory Comput.* **2023**, *19*, 8732−8742.
(29) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* **2003**, *119*, 3559−3566.
(30) Rico, P. F. Z.; de Pablo, J. J. Sobolev Sampling of Free Energy Landscapes. *arXiv preprint arXiv:2202.01876*. 2022,.
(31) Rahimi, M.; Ramezani-Dakhel, H.; Zhang, R.; Ramirez-Hernandez, A.; Abbott, N. L.; De Pablo, J. J. Segregation of liquid crystal mixtures in topological defects. *Nat. Commun.* **2017**, *8*, 15064.
(32) Koenig, G. M., Jr; Lin, I.-H.; Abbott, N. L. Chemoresponsive assemblies of microparticles at liquid crystalline interfaces. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 3998−4003.
(33) Patel, M.; Radhakrishnan, A. N.; Bescher, L.; Hunter-Sellars, E.; Schmidt-Hansberg, B.; Amstad, E.; Ibsen, S.; Guldin, S. Temperature-induced liquid crystal microdroplet formation in a partially miscible liquid mixture. *Soft Matter* **2021**, *17*, 947−954.
(34) Sidky, H.; de Pablo, J. J.; Whitmer, J. K. In silico measurement of elastic moduli of nematic liquid crystals. *Physical review letters* **2018**, *120*, 107801.
(35) Bezrodna, T.; Nesprava, V.; Melnyk, V.; Klishevich, G.; Curmei, N.; Gavrilko, T.; Roshchin, O.; Baran, J.; Drozd, M. Conformation-dependent molecular association and spectral properties of 4-pentyl-4'-cyanobiphenyl liquid crystal in different phases. *Low Temperature Physics* **2023**, *49*, 302−309.
(36) Goodby, J. W. 4'-pentyl-4-cyanobiphenyl-5CB. *Liq. Cryst.* **2024**, 1−24.
(37) Schwantes, C. R.; Pande, V. S. Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9*, 2000−2009.
(38) Chen, W.; Ferguson, A. L. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *Journal of computational chemistry* **2018**, *39*, 2079−2102.
(39) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018**, *9*, 5.
(40) Chen, W.; Sidky, H.; Ferguson, A. L. Capabilities and limitations of time-lagged autoencoders for slow mode discovery in dynamical systems. *J. Chem. Phys.* **2019**, *151*, 064123.
(41) Sivaraman, G.; Krishnamoorthy, A. N.; Baur, M.; Holm, C.; Stan, M.; Csányi, G.; Benmore, C.; Vázquez-Mayagoitia, Á. Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide. *npj Computational Materials* **2020**, *6*, 104.
(42) Ortner, C. On the Atomic Cluster Expansion: interatomic potentials and beyond. *arXiv preprint arXiv:2308.06462*. 2023.
(43) De Gennes, P.-G.; Prost, J. *The physics of liquid crystals*; Oxford University Press: 1993.
(44) Kléman, M.; Lavrentovich, O. D. *Soft Matter Physics: An Introduction*; Springer: New York, 2003.
(45) Poltavsky, I.; Tkatchenko, A. Machine learning force fields: Recent advances and remaining challenges. *J. Phys. Chem. Lett.* **2021**, *12*, 6551−6564.
(46) Pinheiro, M.; Ge, F.; Ferré, N.; Dral, P. O.; Barbatti, M. Choosing the right molecular machine learning potential. *Chemical Science* **2021**, *12*, 14396−14413.
(47) Montes de Oca Zapiain, D.; Wood, M. A.; Lubbers, N.; Pereyra, C. Z.; Thompson, A. P.; Perez, D. Training data selection for accuracy and transferability of interatomic potentials. *npj Computational Materials* **2022**, *8*, 189.
(48) Karabin, M.; Perez, D. An entropy-maximization approach to automated training set generation for interatomic potentials. *J. Chem. Phys.* **2020**, *153*, 094110.
(49) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemotte, L. Enhanced sampling methods for molecular dynamics simulations. *arXiv preprint arXiv:2202.04164*. 2022.
(50) Herr, J. E.; Yao, K.; McIntyre, R.; Toth, D. W.; Parkhill, J. Metadynamics for training neural network model chemistries: A competitive assessment. *J. Chem. Phys.* **2018**, *148*, 241710.
(51) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148*, 241733.
(52) Yang, M.; Bonati, L.; Polino, D.; Parrinello, M. Using metadynamics to build neural network potentials for reactive events: the case of urea decomposition in water. *Catal. Today* **2022**, *387*, 143−149.
(53) Stocker, S.; Gasteiger, J.; Becker, F.; Günnemann, S.; Margraf, J. T. How robust are modern graph neural network potentials in long and hot molecular dynamics simulations? *Machine Learning: Science and Technology* **2022**, *3*, 045010.
(54) Reif, E.; Yuan, A.; Wattenberg, M.; Viegas, F. B.; Coenen, A.; Pearce, A.; Kim, B. Visualizing and measuring the geometry of BERT. *Advances in Neural Information Processing Systems* **2019**, *32*.
(55) Kiarashinejad, Y.; Abdollahramezani, S.; Adibi, A. Deep learning approach based on dimensionality reduction for designing electromagnetic nanostructures. *npj Computational Materials* **2020**, *6*, 12.
(56) Anishchenko, I.; Pellock, S.; Chidyausiku, T.; Ramelot, T.; Ovchinnikov, S.; Hao, J.; Bafna, K.; Norn, C.; Kang, A.; Bera, A.; et al. De novo protein design by deep network hallucination. *Nature* **2021**, *600*, 547−552.
(57) Trozzi, F.; Wang, X.; Tao, P. UMAP as a dimensionality reduction tool for molecular dynamics simulations of biomacromolecules: a comparison study. *J. Phys. Chem. B* **2021**, *125*, 5022−5034.
(58) Kulichenko, M.; Barros, K.; Lubbers, N.; Li, Y. W.; Messerly, R.; Tretiak, S.; Smith, J. S.; Nebgen, B. Uncertainty-driven dynamics for active learning of interatomic potentials. *Nature Computational Science* **2023**, *3*, 230−239.
(59) Yoo, P.; Sakano, M.; Desai, S.; Islam, M. M.; Liao, P.; Strachan, A. Neural network reactive force field for C, H, N, and O systems. *npj Computational Materials* **2021**, *7*, 9.
(60) Henderson, R. L. A uniqueness theorem for fluid pair correlation functions. *Phys. Lett. A* **1974**, *49*, 197−198.
(61) Sadati, M.; Ramezani-Dakhel, H.; Bu, W.; Sevgen, E.; Liang, Z.; Erol, C.; Rahimi, M.; Taheri Qazvini, N.; Lin, B.; Abbott, N.; et al. Molecular structure of canonical liquid crystal interfaces. *J. Am. Chem. Soc.* **2017**, *139*, 3841−3850.
(62) Bukusoglu, E.; Wang, X. G.; Martinez-Gonzalez, J. A.; de Pablo, J. J.; Abbott, N. L. Stimuli-Responsive Cubosomes Formed from Blue Phase Liquid Crystals. *Adv. Mater.* **2015**, *27*, 6892−6898.
(63) Carlton, R. J.; Hunter, J. T.; Miller, D. S.; Abbasi, R.; Mushenheim, P. C.; Tan, L. N.; Abbott, N. L. Chemical and Biological Sensing Using Liquid Crystals. *Liquid Crystals Reviews* **2013**, *1*, 29−51.
(64) Carlton, R. J.; Ma, C. D.; Gupta, J. K.; Abbott, N. L. Influence of Specific Anions on the Orientational Ordering of Thermotropic Liquid Crystals at Aqueous Interfaces. *Langmuir* **2012**, *28*, 12796−12805.
(65) Moreno-Razo, J. A.; Sambriski, E. J.; Abbott, N. L.; Hernandez-Ortiz, J. P.; de Pablo, J. J. Liquid-Crystal Mediated Self-assembly at Nanodroplet Interfaces. *Nature* **2012**, *485*, 86−89.
(66) Sadati, M.; Apik, A. I.; Armas-Perez, J. C.; Martinez-Gonzalez, J.; Hernandez-Ortiz, J. P.; Abbott, N. L.; de Pablo, J. J. Liquid Crystal Enabled Early Stage Detection of Beta Amyloid Formation on Lipid Monolayers. *Adv. Funct. Mater.* **2015**, *25*, 6050−6060.

https://doi.org/10.1021/acs.jpca.4c01546
J. Phys. Chem. A XXXX, XXX, XXX−XXX

(67) Amovilli, C.; Cacelli, I.; Campanile, S.; Prampolini, G. Calculation of the intermolecular energy of large molecules by a fragmentation scheme: Application to the 4-n-pentyl-4'-cyanobiphenyl (5CB) dimer. *J. Chem. Phys.* **2002**, *117*, 3003−3012.
(68) Kühne, T.; Iannuzzi, M.; Del Ben, M.; Rybkin, V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R.; Schütt, O.; Schiffmann, F. CP2K: An electronic structure and molecular dynamics software package-Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, 194103.
(69) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(70) 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.
(71) Brown, P. A.; Kołacz, J.; Fischer, S. A.; Spillmann, C. M.; Gunlycke, D. Insertion of the Liquid crystal 5CB into Monovacancy graphene. *Molecules* **2022**, *27*, 1664.
(72) Case, D.; Betz, R.; Cerutti, D.; Cheatham, T. I.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A.; Homeyer, N.; et al. *AMBER 2016*; University of California: San Francisco, CA, 2016.
(73) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology* **2017**, *13*, No. e1005659.
(74) Rico, P. F. Z.; Schneider, L.; Pérez-Lemus, G. R.; Alessandri, R.; Dasetty, S.; Nguyen, T. D.; Menéndez, C. A.; Wu, Y.; Jin, Y.; Xu, Y.; et al. PySAGES: Flexible, advanced sampling methods accelerated with GPUs. *npj Computational Materials* **2024**, *10*, 35.
(75) Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N; Hammer, B.ør.; Hargus, C.; et al. The atomic simulation environment—a Python library for working with atoms. *J. Phys.: Condens. Matter* **2017**, *29*, 273002.

https://doi.org/10.1021/acs.jpca.4c01546
J. Phys. Chem. A XXXX, XXX, XXX−XXX