Regularized by Physics: Graph Neural Network Parametrized Potentials for the Description of Intermolecular Interactions

Moritz Thürlemann, Lennard Böselt, and Sereina Riniker*

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: Simulations of molecular systems using electronic structure methods are still not feasible for many systems of biological importance. As a result, empirical methods such as force fields (FF) have become an established tool for the simulation of large and complex molecular systems. The parametrization of FF is, however, time-consuming and has traditionally been based on experimental data. Recent years have therefore seen increasing efforts to automatize FF parametrization or to replace FF with machine-learning (ML) based potentials. Here, we propose an alternative strategy to parametrize FF, which makes use of ML and gradient-descent based optimization while retaining a functional form founded in physics. Using a predefined functional form is shown to enable interpretability, robustness, and efficient simulations of large systems over long time scales. To demonstrate the strength of the proposed method, a fixed-charge and a polarizable model are trained on *ab initio* potential-energy surfaces. Given only information about the constituting elements, the molecular topology, and reference potential energies, the models successfully learn to assign atom types and corresponding FF parameters from scratch. The resulting models and parameters are validated on a wide range of experimentally and computationally derived properties of systems including dimers, pure liquids, and molecular crystals.
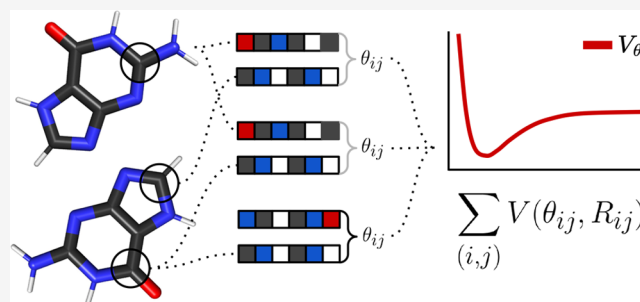
1. INTRODUCTION

Computer based simulations are a powerful tool for the investigation of chemical systems.[1−3] Performing such simulations requires an accurate description of intermolecular forces.[4] However, due to the computational complexity of *ab initio* methods[5−9] or density functional theory (DFT),[10,11] an exact description is out of reach for most systems, particularly systems of biological relevance.[12,13] As a result, a variety of approximate methods has been developed, which can be broadly categorized into three classes: semiempirical methods, classical force fields (FF), and machine learning (ML) based models. Semiempirical methods explicitly describe the electronic structure.[14,15] However, various approximations are introduced to reduce computational costs. Existing methods attempt to compensate for these approximations by introducing a relatively small number of empirical parameters.[16−18] Classical force fields, on the other hand, forego an explicit description of the electronic structure and employ instead a predefined functional form and associated parameters, which together encapsulate aspects of a given interaction.[19,20] Due to their simplicity, they can be evaluated much more efficiently than *ab initio* or semiempirical methods, but generally require a larger number of parameters and extensive parametrization. In recent years, ML-based models have emerged as a third alternative,[21] which assume fewer inductive biases, but require an even larger number of parameters compared to the two previous approaches. Even though very promising results have

been reported for ML potentials,[21−26] their application to condensed-phase systems and the prediction of experimentally measured properties has been fairly limited.[27−32]

The relative scarcity of application of ML potentials to propagate molecular dynamics (MD) simulations is likely a result of insufficient data efficiency due to a lack of inductive biases as well as difficulties posed by the presence of a large number of relatively weak and long-ranged interactions in condensed-phase systems.[33] In addition, doubts have been raised if ML potentials are robust enough to perform MD simulations.[34] There are also discussions whether ML potentials describe features of the potential-energy surface (PES), such as its curvature, sufficiently accurate to perform MD simulations.[35] With their predefined functional form, which is physically motivated, FF have become an established tool to simulate large condensed-phase systems such as solvated proteins over long time-scales.[1] While the FF formalism provides a computationally efficient, interpretable and robust way to describe forces in molecular systems, this
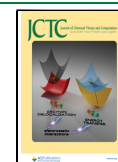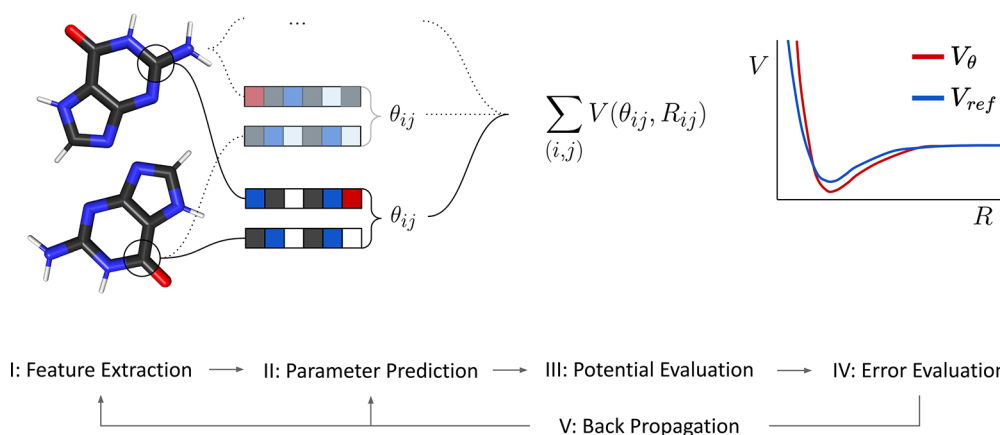
Figure 1. Overview of the proposed parametrization workflow: (I) Atomic environments are encoded as feature vectors with a GNN. (II) Parameters θ_{ij} are predicted for tuples of atomic features. (III) In conjunction with predefined parametric interaction potentials V, the predicted parameters θ_{ij} are used to evaluate the potential energy of a state. (IV) The prediction is evaluated against reference properties. (V) Using automatic differentiation, errors with respect to the target properties are backpropagated to improve the quality of the predicted parameters.

robustness comes at a price. Most of the commonly used FF do not account for phenomena such as charge anisotropy or polarization. In addition, the development of a FF is still a nontrivial process, despite advances in automation over the past years.[36−42] The OpenFF initiative in particular has initiated a grand effort to fully automate this process, including atom-typing,[43,44] data generation,[45] parametrization,[46] and validation.[47]

While FF were historically and continue to be (partly) fitted to experimental data, the advances in computational power and improved scaling of methods based on quantum-mechanical (QM) calculations has opened up new opportunities.[19] As a readily accessible data source, focus has shifted to parametrization with respect to QM reference data such as torsion profiles or interaction components. In addition, there is an increased effort to extract FF parameters directly from electron densities.[48−54] As an example, van Vleet et al.[55] developed an *ab initio* FF based on a parametrization formalism using Slater functions from which certain parameters were directly derived. An alternative approach is to keep a predefined functional form to describe intermolecular potentials but obtain the parameters from a ML model.[56−58] In a similar fashion, Wang et al.[59] and Harris et al.[60] investigated graph neural networks (GNN) and graph-convolutional neural networks (GCNN) in combination with automatic differentiation as a method to parametrize FF. Focusing on intramolecular interactions, they could show that GNN can be used to predict FF parameters from potential energies and recover human-defined atom types. Furthermore, Li et al.[61] demonstrated for the case of a ML-DFT functional that inclusion of physical priors in the learning process by solving the Kohn−Sham equations during training resulted in improved generalization. Finally, we also point out recent efforts which use ML in a complementary fashion to extract symbolic expressions from data.[62−66]

In this work, we build on these developments to propose a universal framework for the parametrization of FF, focusing on intermolecular interactions. Besides generalizing the parametrization process, we describe a formalism for end-to-end differentiable FF, taking advantage of learned atom types. The proposed approach is applied to the parametrization of a nonpolarizable as well as a polarizable FF. Multipoles and monopoles used to describe electrostatic interactions were obtained separately from our previously introduced equivariant

GNN model without further modifications.[67] For both the nonpolarizable and polarizable FF, the model is trained on the PES of dimer interaction potentials of the recently published DES5M data set[68] and a data set of intermolecular potentials of molecular crystals, which was generated for this work. Hence, the models learn the parameters from scratch to reproduce the given PES. We show that a fixed-charge FF parametrized in such a manner can be used to reproduce experimental condensed-phase properties. We find that the resulting models outperform comparable models for the considered test cases.

2. THEORY

2.1. Formalism. Assuming a chemist's viewpoint, molecules can be interpreted as graphs $G = (A, B)$ with nodes (atoms) A and edges (bonds) B. Accordingly, a FF consists of a function $\mathcal{G}: G \to \theta$, which assigns FF parameters θ to a molecular graph G and a function

$$\mathcal{V}_\theta(x) = \sum_i \mathcal{V}_{\theta,i}(x) \tag{1}$$

where the total potential energy of a state x is decomposed into $\mathcal{V}_{\theta,i}$. The function $\mathcal{G}$ that assigns parameters θ can be understood as a parametrized, or learnable, function itself. For commonly used FF, $\mathcal{G}$ is generally expert-devised and depends only on atomic features such as element types and hybridization states. For FF that are parametrized based on electron densities, $\mathcal{G}$ partitions and assigns parameters to an electron density.

Given a parametrization function $\mathcal{G}$, a FF can thus be interpreted as a function $\mathcal{V}_\theta$ with parameters θ and a functional form $\mathcal{V}$, $\mathcal{V}_\theta: X \to V$, which maps a PES $V \in \mathbf{R}$ to the states $X \in \mathbf{R}^n$ of a system, with n denoting the dimensionality of the system. A system can thus be propagated in time by integrating the negative derivative of the potential energy $v \in V$ with respect to its current state $x \in X$,

$$\mathcal{F}(x) = -\nabla_x v = -\nabla_x \mathcal{V}_\theta(x), \tag{2}$$

where $F \in \mathbf{R}^n$ is the gradient field of the negative potential energy. Propagating the system for an appropriate amount of time, a system property P can be derived as $\mathcal{P}: X \to P$ with a function $\mathcal{P}$, which assigns a property to a state or set of states

of the system. Generally, system properties can be scalar, vectorial, or tensorial. They may be defined for each configuration or ensemble average, and depend on the functional form of the FF and its parametrization $\mathcal{V}_\theta$ (Figure 1). Given a system property obtained from a simulation, P_{pred} and a reference value P_{ref} (e.g., from experiment or a QM reference calculation), a loss L can be defined as $\mathcal{L}: (P_{pred}, P_{ref}) \rightarrow L$ with a loss function $\mathcal{L}$. Due to its generality, any computable property can be used as target. Examples include the potential energy, gradients, or Hessians from QM reference calculations, but also experimental properties would be possible such as geometrical constraints from NMR or crystallography, vibrational spectra, or ensemble properties such as enthalpies of phase transitions.

2.2. Parametrization. Given a loss function $\mathcal{L}$ and parameters θ, a FF may be optimized to yield more accurate properties. Specifically, the derivative of the loss function with respect to the FF parameters,

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial \mathcal{L}}{\partial \mathcal{P}_\theta} \cdots \frac{\partial \phi}{\partial \theta}, \tag{3}$$

can be used to update FF parameters successively with gradient descent based optimization methods analogously to back-propagation used in deep learning.[69] With the help of automatic differentiation libraries, these gradients can be evaluated with minimal additional programming effort and computational cost.[70,71]

2.3. Graph Neural Networks as Universal Parametrization Functions. GNN are ML models parametrized by artificial neural networks (ANN) that process graph-structured data. In the commonly used form, node, edge and/or global features are iteratively refined based on the current features. GNN models differ mainly by the features used, the way the underlying graph is constructed, and the updating or feature-refinement process applied.[72−75]

Considering a molecular graph $G = (A, B)$ with nodes $a_i \in A$ and edges $b_{ij} \in B$, message passing can be defined as

$$m_{ij} = \phi_b(h_i^l, h_j^l, b_{ij})$$
$$m_i = \sum_{j \in N(i)} m_{ij}$$
$$h_i^{l+1} = \phi_h(h_i^l, m_i), \tag{4}$$

with $h_i^l \in \mathbf{R}^n$ describing the hidden feature vector of node a_i after l graph processing steps, $b_{ij} \in \mathbf{R}^n$ the bond feature of bond b_{ij} between node i and j, $N(i)$ denoting the set of neighbors of a_i and m_{ij} the message from node j to node i. ϕ_b and ϕ_h represent ANN parametrized edge and node update functions. The superscript l denotes the current layer or the current iteration in the recurrent realization. After n iterations, the refined node feature h_i^n is used as an atomic-environment descriptor in subsequent steps. We note that h_i^n does not necessarily have to be obtained from a GNN. In principle, any other descriptor can be used, for instance atom-based topological fingerprints. However, GNN present themselves as a natural choice to work with graph-structured data, which in turn is ideally suited for a classical description of molecules.

2.4. Parameter Prediction. To predict atomic parameters, learned atom features h_i^n obtained from the GNN serve as descriptors of the atomic environment. Assuming that FF parameters are a function of the atomic environment, such

features can be used to distinguish atom types and to assign FF parameters. Hence, atomic parameters are predicted as

$$\theta_i = \phi_{atom}(h_i^n), \tag{5}$$

with ϕ_{atom} denoting the readout function that assigns the parameter θ_i for a given interaction to an atom type. In general, FF also describe how parameters of two particles are combined to describe a given interaction. Standard biomolecular FF use combination rules to derive parameters of two distinct atom types.[19] By design, combination rules must be symmetric under arbitrary permutations of types present in the interaction. Instead of using established combination rules such as arithmetic or geometric means, pairwise parameters are predicted as a function of two atom types. Specifically, the following approach was chosen

$$\theta_{ij} = \phi_{pair}(h_i^n, h_j^n) + \phi_{pair}(h_j^n, h_i^n) \tag{6}$$

In this case, ϕ_{pair} is an ANN parametrized function, which assigns the parameters of a specific interaction to a pair of atomic environments. This approach guarantees symmetry with respect to permutation of two atoms and allows for more complex combination rules.

2.5. Force-Field Parametrization Function. Combining the previously introduced concepts with the formalism described in section 2.1, the FF parametrization $\mathcal{G}_\phi$ can be defined as the combination of the following two components: (i) a typing function to assign atom types to a given system, and (ii) a combination rule, which returns parameters for a set of atoms partaking in a given interaction. In the current work, the function assigning atom types is modeled with a GNN consisting of a node and edge update layer, ϕ_h and ϕ_b, and a combination rule ϕ_{Atom} or ϕ_{Pair} for atomic and pairwise parameters, respectively. In the context of the GNN formalism, ϕ_θ can be considered a readout function, which maps the hidden state of node features to a label. In our case, FF parameters are mapped to interactions between given atom types. Evidently, established FF atom type definitions and parametrization procedures can be viewed as a special case of the described formalism. For example, the SMARTS patterns used for chemical perception in OpenFF[44] can be cast as graph-based operations that account for features such as the element, its coordination number, bonded neighbors, or subgraph features. In the case of models, which derive parameters from electron densities, the partition function used to decompose the electron density into atomic contributions takes up the role of the typing functions. In these cases, the combination rules are often derived from first principle considerations or empirically fitted.[48,50]

2.6. Models. Besides the aforementioned components, which assign parameters to a given interaction or atom, a FF must further define potential-energy terms and an associated functional form. The chosen functional form reflects the assumptions of the model, and thus determines the accuracy of the model, its computational cost, and its capabilities.

In this work, we consider two models. The first model is based on three interactions: (i) repulsive, (ii) attractive, and (iii) electrostatic. The first two components follow the functional form of the Mie potential[76] with a repulsive C_9 term and an attractive C_6 term. The electrostatic component is described with Coulomb's law and fixed partial charges. Hence, this model assumes an isotropic-pairwise-additive form, and we will refer to it as "IPA model". The second model considered in

Table 1. Mean Absolute Error (MAE) in [kJ/mol] for the IPA$_d$, IPA$_{d+c}$ and ANA$_d$ Models for Intermolecular Interactions in Dimers^a

Data set	Data points	IPA$_{d+c}$	IPA$_d$	ANA$_d$	CGenFF	PM6-DH2	PBE0-D3
S66x8[98]	528	2.9	2.3	1.1		3.3[105]	1.0[106]
SSI[91]	2596	1.4	1.1	1.0	1.3	1.1	0.5
BBI[91]	100	3.3	0.8	1.1	2.1	2.9	0.3
UBQ[95]	81	3.9	1.7	1.4	3.7	1.4	
ACHC[97]	54	2.7	5.8	2.7			1.8
JSCH[96]	123	5.9	6.5	2.3			
HSG[94]	16	1.8	1.3	0.9	1.3	1.8	1.3
D1200[99]	401	4.2	2.1	1.9			
D442x10[99]	1230	4.6	3.1	2.9			
R739x5[100]	1370	4.5	3.7	3.6			
HB375x10[101]	3750	4.4	4.9	1.9			
HB300SPXx10[102]	640	3.5	3.8	2.3			
S7L[103,104]	7	8.8	13.9	10.3			

^{a}Test sets from the Biofragment database (SSI, BBI, UBQ, ACHC, JSCH, HSG)[91−97] and the non-covalent interaction (NCI) atlas (S66x8, D1200, D442x10, R739x5, HB375x10, HB300SPXx10).[98−102] IPA$_{d+c}$ refers to the fixed-charge model, which included training on crystal structures. The IPA$_d$ and the polarizable ANA$_d$ models were trained exclusively on dimer interaction potentials from the DES5M data set.[68] The full error statistics are given in Tables S1−S3 in the Supporting Information. Where available, values for the classical force field CGenFF,[107] the semi-empirical model PM6-DH2,[108] and the DFT method PBE0-D3[109−111] are reported for comparison. Reference values were taken from the publication of the corresponding data set if not indicated otherwise. Values were converted to kJ/mol using a factor of 4.184. Note that PBE0-D3 does not use the same basis sets in all cases. Values for the largest available basis set (def2-QZVP or aug-cc-pVTZ) and counterpoise correction were chosen if available.

this work is based on multipole electrostatics and polarizable atoms, therefore violating isotropy and pairwise additivity. We will refer to this anisotropic-nonadditive model as "ANA model". The IPA model is similar to the functional form of the most widely used fixed-charge FF.[19,46,77−81] The functional form of the ANA model, on the other hand, is an attempt to explore the limits of a purely classical model through an implicit description of the electron distribution based on atomic multipoles and induced dipoles. Its functional form is similar to the class of polarizable and QM derived FF.[56,57,82−87] A detailed description of the functional forms and the included terms is provided in the Methods section.

In addition to the two functional forms used, we also investigated two training data regimes. The IPA$_d$ and ANA$_d$ models were only trained on dimer interaction potentials in vacuum using the SNS-MP2[88] reference values of the DES5M data set.[68] The model labeled with IPA$_{d+c}$ was additionally trained on intermolecular potentials of molecular crystals, which were obtained with B86bPBE-XDM.[89,90] The performance of the IPA$_d$, IPA$_{d+c}$, and ANA$_d$ models was investigated for a wide range of systems and environments. For the IPA$_{d+c}$ model, emphasis is put on the performance in condensed-phase systems, which are generally challenging for nonclassical methods or ML-based models. With its more sophisticated functional form, the ANA model is applied to intermolecular potentials of small-molecule dimers, which permits a direct comparison with first-principle and DFT methods. Finally, further explorations of parameters learned by both approaches are presented. We note that the ANA model has not been applied to condensed-phase systems.

3. RESULTS AND DISCUSSION

3.1. Intermolecular Potentials in Vacuum. Small-molecule dimers are accessible to highly accurate wave function methods. As such they present a valuable validation case to probe the accuracy of the description of specific interactions. For this purpose, several established noncovalent

interaction benchmarks were taken as test sets from the Biofragment database[91−97] and the noncovalent interaction (NCI) atlas[98−102] (Figure 5). In addition, the models are tested on the S7L[103] data set. For systems in the S7L data set, estimated CCSD(T) binding energies from a recent study[104] were used as reference. In all cases, only systems consisting of neutral monomers with more than two atoms were included. The performance results on the benchmarking sets with a total of 10,896 data points are summarized in Table 1. The full error statistics are given in Table S1−S3 in the Supporting Information.

The results of the three models investigated in this study are shown in Table 1. Consistent results over these diverse sets are observed. For comparison, reference values are given for the data sets in the Biofragment database where available. Three methods were chosen to represent classical force fields (CHARMM General FF (CGenFF)[107]), semiempirical models (PM6-DH2[108]), and DFT methods (PBE0-D3[109−111]). The IPA models perform comparable to previously reported results for empirical (FF) or semiempirical models. In general, the simpler models are outperformed by the ANA$_d$ model, which achieves for some data sets an accuracy comparable to dispersion corrected hybrid functionals like PBE0-D3BJ. Exceptions are the BBI data sets, where the IPA$_d$ model performs better than ANA$_d$, and the S7L data set for which IPA$_{d+c}$ outperforms the other models.

Comparison between IPA$_{d+c}$ and IPA$_d$ suggests that additional training on crystal data points does not necessarily lead to a strong negative impact on the description accuracy for dimers in vacuum. This is surprising for two reasons: (i) The method used to compute the potential energy of crystal structures is considerably less accurate than the methods used to compute the dimer interaction potentials in the bench-marking data sets and the DES5M training data set. (ii) Polarization effects, which the IPA models must implicitly take into account, differ between vacuum and condensed-phase environments. This effect is most pronounced for the S7L data set, where IPA$_{d+c}$ outperforms the IPA$_d$ and ANA$_d$ models.

Interestingly, training on crystal potential energies appears to provide a regularizing effect for the IPA_{d+c} model, offering not the most accurate but the most consistent results over a wide range of applications beyond dimers (see below).

The biggest deviations are observed for the large systems in the S7L data set. While not larger in relative terms, it is interesting to note that the error for the two models trained on dimer data only (i.e., IPA_d and ANA_d) is almost completely due to a systematic overestimation of the binding energy, with a mean error equal to the mean absolute error and with errors for each system roughly proportional to the magnitude of the interaction (Table S1 in the Supporting Information).

For the ANA_d model, which includes an explicit treatment of polarization, this observation may indicate an imbalance between pairwise and many-body interactions that is not present for the small-molecule dimers in the DES5M data set. Finding a way to improve this balance might be an important step to construct potentials that generalize from dimers to larger assemblies and condensed-phase systems. Developing additional benchmark data sets with a broader coverage of the space from medium to large structures and resolving the disagreement between CCSD(T) and fixed-node diffusion Monte Carlo for large structures, which was brought up in ref 104 could be beneficial to validate such efforts.

3.2. IPA_{d+c} Model—Crystalline Systems.

Reproducing properties of liquid or crystalline systems is an important validation task for intermolecular potentials. QM methods are in general not feasible to simulate such systems, and the wide range of interactions poses considerable challenges to ML models.

Intermolecular Potentials in Crystals. Intermolecular potential energies calculated for a wide range of molecular crystals were used to parametrize the IPA_{d+c} model in addition to the dimers in vacuum. Table 2 shows the mean absolute

Table 2. Mean Absolute Errors (MAE) [kJ/mol] for Intermolecular Potentials of Molecular Crystals at Equilibrium Calculated with the IPA_{d+c} and the IPA_d Models[a].

Data set	IPA_{d+c}	IPA_d
Training	5.3	28.7
Test	5.3	28.2

[a]The training set and test set contained 10,000 and 1,489 data points, respectively.

error on the molecular crystals for the IPA_{d+c} and the IPA_d models. Only equilibrium structures of the crystals were included. The training set and test set contained 10,000 and 1,489 data points, respectively. As expected, the IPA_d model performs considerably worse than the IPA_{d+c} model. As the latter model performs also comparatively well on dimers in vacuum, these results point at a general advantage for the training of IPA-type models on both gas-phase and condensed-phase data. For this reason, only the IPA_{d+c} model is used in the following sections.

To obtain a picture of the range of van der Waals parameters predicted by the model, Figure 2 shows the C_6 and C_9 parameters from the IPA_{d+c} model for all pairwise interactions in the crystal data set. Further information is given in Figures S2 and S3 in the Supporting Information, showing the resulting well-depth and the minimum distance by atom pairs. The nature of the electrostatic interaction is indicated by
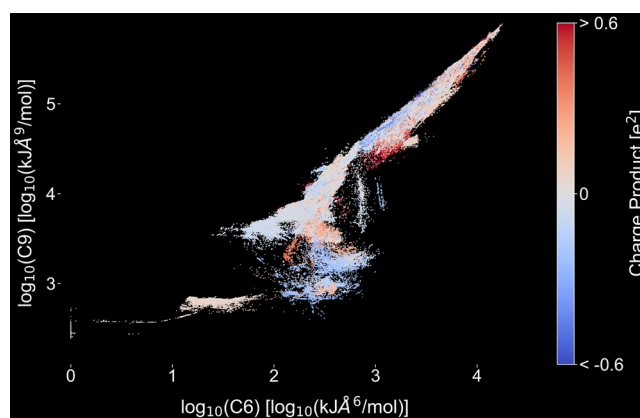

Figure 2. Predicted C_6 and C_9 parameters by the IPA_{d+c} model on log scale for all atom pairs in the data set of intermolecular potentials of crystals (11,489 molecules). The color indicates the strength and sign of the electrostatic interaction (blue = attractive, red = repulsive).

the color, with blue for attractive interactions and red for repulsive interactions. Interestingly, a large part of the correlation between C_6 and C_9 parameters might be captured by a power law. This observation could potentially be used to construct FF based on a single parameter and appropriate scaling laws, for instance based on a notion of atomic volumes. Furthermore, the predicted C_6 and C_9 parameters form a continuum over large ranges. While there are some distinct islands, in particular interactions with hydrogens in the bottom left, the results in Figure 2 are nevertheless an indication that the model takes advantage of continuous atom types.

Lattice Enthalpies: X23 and G60. Performance on crystal structures was evaluated on the X23 benchmark[112,113] using the revised values from ref 114 as well as the G60 data set[115] (Table 3). Note that these data sets differ in the way the

Table 3. Mean Absolute Error (MAE) and Mean Error (ME) in [kJ/mol] for the IPA_{d+c} Model on the X23 and G60 Benchmark Sets Based on Revised Values from ref 114 and reference values collected in ref 115[a].

Data set	MAE		ME	
	Nonrelaxed	Relaxed + intra	Nonrelaxed	Relaxed + intra
X23	7.1	6.1	4.2	2.2
G60	9.5	12.1	5.4	9.5

[a]The full error statistics are provided in Table S4 and S5 in the Supplementary Information.

experimental heat of sublimation was corrected for vibrational contributions. Two settings were considered: For setting (1), labeled with 'Nonrelaxed', the lattice energy was approximated with the intermolecular potential for the nonrelaxed experimental geometries. In setting (2), labeled with 'Relaxed + intra', structures were relaxed and the potential-energy difference between the gas-phase minimum conformation and the crystalline phase minimum was included. For the second setting, six systems had to be excluded due to problems with the bonded terms from OpenFF (i.e., X23: CO_2 and UREAXX12; G60: CTMTNA03, METNAM08, MTNANL, OCHTET13).

For both data sets, the cohesive energy is systematically underestimated with the IPA_{d+c} model (i.e., positive mean errors in Table 3) and the MAE is above chemical accuracy.

Nevertheless, the IPA$_{d+c}$ model reproduces lattice energies more accurately than most existing models reported so far in the literature, e.g., DFTB-D3 with a MAE of 10.38 kJ/mol for the X23 data set,[116] except for some of the best performing dispersion corrected DFT functionals such as PBE0-MBD with a MAE of 3.9 kJ/mol on the X23 data set.[112,113] This shows that the accurate description of the lattice energy of molecular crystals remains a particularly challenging problem, which will continue to serve as an important reality check, specifically for models that are not parametrized on condensed-phase data.

3.3. IPA$_{d+c}$ Model—Pure Liquid Properties. Properties of pure organic liquids are commonly used to validate classical FF, serving in many cases also as parametrization targets (see, e.g., refs 79, 117, 118). To further explore the performance of the IPA$_{d+c}$ model beyond vacuum and crystals, applications to the liquid phase are shown in the following. Three benchmarks covering a wide range of systems and properties are considered for this: (i) 13 sulfur compounds taken from the publication of the OPLS4 release,[119] (ii) 29 molecules containing H, C, O taken from a recent investigation of condensed-phase parametrization of OpenFF,[118] and (iii) 57 organic compounds from the validation of the GROMOS 2016H66 FF.[117] Unlike the referenced FF, parametrization of the IPA$_{d+c}$ model did not include experimental liquid properties such as the density or heat of vaporization. Therefore, pure liquid properties present an interesting test case for this model. For the liquid simulations with the IPA$_{d+c}$ model, bonded interactions (i.e., bonds, angles, dihedrals, and 1−4 nonbonded interactions) were treated with OpenFF 2.0,[120] while all other interactions (i.e., nonbonded terms) were treated with the IPA$_{d+c}$ model. A detailed description of the simulation protocol is given in section 5.13.

Sulfur Compounds. Systems containing sulfur are challenging for fixed-charge FF due to the polarizability of sulfur and the presence of higher-order multipole components. Recent work on the OPLS4 FF[119] improved the performance on several challenging motifs, including sulfur interactions and σ-holes. Specifically, OPLS4 improved the RMSE of the heat of vaporization (H_{vap}) for 13 sulfur-containing systems by more than 1 kJ/mol (i.e., 0.3 kcal/mol) compared to the previous OPLS3 version.[119] With an RMSE of 2.5 kJ/mol for H_{vap}, the IPA$_{d+c}$ model performs comparable to OPLS4 (Table 4). This

Table 4. Root-Mean-Square Error (RMSE) for the Heat of Vaporization (H_{vap}) and Density (ρ) for 13 Sulfur-Containing Compounds Investigated in Reference 119[a].

Property	IPA$_{d+c}$	OPLS4[119]
H_{vap} [kJ/mol]	2.5	2.6
ρ [kg·m^{-3}]	26.0	

[a]Values for OPLS4 from ref 119 are given as comparison and were converted from kcal/mol to kJ/mol by a factor of 4.184. Note that densities were not reported for OPLS4 in ref 119. The full error statistics and scatter plots are provided in Table S6 and Figure S4 in the Supporting Information. The individual numerical values are given in Table S11.

result is remarkable for several reasons. First, unlike the IPA model, OPLS4 uses virtual-sites to represent lone pairs and anisotropic Lennard-Jones interactions, which were specifically introduced to improve the description of systems containing sulfur and halogens. Second, the OPLS FF family was specifically parametrized with respect to liquid properties

such as H_{vap},[79] whereas the IPA$_{d+c}$ model was only trained on QM intermolecular potential energies. Third, the intramolecular potential of OPLS4 is jointly optimized with the intermolecular potential, allowing for a higher degree of consistency between the two parts. It is likely that considering the above points in future work on the IPA model could result in further improvements.

Test Systems from OpenFF. To gain a better understanding for the role of the bonded terms taken from OpenFF, results for 29 pure liquids from a recent benchmark of OpenFF are presented here.[118] The compounds contain only H, C, and O. The referenced work is particularly interesting for its investigation of opposing forces during the parametrization with respect to mixing enthalpies, vaporization enthalpies, and densities. Only pure liquid properties were considered here.

As can be seen in Table 5, similar errors are observed for the IPA$_{d+c}$ model and the standard OpenFF 1.0. Reoptimization of

Table 5. Root-Mean-Square Error (RMSE) for Pure Liquid Properties of 29 Systems Used As Test Set in ref 118[a]

Property	IPA$_{d+c}$	OpenFF 1.0[118]	OpenFF ('Pure only')[118]
H_{vap} [kJ/mol]	9.2	9.9	7.5
ρ [kg·m^{-3}]	32.1	30.0	18.0

[a]Values for OpenFF were taken from the referenced publication. The label 'Pure only' refers to a version of OpenFF 1.0, which was re-optimized on the densities and heats of vaporization of the training set in ref 118. The full error statistics and scatter plots are provided in Table S7 and Figure S5 in the Supporting Information. The individual numerical values are given in Table S12.

the FF with respect to the pure liquid properties of the training set (also compounds containing only H, C, and O) improved the accuracy of OpenFF 1.0 considerably[118] (Table 5). The authors observed thereby opposing gradient components for the simultaneous optimization with respect to densities and heats of vaporization. No further liquid properties were considered in ref 118 (such as dielectric permittivity, thermal expansion coefficient, etc.). It would thus be interesting to see the performance of the reoptimized OpenFF (termed 'Pure only') on other properties.

The fact that all models shown in Table 5 use the same functional form with very similar bonded terms may indicate that liquid properties cannot be reproduced more accurately without either improving the description of the bonded interactions and/or extending the functional form, for instance through the use of a polarizable model and multipoles. We note that the simulations with the IPA$_{d+c}$ model used bonded terms from OpenFF 2.0, while OpenFF 1.0 was employed in ref 118.

Test Systems from GROMOS 2016H66. The 57 pure liquids from the GROMOS 2016H66[117] release include extended coverage of the chemical space and additional properties such as the isobaric thermal expansion coefficient (α), the static relative dielectric permittivity (ϵ), and the isothermal compressibility (κ). For the considered properties, the IPA$_{d+c}$ model performs comparable to the 2016H66 FF (Table 6). While H_{vap} is less accurately reproduced by the IPA$_{d+c}$ model, smaller errors are observed for the remaining properties. Note that 2016H66 was parametrized on the H_{vap} and density values of 27 of the considered 57 molecules. The observation that the IPA$_{d+c}$ model outperforms 2016H66 on properties that were not used for its parametrization may be an indication that the

Table 6. Root-Mean-Square Error (RMSE) for Pure Liquid Properties of 57 Systems Used in the Calibration and Validation of the GROMOS 2016H66 FF[117a].

Property	IPA_{d+c}	GROMOS 2016H66[117]
H_{vap} [kJ/mol]	4.5	3.5
ρ [kg·m^{-3}]	26.3	32.4
α [10^{-4} K^{-1}]	1.7	4.4
ϵ [1]	12.8	14.0
κ [10^{-5} bar^{-1}]	1.8	3.6

[a]Values for GROMOS 2016H66 were taken from the referenced publication. The full error statistics and scatter plots are provided in Table S8 and Figure S6 in the Supporting Information. The individual numerical values are given in Table S13.

'bottom-up' approach of the IPA model, focusing on the reproduction of the PES, is a promising parametrization strategy. It further demonstrates that there is still room for improvement for fixed-charge FF.

3.4. ANA Model—Learned Parameters. *3.4.1. Comparison with Experimental Dispersion and Polarizability.* As the ANA approach cannot be used yet for condensed-phase simulations, we validated the ANA_d model by comparing the predicted molecular polarizabilities and intermolecular C_6 dispersion coefficients to experiment. The data set consists of molecular polarizability values for 87 compounds[121−126] and C_6 coefficients for 231 molecular pairs.[127] Since the ANA model predicts atomic parameters, the molecular polarizability is obtained as the sum of all atomic contributions. Dispersion coefficients are summed over all intermolecular atom pairs as in previous work.[90,110,128]

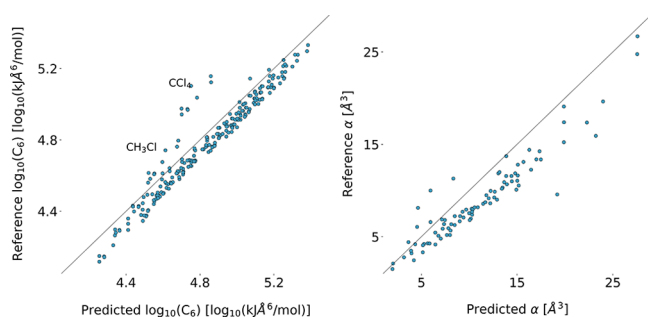Figure 3 shows the comparison between predicted and experimental values. Mean absolute relative errors of 19.4%

Figure 3. Comparison of the C_6 dispersion coefficients (right) and molecular polarizabilities (left) obtained from experiment and predicted by the ANA model. Experimental values for the molecular polarizability of 87 compounds were taken from refs 121−126, and experimental C_6 coefficients of 231 molecular pairs were taken from ref 127.

and 2.1% and Spearman correlation coefficients of 0.95 and 0.94 were observed for the C_6 coefficients and the molecular polarizabilities, respectively. For the C_6 coefficients, two sets of outliers were found. One set includes interactions with tetrachloromethane (labeled with CCl$_4$) while the second set includes interactions with chloromethane (labeled as CH$_3$Cl).

In Figure 3, we observe a systematic overestimation of C_6 coefficients by the ANA_d model, which could be a result of two factors: First, the value of the damping function in the Becke-Johnson scheme[87,129] is related to the dispersion coefficients through eqs 19 and 18. Since all dispersion coefficients are treated as free parameters, compensation effects may arise. Second, the magnitude for higher-order coefficients might be too weak compared to other components. As a result, the relationship assumed in eq 19 might not hold in the same way as for coefficients obtained from DFT densities. Similarly, a weak overestimation is found for the molecular polarizabilities. As noted previously, this may be caused by the neglect of higher-order polarization terms or by an imbalance between pairwise short-range induction, the dispersion potential, and many-body polarization terms. Inclusion of induced dipoles in the exchange potential and the parameter used for the Thole damping might also play a role. In addition, it should be noted that topology-based parameters must compensate for damping effects due to the surrounding environment, which are not taken into account but may play an important role.[130]

3.4.2. Variance of Parameters. In addition to the comparison with experimental properties, the dependence on the random initialization of the models was investigated. For this purpose, ten ANA_d models were trained using different initial weights for the underlying deep neural networks. Of those ten models, half did not converge. This phenomenon indicates that "failed" parameter sets can occur from which the training cannot recover. While not of concern in the present work, this issue might easily be fixed by conditioning, regularizing parameter ranges, pretraining on a set of known parameters, or improved model normalization. Only the five models that converged (labeled A-D) were used for the following analysis.

For the systems in the S66x8 data set, we find a relative standard deviation of 3% for the C_6 coefficients and 11% for the molecular polarizabilities. The magnitude of each FF term and the relative standard deviation of the five models A-D averaged over all equilibrium data points in the S66x8 data set are shown in Table 7. In general, the FF components and

Table 7. Magnitude and relative standard deviation (RSD) [%] of the molecular polarizability α, the cumulative C_6 coefficient, and different FF terms. All FF terms are attractive except for V^{ex}.

Model	α	C_6	V^{ele}	V^{disp}	V^{ind}	V^{ex}
	[Å^3]	[MJÅ^6/mol]	[kJ/mol]	[kJ/mol]	[kJ/mol]	[kJ/mol]
A	12.7	61.8	27.6	17.6	9.9	31.9
B	13.1	62.0	27.5	17.2	8.8	31.4
C	15.8	59.4	27.6	17.7	9.5	32.1
D	14.1	60.8	28.4	17.3	8.5	32.2
E	15.6	60.4	27.6	16.8	10.3	32.2
RSD	10.9	3.0	6.5	5.6	23.9	7.2

parameters are relatively consistent with relative deviations in the single digit percentage range. The largest deviations were observed for the molecular polarizability α and the resulting V^{ind} component. It is important to note here that V^{ind} also contains the pairwise charge-transfer term and that the resulting induced dipoles also influence V^{ex}. Especially models D and E show evidence for compensatory effects, i.e., in model D a stronger electrostatic term appears to compensate for a weaker induction, while in model E a weaker dispersion is balanced with stronger induction. Although relatively small, such fluctuations could indicate that these FF terms are underdetermined, which could in turn explain the observed issues with transferability to larger systems seen for the S7L data set.

Finally, we noticed that averaging over the predictions of all five models slightly improves the MAE for the S66x8 model to 1.0 kJ/mol compared to 1.1 kJ/mol for the single ANA_d model reported in Table 1. However, averaging over model parameters results in an increase of the error to 1.7 kJ/mol. These observation suggests a relatively fine balance between the FF terms of the model, underscoring the importance of a parameter set, which is consistent within itself. In practice, an ensemble prediction might be helpful to quantify the model uncertainty. The presence of a large variance for a specific system may warrant further training.

Overall, the results show that the ANA_d model predicts physically meaningful parameters from scratch given a physically motivated functional form. As such, including constraints through the use of predefined functional forms or known (in)equalities and parameter relations might be the method of choice to regularize ML models applied to physical problems. Further improvements may take into account the relationship between polarizabilites and dispersion coefficients with atomic volume ratios, which has been shown in several studies.[128,130−133] Using an independent model to handle polarizabilities and dispersion coefficients could not only introduce sensitivity to the surrounding environment and conformational changes but also reduce the number of fitted parameters and the resulting interdependencies.

4. CONCLUSION

In summary, the presented results attest to the power and feasibility of the proposed approach to use ML to predict the parameters of a physically motivated functional form of a FF. On one hand, the development of FF may benefit from the use of the proposed parametrization formalism due to the small computational costs of the model training, which takes only a couple of hours on a standard desktop computer, and a fully automated workflow that can take advantage of large data sets. On the other hand, our results show that the inclusion of information about the underlying physical interactions might be highly beneficial for ML models, particularly when applied to systems that exhibit a large number of weak and long-ranged interactions. Both models learned all parameters from scratch using only information about the PES obtained from reference calculations. To explore whether the ML models can truly learn parameters from the atomic environments alone, no baseline parameters were used in the present work. However, introducing such baseline parameters and additional physical models may improve the models in the future. Atomic polarizabilities and dispersion coefficients obtained from atomic volume ratios might be particularly suitable candidates for such a strategy. We found that the IPA_{d+c} model, i.e., a fixed-charge FF parametrized on intermolecular potentials of dimers in vacuum and molecular crystals from DFT calculations, can provide consistent results over a wide range of systems, ranging from dimer interaction potentials in vacuum to pure liquid properties and molecular crystals. In particular, the implicit description of the electron distribution through atomic multipoles offers a promising alternative to semiempirical models, providing a comparable or even higher accuracy for the considered test systems at smaller costs. Automatic differentiation presents a powerful tool for the development of parametrized models, which could also be applied to problems other than molecular interactions. ML-based techniques, such as the presented GNN-based atom typing, may thus ideally complement FF parametrization.

5. METHODS

In the following sections, the components used in the respective models are described. For consistency, the following notation is used: Capitalized letters refer to pairwise parameters, i.e., parameters given for a pair of atoms. Small letters are used for atomic parameters, i.e., parameters assigned to one specific atom. Subscripts are used to further clarify interaction partners. In general, the indices i, j iterate over each unique pair of atoms.

5.1. Potential-Energy Terms: IPA Model. The IPA model for the nonbonded potential energy includes three components: attractive−repulsive described with the Mie potential,[76] and the electrostatic interaction described with atomic partial charges interacting through Coulomb's law,

$$V^{\text{pot,IPA}} = V^{\text{Mie}} + V^{\text{ele}} \tag{7}$$

As the simplest form with two parameters that can reproduce the qualitative features of the dissociation of uncharged atoms, the Mie potential[76] is used,

$$V_{ij}^{\text{Mie}}(r_{ij}) = \frac{C_n}{(r_{ij})^n} - \frac{C_m}{(r_{ij})^m}, \tag{8}$$

where C_n and C_m denote the coefficients to describe the strength of the repulsion and the attraction, respectively, and r_{ij} is the distance between two atoms i and j. For $n = 12$ and $m = 6$, the well-known Lennard-Jones potential is obtained.[134] While the attractive part is often set to $m = 6$ motivated by the asymptotic behavior of the dispersion interaction, there is no such obvious choice for the repulsive part. In the past, $n = 12$ was commonly chosen for its computational efficiency. In this work, a softer $n = 9$ repulsive interaction is used as in ref 135. This choice allows for a more accurate description of the repulsive interaction while still retaining computational efficiency and comparability with the more common $C_{12}−C_6$ formulation.

Electrostatic interactions in the IPA model are treated on the basis of atomic monopoles (partial charges),

$$V_{ij}^{\text{ele}}(r_{ij}) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \tag{9}$$

with monopoles q_i and q_j, and the vacuum permittivity ϵ_0. In this work, the monopoles are obtained from our previously introduced equivariant GNN model,[67] which was trained on minimal basis iterative Stockholder multipoles (MBIS).[136]

5.2. Predicted Parameters: IPA Model. The IPA model predicts a set of C_6 and C_9 parameters for each atom pair, whose features were represented with the permutation-invariant pairwise feature combination shown in eq 6. No prior knowledge such as baseline or default parameters is used in the training of the IPA model. As mentioned above, the monopoles q_i are obtained from a separate GNN model[67] and not further modified for the present work.

The atomic monopoles and the parameters of the Mie potential are considered fixed parameters, i.e., they remain constant during a simulation and do not change in response to changes in the molecular geometry. This setting was chosen to be comparable with existing FF.[19]

5.3. Potential-Energy Terms: ANA Model. Unlike the IPA model, the second model includes explicit treatment of polarization effects and anisotropy, resulting in an anisotropic and nonadditive model, abbreviated as 'ANA' model. The

model is motivated by the desire to develop a fully classical description that performs with an accuracy expected from semiempirical methods. In addition, we aim to demonstrate the power of the proposed parametrization strategy through its application to a model with several interdependent components. At its core, the ANA model is based on an implicit description of the electronic structure through the use of atomic multipoles and a polarization model. Adding additional interaction terms allows for a more detailed decomposition of the total energy. Consistent with the decomposition by symmetry adapted perturbation theory (SAPT),[137] components for the dispersion, electrostatic, induction, and exchange potential energy are used. As a further benefit, it is possible to include SAPT terms in the fitting procedure.

The functional form of the nonbonded potential energy of the ANA model is inspired by previous work on polarizable FF and intermolecular potentials.[57,82,83,138]

$$V^{\text{pot,ANA}} = V^{\text{ele}} + V^{\text{disp}} + V^{\text{ind}} + V^{\text{ct}} + V^{\text{ex}} \tag{10}$$

The dispersion interaction (V^{disp}) is described with dispersion coefficients in conjunction with the Becke-Johnson damping model.[87,129] Following the model used in AMOEBA/AMOEBA+, induction is included through the Thole damping model (V^{ind}) and a charge-transfer potential (V^{ct}).[82,83,139,140] Exchange and electrostatic interactions (V^{ex} and V^{ele}) are based on anisotropic potentials derived for atomic multipoles following the work by Rackers et al.[84−86,140] The potential energy terms are described in more detail in the following paragraphs.

Electrostatics. The electrostatic interaction is described through the use of multipoles up to quadrupoles. Multipoles were in all cases obtained from our previously published equivariant GNN[67] for the prediction of atomic multipoles, which was trained on MBIS reference data.[136] The implementation follows the formalism in refs 141−143. According to the formalism introduced in ref 141, the total potential energy due to the interaction of point multipoles at site i and site j is obtained as

$$V_{ij}^{\text{multi}}(r_{ij}) = \sum_{l=0}^{4} B_l(r_{ij}) G^l(\vec{r}_{ij}) \tag{11}$$

with $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$ and $r_{ij} = |\vec{r}_{ij}|$. The radial functions $B_l(r)$ are defined as $B_l(r) = (2l-1)!!/r^{2l+1}$ and the coefficients $G^l(\vec{r})$ arise through the interactions between components of two multipole sites. For the considered case of a treatment up to quadrupoles, terms up to $l = 4$ are included. Coefficients $G^l(\vec{r})$ can be thought of as contributions due to the interactions of multipoles of a given order. A list of $G^l(\vec{r})$ is given in ref 141.

Deficiencies of the multipole description at short ranges are compensated through the use of the charge penetration model introduced in ref 84. In this model, effects of interactions between charge distributions are treated through the use of damping functions. Specifically, a damping function for the interaction between a charge distribution and a point charge

$$f^{\text{damp}}(r_{ij}) = 1 - \exp(-b r_{ij}) \tag{12}$$

and a damping function for the interaction between two charge distributions

$$f^{\text{overlap}}(r_{ij}) = 1 - \frac{b_j^2}{b_j^2 - b_i^2}\exp(-b_i r_{ij})$$
$$- \frac{b_i^2}{b_i^2 - b_j^2}\exp(-b_j r_{ij}) \tag{13}$$

are introduced. The damping parameter b_i describes the extent of an exponentially decaying charge distribution centered on atom i. The above damping functions give rise to damping coefficients of a given order $\lambda_l(r)$ which are given in the Supporting Information of ref 84. Combining the damping coefficients with the radial functions $B_l(r)$ gives rise to the damped radial functions

$$B_l^{\text{damp}}(r_{ij}) = \lambda_l(r_{ij}) B_l(r_{ij}) \tag{14}$$

In the charge penetration model, the standard radial functions are replaced with the damped radial functions. In addition, core−core, core−multipole, and multipole−core interactions are included. Thus, the complete description of the electrostatic potential is obtained as

$$V^{\text{ele}}(r_{ij}) = B_0 z_i z_j + \sum_{l=0}^{2} B_l^{\text{damp}}(r_{ij}) G_{CM}^l(\vec{r}_{ij})$$
$$+ \sum_{l=0}^{2} B_l^{\text{damp}}(r_{ij}) G_{MC}^l(\vec{r}_{ij})$$
$$+ \sum_{l=0}^{4} B_l^{\text{overlap}}(r_{ij}) G_{MM}^l(\vec{r}_{ij}) \tag{15}$$

with z_i representing the core charge of atom i. $B_l^{\text{damp}}(r)$ and $B_l^{\text{overlap}}(r)$ label the aforementioned damping coefficients for a single site and for a pair of charge distributions, respectively. G_{CM}^l, G_{MC}^l and G_{MM}^l describe the core-multipole, multipole-core, and multipole-multipole interactions, respectively. The core-multipole and multipole-core coefficients are obtained by replacing the monopole with the respective core charge.

Dispersion. The dispersion interaction is described based on the formalism used in the XDM model proposed by Becke and Johnson,[87]

$$V^{\text{disp}}(r_{ij}) = - \sum_{n=6,8,10} f_n(r_{ij}) \frac{C_n}{(r_{ij})^n} \tag{16}$$

with dispersion coefficients C_n and a damping function

$$f_n(r_{ij}) = \frac{(r_{ij})^n}{(R^{\text{vdW}})^n + (r_{ij})^n} \tag{17}$$

depending on a damping parameter R^{vdW}, which is obtained as follows,[129]

$$R^{\text{vdW}} = a_1 R_c + a_2 \tag{18}$$

with two positive parameters a_1 and a_2 and

$$R_c = \frac{1}{3}\left[\left(\frac{C_8}{C_6}\right)^{1/2} + \left(\frac{C_{10}}{C_6}\right)^{1/4} + \left(\frac{C_{10}}{C_8}\right)^{1/2} \right] \tag{19}$$

In the present work, a_1 was set to 1 and a_2 to 0, i.e., $R^{\text{vdW}} = R_c$.

Induction. Polarization is treated based on the Applequist model[144] including the modification proposed by Thole[139] and follows the formalism described by Stone.[4] The long-ranged

component is treated on the basis of atomic dipoles, which are induced by the external electric field and scaled by the atomic polarizability[4]

$$M^{(1),\text{ind}} = B^{-1}F^D \tag{20}$$

where F^D gathers the electric field components at each atom formed by the static multipoles of the surrounding molecules, i.e., only intermolecular contributions. The polarizability matrix B is formed as[4]

$$B = \begin{cases} \alpha_{ij}^{-1} & \text{for } i = j \\ - T_{ij} & \text{for } i \neq j \end{cases} \tag{21}$$

with the atomic polarizability α_i and the elements T_{ij} of the dipole–dipole interaction matrix. The $3N \times 3N$ polarizability matrix is inverted to obtain the induced dipoles. Given self-consistently induced dipoles, the potential energy due to induction is given as the inner product with the external field[4]

$$V^{\text{ind}} = -\frac{1}{2}M^{(1),\text{ind}}F^D \tag{22}$$

To prevent the divergence of induced dipoles ('polarization catastrophe'), elements of the polarizability matrix B are damped based on the modifications proposed by Thole analogously to the damping function used to model charge penetration effects.[139]

The exponential damping function as used in AMOEBA was used for this purpose[82]

$$f^{\text{Thole}}(r_{ij}) = 1 - \exp(-au^3(r_{ij})) \tag{23}$$

using a damping factor a and the polarizability-normalized distance

$$u(r_{ij}) = \frac{r_{ij}}{(\alpha_i\alpha_j)^{1/6}} \tag{24}$$

As in the original AMOEBA FF, the damping factor a was globally set to 0.39.[82] In addition, a charge transfer potential was added to improve the treatment of polarization at short ranges. This potential is based on work proposed for the AMOEBA+ FF[83] using an exponential form

$$V^{ct}(r_{ij}) = -A \exp(-Cr_{ij}) \tag{25}$$

where A describes the strength of the interaction and C is used to approximate the degree of electron density overlap between the respective atom pair.

Exchange. The exchange interaction is treated with the anisotropic repulsion model proposed by Rackers et al.[86] Using atomic multipoles, their work derives a description for the overlap between two atoms analogously to the electrostatic interaction between atomic multipoles leading to the following expression consistent with the expression obtained by Salem.[145] Specifically, the damping function used to construct the damped radial functions B_l^{damp} as shown above is replaced with the following damping function

$$f^{\text{damp}}(r_{ij}) = \frac{\sqrt{r_{ij}}}{b^3}\left(1 + \frac{br_{ij}}{2} + \frac{1}{3}\left(\frac{br_{ij}}{2}\right)^2\right)\exp\left(\frac{-br_{ij}}{2}\right) \tag{26}$$

for the case $b_i = b_j$ and

$$f^{\text{damp}}(r_{ij}) = \frac{1}{2X^3\sqrt{r_{ij}}}\left(b_i(rX - 2b_j)\exp\left(\frac{-b_jr_{ij}}{2}\right) \right. $$
$$\left. + b_j(rX + 2b_i)\exp\left(\frac{-b_ir_{ij}}{2}\right)\right) \tag{27}$$

for the case $b_i \neq b_j$. With $X = \left(\frac{b_i}{2}\right)^2 - \left(\frac{b_j}{2}\right)^2$ leading to the radial function for the exchange potential

$$B_0^{\text{damp}}(r_{ij}) = \frac{b_i^3 b_j^3}{r_{ij}}f^{\text{damp}}(r_{ij})^2 \tag{28}$$

with higher order radial functions $B_0^{\text{damp}}(r)$ following analogously to the damped radial functions presented in the description of the electrostatic potential.

The overlap defined as

$$S_{\text{total}}^2(r_{ij}) = \sum_{l=0} B_l^{\text{damp}}(r_{ij})G^l(\vec{r}_{ij}) \tag{29}$$

is then used to obtain the exchange potential-energy contribution,

$$V^{\text{ex}}(r_{ij}) = \frac{k_ik_j}{r_{ij}}S^2(r_{ij}) \tag{30}$$

with k_i being the relative size of atom i, and S^2 being the multipole derived orbital overlap.

We note that b used to damp the exchange interaction (eqs 26 and 27) and the electrostatic interaction (eqs 12 and 13) are treated as independent parameters despite relating to the same underlying feature, i.e., an exponentially decaying charge distribution. Induced dipoles are added to the static dipoles. As in the original work,[86] the monopole is replaced with an additional atomic parameter q^{val}, which weights the influence of the multipole interaction coefficients present in $G^l(\vec{r})$. Following ref 86, q^{val} is set to 1 for all hydrogens and >2 for all other elements. This parameter is added to the negative monopole, yielding the final q^{ex} parameter, which is used in place of the monopole used to compute the multipole interaction coefficients in $G^l(\vec{r})$.

5.4. Predicted Parameters: ANA Model. As in the IPA model, atomic multipoles were obtained using our previously developed equivariant GNN.[67] In total, five atomic parameters and five pairwise parameters were predicted by the ANA model.

Atomic parameters include the atomic polarizability (α in eq 20), the exponential factor for the electrostatic damping function (b in eqs 12 and 13), the exponential factor used in the damping function of the exchange potential (b in eqs 26, 27, and 28), the scaling factor used to adjust the strength of the exchange potential (k in eq 30), and the number of valence electrons (q^{val}), which is added to the negative atomic monopole to obtain a scalar that replaces the atomic monopole in the anisotropic exchange potential. Further, pairwise C_6, C_8, and C_{10} parameters were independently predicted for each atom pair as well as the exponents for the short-range induction potential (A and B in eq 25).

5.5. Differentiable Force Field. To achieve end-to-end differentiability, all FF terms and parametrization models were implemented in TensorFlow (version 2.6.2),[70,146] taking advantage of its automatic differentiation capabilities as well

as (batched) GPU accelerated computation. The particle-mesh-Ewald method (PME)[147,148] implemented in OpenMM (version 7.7)[149] was used to obtain the long-range electrostatic contributions for periodic systems.

5.6. Software Used. Pipelines were written with Python (3.9.5)[150] and Numpy (1.19.5).[151] Plots and visualizations were created with Matplotlib (3.5.1)[152] and Seaborn (0.11.2).[153] Trajectories were processed and analyzed with MDTraj (1.9.7).[154] RDKit (2021.09.2) was used to manipulate molecules and generate conformations.[155,156]

5.7. Graph Neural Networks (GNN). GNN models were implemented with TensorFlow (2.6.2)[70,146] and the Graph-Nets library (1.1.0)[73] using the InteractionNetwork model.[75] The parametrization model consisted of a GNN and a readout layer or combination-rule layer. Node and edge features were initially embedded as 64-dimensional vectors. The GNN consisted of independent graph-updating layers, which were composed of two fully connected feed-forward layers with 64 units, each combined with the Mila nonlinearity using $\beta = -1$.[157] Each edge and node update layer consisted of the following module ([64, Mila, 64, Mila]) for the IPA model and ([64, Mila]) for the ANA model. For both models, three graph-updating layers were used. The impact of the number of graph-updating layers on the model accuracy is shown in Figure S1 in the Supporting Information. The GNN module was followed by a readout module/combination rule para-metrized by two fully connected layers and an output layer with n output neurons equivalent to the number of predicted parameters [64, Mila, 64, Mila, n, Softplus+ϵ]. For the ANA model, a small term ($\epsilon = 10^{-3}$) was added to the output of the Softplus activation to avoid numerical instabilities, and two independent readout modules were used. Layer weights were initialized with the method introduced by He.[158]

Graph Construction. The approach applied in this work is based on graphs constructed from the molecular topology, referred to as 'topological graphs'. Topological graphs do not include information about the Euclidean distance between atoms but only atomic connectivity. Including geometrical information could be advantageous for certain applications but would require frequent recalculation of parameters, which would limit the performance. In addition, model robustness might suffer from insufficient sampling of intramolecular degrees of freedom. On the other hand, topological graphs may be ill-defined for certain cases and are unable to describe phenomena such as bond forming and breaking. For the envisioned application, i.e., the classical description of molecular motion, the shortcomings of topological graphs are acceptable while presenting a robust and efficient solution. We note that some degree of conformational dependence is present in the overall approach due to the GNN model used for the prediction of atomic multipoles.[67]

No chemical concepts such as bond types or hybridization states were included in the graphs. Hence, graphs only contained a description of the element type of an atom and the presence of a covalent bond between two atoms. Since topological information is not available for all data sets, graphs were constructed from monomer coordinates. Each graph was built by adding a node for each atom and an edge between bonded nodes. Bonds were added by first assigning hydrogen and halogen atoms to its nearest neighbors. For all other elements, all nearest neighbors within a given cutoff were assigned as bonded neighbors to the respective central atom. For C, N, O, and S, a cutoff of 2.0 Å, 1.8 Å, 1.8 Å, and 2.25 Å,

respectively, was used. Element types were encoded as one-hot vectors serving as initial node features. Edge features were built by concatenating the node features of the binding partners. Construction of the graphs is illustrated in Figure 4. We note
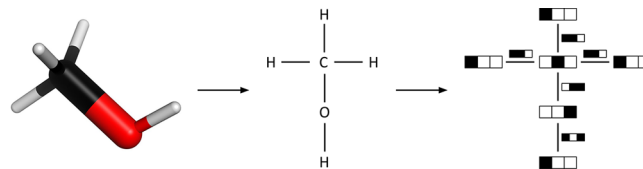

Figure 4. Example of the graph construction: For a given molecule, the topology (shown in the middle) is extracted using the heuristics described in the subsection 'Graph Construction'. The topology is turned into a graph (shown on the right), which is used as the input to the GNN. The nodes of the graph are labeled with a one-hot vector encoding the element. Each edge is labeled by the combination of the two one-hot vectors of the bonding partners. Neither distances or topological information such as bond-types or hybridization are included in the graph on the right side.

that no distance information or chemical concepts, such as bond types, were included in the graph except for the aforementioned assumptions regarding the extraction of bonded neighbors.

Loss Weighting. Models were optimized by minimizing the mean square error L between the predicted intermolecular potential energy $V^{\text{pot},\theta}$ and a reference intermolecular potential energy $V^{\text{pot,ref}}$, which was used as the target property

$$L = \frac{1}{N} \sum_i^N w_i (V^{\text{pot,ref}}(x_i)) \cdot [V^{\text{pot},\theta}(x_i) - V^{\text{pot,ref}}(x_i)]^2 \tag{31}$$

with N denoting the batch size, and i iterating over each sample of the batch. w_i is a scalar used to weight the contribution of each sample. The importance of each sample was scaled according to its Boltzmann weight in the following manner

$$w_i = \begin{cases} 1 & \text{if } r_i > r_{\text{eq}} \\ \exp[-\beta(V^{\text{pot,ref}}(x_i) - V^{\text{pot,ref}}(x_{\text{eq}}))] & \text{if } r_i \le r_{\text{eq}} \end{cases} \tag{32}$$

Samples beyond the equilibrium distance r_{eq} were weighted with $w_i = 1$, and samples closer than the equilibrium distance were weighted with the Boltzmann weight of the difference between the potential of the given sample and the potential of the equilibrium sample for a given system. The inverse-temperature β was used to determine the relative importance. To give more importance to near-equilibrium energy samples toward the end of the training procedure, an exponential decay was used to simulate annealing

$$T_n = T_0 \exp(-\gamma n) \tag{33}$$

with a decay rate γ and an initial temperature T_0 and a stopping temperature T_{min}.

Model Optimization. Model parameters were optimized using ADAM and the same exponential decay schedule used for annealing with learning rates (4×10^{-4}, 4×10^{-6}).[159] Gradients were clipped by their global norm with a clip norm of 1.[160]

5.8. Training Data. In the following sections we provide an overview of the data sets used. Generally, only neutral

molecules with elements included in {H, C, N, O, F, S, Cl} were used.

Data Set I: Dimers in Vacuum. All models were initially fitted to a recently published DES5M data set of small-molecule dimer dissociation curves published by Donchev et al.[68] Both, the SAPT0 components[137,161] and the spin-network-scaled-MP2[9,88,162] total intermolecular potential were used during training.

Data Set II: Crystal Intermolecular Potentials. To fit intermolecular potentials in the crystalline phase, a new data set with DFT energies of molecular crystals was built. The selection of the structures and the relaxation is described first. Calculations for this data set were performed on the Euler cluster of ETH Zürich.

Selection. Experimental crystal structures from the CSD, which satisfy the following requirements, were selected:

1. A single molecule in the asymmetric unit
2. Up to 100 atoms in the unit cell
3. No disorder or missing coordinates
4. Unit-cell volume up to 1600 A^3
5. Elements in {H, C, N, O, F, Cl, S}

A total of 35,577 structures were found to satisfy these requirements.

Relaxation. Of those structures, 32,811 were successfully relaxed in less than 192 CPU hours under the following settings. Structures were relaxed with the L-BFGS optimizer using QuantumEspresso (QE, 6.8) under the PBE functional and XDM dispersion correction.[90,129,163−167] Coordinates were relaxed under fixed lattice parameters with default QE settings, i.e., an energy convergence threshold of 10^{-4} Ry and a force convergence threshold of 10^{-3} a.u. The plane wave cutoff was set to 70 Ry and the charge density cutoff to 560 Ry. A uniformly spaced k-point grid was used with the number of k-points for each dimension chosen such that $n_{k_i} = \left\lceil \frac{25}{|x_i|} \right\rceil$. The publicly available projector augmented-wave (PAW) pseudo-potentials (PP)[168] were used.

Calculation of Intermolecular Potentials. For a subset of 10,489 minimized structures obtained from the previous steps, five additional geometries were generated by scaling the unit cell by factors of (0.95, 0.975, 1.0, 1.05, 1.1) without modifying the intramolecular geometry. For each of the five resulting geometries, a single-point calculation was performed using the XDM dispersion corrected B86bPBE functional.[166,169] In previous work, B86bPBE-XDM was shown to accurately reproduce lattice energies with a small systematic error.[89] PAW PP for the B86bPBE functional were generated with the pslibrary (1.0).[168] The same k-points scheme was used, while the plane wave cutoff and the charge density cutoff were set to 80 and 800 Ry, respectively. Monomers were calculated with a single k-point sampled at Γ in a cubic box with lengths chosen such that the minimal distance between atoms of the central cell and its periodic images was larger than 12 Å. The intermolecular potential energies used to train the IPA$_{d+c}$ model were then obtained as

$$\Delta V_{\text{inter}} = \frac{V_{uc}}{Z} - V_g \tag{34}$$

The crystal data set is available in the ETH research collection (10.3929/ethz-b-000549359).

5.9. General Training Strategy. The recently published DES5M data set[68] (see section 5.8) was used as main data

source for the model training (Figure 5). The data set includes spin-network-scaled MP2 (SNS-MP2)[9,88,162] intermolecular

	IPA$_{d+c}$	IPA$_d$	ANA$_d$
DES5M		Training 100'000xN Validation 13'800xN	
Crystals Intermolecular	Training 10'000x5 Validation 1'489x5		
NCI & Biofragment DB		Test 10'896	

Figure 5. Data sets used for training, validation, and testing of the IPA and ANA models. The number of molecular systems in the respective data set is given below the set indication. For dimers in the DES5M data set,[68] a variable number of samples is found, indicated by 'xN'. Details on the data sets are given in section 5.8.

potentials and SAPT0 components[137,161] for a large number of small-molecule dimers in vacuum. In total, 113,800 dimer sets were included with 100,000 sets randomly selected for training and the remaining 13,800 used for validation. Training was performed over 512 epochs. During each epoch, a total of 1024 batches were presented. Each batch contained one dimer set, i.e., all interaction potentials of one dissociation curve or set of clusters for the same two monomer molecules.

5.10. Training Strategy: IPA$_d$ Model. The IPA model was first optimized against dimer intermolecular potential energies in vacuum from the DES5M data set[68] through minimization of the expression found in eq 31, with SNS-MP2 energies serving as $V^{\text{pot,ref}}$. Annealing was performed for the loss weight term $w_i(V^{\text{pot,ref}}(x_i))$ using the described exponential decay with $T_0 = 8000$ K, $\gamma = -8 \times 10^{-3}$, and n being incremented after every epoch. Annealing was stopped at $T_{\text{min}} = 400$ K. Annealing was of particular importance for the IPA model as the attraction-repulsion potential used is not able to describe very short-ranged interactions accurately.

The IPA model was regularized with the following term

$$L_{C9} = \log(\exp(C_9 - C_6) + 1)^{-1} \tag{35}$$

During the training on dimers, an additional term $L_{C6} = C_6^{-1}$ was added. Both regularization terms were averaged over all interaction pairs of a molecule. The model obtained in this manner is referred to as IPA$_d$.

5.11. Training Strategy: IPA$_{d+c}$ Model. The IPA$_d$ model was optimized in a second step with respect to intermolecular potentials of molecular crystals. Intermolecular potentials were calculated with Quantum Espresso[163−165] using the B86bPBE functional in conjunction with the XDM dispersion correction.[89,90,129,166,167] For this step, a total of 11,489 molecules from the CSD database were used.[170] 10,000 molecules were randomly assigned to the training set, and the remaining 1489 molecules formed the validation set. Each molecule entailed five intermolecular potentials, which served as one batch. These five potentials were obtained from the relaxed crystal structure and through the expansion/contraction of the lattice of the relaxed structure. Further details on this data set are given in section 5.8.

Again, the mean square error between the predicted and the reference intermolecular potentials was minimized. The electrostatic component was independently calculated with the PME method[147] and not optimized. Fixed partial charges

were obtained as described in section 5.13. The loss was weighted with

$$w_i = \exp[-\beta(V^{\text{pot,ref}}(x_i) - V^{\text{pot,ref}}(x_{\text{eq}}))/n] \qquad (36)$$

Unlike the weighting function used for dissociation curves in eq 31, the weight was only set to one for the equilibrium structure. In addition, the potential-energy terms were scaled by the number of atoms in the respective molecule. Loss weighting is necessary since contracting/expanding the relaxed crystal structures resulted in highly unfavorable structures in certain cases. Training was performed over 512 epochs. During each epoch, 512 randomly sampled batches were presented. Each batch contained five intermolecular potentials of one specific molecule. Throughout, the temperature was set to $T_0 =$ 128 K. A cutoff of 10 Å was used for nonbonded interactions.

5.12. Training Strategy: ANA Model. Optimization of the ANA_d model followed the same procedure used to optimize the IPA_d model based on the dimer data set, except for the following differences. Since the functional form of the ANA model allows for a more accurate description of short-range interactions, the annealing schedule was modified to $T_0 = 40{,}000$ K, $\gamma = -7.5 \times 10^{-3}$, and $T_{\min} = 2000$ K. Furthermore, the L_{C_6} and L_{C_9} terms were replaced with additional loss terms for the energy components as the ANA model permits a decomposition of the total energy into components, which can be related to SAPT components.[137,161]

$$
\begin{aligned}
L_{\text{SAPT}} = \kappa \sum_\lambda \frac{1}{N} \sum_i^N w_i(V^{\text{pot,ref}}(x_i)) \\
[V_\lambda^{\text{pot},\theta}(x_i) - V_\lambda^{\text{pot,SAPT}}(x_i)]^2
\end{aligned}
\qquad (37)
$$

with λ iterating over all SAPT components, i.e., exchange, induction, electrostatic, and dispersion. The weights w_i were calculated for the total reference potential.

The SAPT loss term was scaled by a factor κ and added to the total loss in eq 31. The weight for the total energy component with respect to the SNS-MP2 calculation was kept at 1. κ was initially set to 0.5 and reduced to 0.01 after 256 epochs had passed. We note that the inclusion of the SAPT loss term is a double-edged sword. Preliminary investigations indicated that the SAPT loss term serves on one hand as a regularizer, which also accelerates convergence. On the other hand, larger values for κ limit the degree of error cancellation between potential-energy terms. It is further important to keep in mind that the accuracy of the employed SAPT0 method is lower than SNS-MP2. Specifically, for the considered subset of the DES5M data set, a MAE of 4.19 kJ/mol and 0.77 kJ/mol was found when using the weighting in eq 36 with $T = 2000$ K. With a mean error of -4.11 kJ/mol and -0.71 kJ/mol, SAPT0 overbinds relative to the SNS-MP2 results. This observation is consistent with previous benchmarks.[137]

5.13. Calculation of Condensed-Phase Properties. Condensed-phase simulations as well as the evaluation of the electrostatic potential for intermolecular potentials of crystals were performed under periodic boundary conditions using OpenMM (7.7).[149] Bonded terms were parametrized with OpenFF 2.0[120] as our model does currently not provide bonded terms. The C_6-C_9 potential was implemented using the CustomNonbondedForce class in OpenMM.

For the IPA_{d+c} model, monopoles were predicted for each conformation of an ensemble generated with the ETKDG conformation generator[156] implemented in the RDKit.[155] A

RMS pruning threshold of 0.1 Å was used, and up to 32 conformations were generated. Monopoles were predicted for each conformation using our previously introduced equivariant GNN.[67] Fixed partial charges were then obtained by averaging over all monopoles obtained for the conformational ensemble. Charges remained fixed during the simulation, and the same charges were used for the condensed-phase as well as the vacuum simulations. The 1,4-electrostatic interactions were scaled with the same factor as in OpenFF. The 1,4-Lennard-Jones interactions were described with the C_6-C_{12} parameters from OpenFF 2.0, and a scaling factor of 0.5 was used. Long-range electrostatics beyond the cutoff were treated with the smooth particle-mesh-Ewald (PME) method.[148] C_6-C_9 terms were included up to a distance of 10 Å. No long-range correction or shifting function was used for the van der Waals interactions.

Initial configurations were generated using packmol[171] and conformations generated with ETKDG.[155,156] The number of molecules was chosen such that a cubic box with side lengths 50 Å at the experimental density would be filled. Configurations were sampled from an NPT ensemble at 298.15 K using a Langevin integrator[172] with a time step of 2 fs and a collision frequency of 1 ps^{-1}. To maintain constant pressure, a Monte Carlo barostat[173] with a target pressure of 1 bar and a trial move every 25th step was applied to the system. Bonds with hydrogen atoms were fixed at the equilibrium distance with the LINCS algorithm.[174]

Each box was equilibrated for 2 ns followed by a 20 ns production run, with system data being saved to disk every 4 ps. System properties were averaged over the whole production run. To obtain the potential energy in the gas phase, a single molecule was simulated in vacuum using the same settings and simulation times as for the condensed-phase simulation.

Lattice Energy. In the present work, the lattice energy V^{lattice} is approximated as

$$V^{\text{lattice}} \approx \frac{V^{\text{pot,inter}}}{Z} \qquad (38)$$

ignoring the contribution of the intramolecular interactions. Z refers to the number of molecules in the unit cell, and $V^{\text{pot,inter}}$ to the total intermolecular potential energy for a unit cell with Z molecules under periodic boundary conditions. $V^{\text{pot,inter}}$ was calculated for the experimental geometries without relaxation. A list of CSD codes is given as Supporting Information.

Heat of Vaporization. The heat of vaporization was computed from the difference between the mean potential energy in the gas phase $\langle V^{\text{pot, gas}}\rangle$ and the mean potential energy per molecule in the condensed phase $\langle V^{\text{pot, liq}}\rangle$ corrected by a factor of RT,

$$H_{\text{vap}} = \langle V^{\text{pot,gas}}\rangle - \langle V^{\text{pot,liq}}\rangle + RT \qquad (39)$$

where R is the gas constant, and T the absolute temperature.

Density. The condensed-phase density was calculated as the total mass m_{box} in the simulation box divided by its average volume $\langle V_{\text{box}}\rangle$

$$\rho = \frac{m_{\text{box}}}{\langle V_{\text{box}}\rangle}. \qquad (40)$$

Dielectric Permittivity. Dielectric permittivities ϵ were obtained from the fluctuation of the dipole M of the system as described in ref 175,

$$\epsilon = 1 + \frac{4\pi}{3k_{\mathrm{B}}T\langle V_{\mathrm{box}}\rangle}(\langle M^2 \rangle - \langle M \rangle^2) \tag{41}$$

Isothermal Compressibility. Similarly, isothermal compressibilities κ were obtained for fluctuations of the system volume V_{box},[175]

$$\kappa = -\frac{1}{V_{\mathrm{box}}}\left(\frac{\partial V_{\mathrm{box}}}{\partial P}\right)_{N,T} \approx \frac{\langle V_{\mathrm{box}}^2 \rangle - \langle V_{\mathrm{box}}\rangle^2}{k_B T \langle V_{\mathrm{box}}\rangle} \tag{42}$$

where P is the system pressure.

Thermal Expansion Coefficient. Thermal expansion coefficients α were computed via the following relation,[175]

$$\alpha = \frac{1}{V_{\mathrm{box}}}\left(\frac{\partial V_{\mathrm{box}}}{\partial T}\right)_{N,P} \approx \frac{\langle V_{\mathrm{box}}H_l \rangle - \langle V_{\mathrm{box}}\rangle\langle H_l\rangle}{k_B T^2 \langle V_{\mathrm{box}}\rangle} \tag{43}$$

where H_l is the total enthalpy of the box.

◼ ASSOCIATED CONTENT

Data Availability Statement

The intermolecular potentials for 11,489 molecular crystals which were used during training of the IPA_{d+c} model are available in the ETH research collection (10.3929/ethz-b-000549359). A GitHub repository including the weights and models used to produce the results in this work is available at https://github.com/rinikerlab/GNNParametrizedFF.

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00661.

> Additional error statistics and reference values (PDF)
> Three text files with CSD codes used in the intermolecular crystal data set and reference values for molecular polarizabilities and C_6 dispersion coefficients (ZIP)

◼ AUTHOR INFORMATION

Corresponding Author

Sereina Riniker − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland;* ⓞ orcid.org/0000-0003-1893-4031; Email: sriniker@ethz.ch

Authors

Moritz Thürlemann − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland*

Lennard Böselt − *Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.2c00661

Notes

The authors declare no competing financial interest.

◼ ACKNOWLEDGMENTS

◼ REFERENCES

(1) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517−520.

(2) Schlick, T.; Portillo-Ledesma, S. Biomolecular Modeling Thrives in the Age of Technology. *Nat. Comput. Sci.* **2021**, *1*, 321−331.

(3) Warshel, A.; Levitt, M. Theoretical Studies of Enzymatic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(4) Stone, A. *The Theory of Intermolecular Forces*; Oxford University Press, 2013.

(5) Schrödinger, E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys. Rev.* **1926**, *28*, 1049−1070.

(6) Hartree, D. R. The Wave Mechanics of an Atom With a Non-Coulomb Central Field. Part I. *Theory and Methods* **1928**, *24*, 89−110.

(7) Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Phys.* **1930**, *61*, 126−148.

(8) Slater, J. C. The Self Consistent Field and the Structure of Atoms. *Phys. Rev.* **1928**, *32*, 339−348.

(9) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618−622.

(10) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864−B871.

(11) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133−A1138.

(12) Schuch, N.; Verstraete, F. Computational Complexity of Interacting Electrons and Fundamental Limitations of Density Functional Theory. *Nat. Phys.* **2009**, *5*, 732−735.

(13) Friesner, R. A. Ab Initio Quantum Chemistry: Methodology and Applications. *Proc. Natl. Acad. Sci.* **2005**, *102*, 6648−6653.

(14) Levine, I. N. *Quantum Chemistry*; Pearson, 2013.

(15) Pople, J. A.; Beveridge, D. L. *Molecular Orbital Theory*; Citeseer, 1970.

(16) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters. *J. Mol. Model.* **2013**, *19*, 1−32.

(17) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-Consistent-Charge Density-Functional Tight-Binding Method for Simulations of Complex Materials Properties. *Phys. Rev. B* **1998**, *58*, 7260−7268.

(18) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended Tight-Binding Quantum Chemistry Methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11*, No. e1493.

(19) Riniker, S. Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview. *J. Chem. Inf. Model.* **2018**, *58*, 565−578.

(20) Halgren, T. A.; Damm, W. Polarizable Force Fields. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−242.

(21) 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.

(22) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine Learning of Accurate Energy-Conserving Molecular Force Fields. *Sci. Adv.* **2017**, *3*, No. e1603015.

(23) Schütt, K. T.; Kindermans, P.-J.; Sauceda, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.

(24) Klicpera, J.; Groß, J.; Günnemann, S. Directional Message Passing for Molecular Graphs. *arXiv* **2020**, No. arXiv.2003.03123, DOI: 10.48550/arXiv.2003.03123.

(25) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine Learning Unifies the Modeling of Materials and Molecules. *Sci. Adv.* **2017**, *3*, No. e1701816.

(26) Schütt, K. T.; Unke, O. T.; Gastegger, M. Equivariant Message Passing for the Prediction of Tensorial Properties and Molecular Spectra. International Conference on Machine Learning. 2021; pp 9377−9388.

(27) Böselt, L.; Thürlemann, M.; Riniker, S. Machine Learning in QM/MM Molecular Dynamics Simulations of Condensed-Phase Systems. *J. Chem. Theory Comput.* **2021**, *17*, 2641−2658.

(28) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **2020**, *16*, 4192−4202.

(29) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678−3693.

(30) Wang, J.; Chmiela, S.; Müller, K.-R.; Noé, F.; Clementi, C. Ensemble Learning of Coarse-Grained Molecular Dynamics Force Fields With a Kernel Approach. *J. Chem. Phys.* **2020**, *152*, 194106.

(31) Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem., Int. Ed.* **2017**, *56*, 12828−12840.

(32) Lee, K.; Yoo, D.; Jeong, W.; Han, S. SIMPLE-NN: An Efficient Package for Training and Executing Neural-Network Interatomic Potentials. *Comput. Phys. Commun.* **2019**, *242*, 95−103.

(33) Yue, S.; Muniz, M. C.; Calegari Andrade, M. F.; Zhang, L.; Car, R.; Panagiotopoulos, A. Z. When Do Short-Range Atomistic Machine-Learning Models Fall Short? *J. Chem. Phys.* **2021**, *154*, 034111.

(34) 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. *arXiv preprint* **2022**, arXiv:2210:07237.

(35) Folmsbee, D. L.; Koes, D. R.; Hutchison, G. R. Evaluation of Thermochemical Machine Learning for Potential Energy Curves and Geometry Optimization. *J. Phys. Chem. A* **2021**, *125*, 1987−1993.

(36) Morado, J.; Mortenson, P. N.; Verdonk, M. L.; Ward, R. A.; Essex, J. W.; Skylaris, C.-K. ParaMol: A Package for Automatic Parameterization of Molecular Mechanics Force Fields. *J. Chem. Inf. Model.* **2021**, *61*, 2026−2047.

(37) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 1885−1891.

(38) Wang, L.-P.; Chen, J.; Van Voorhis, T. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *J. Chem. Theory Comput.* **2013**, *9*, 452−460.

(39) Seo, B.; Lin, Z.-Y.; Zhao, Q.; Webb, M.; Savoie, B. Topology Automated Force-Field Interactions (TAFFI): A Framework for Developing Transferable Force-Fields. *J. Chem. Inf. Model.* **2021**, *61*, 5013−5027.

(40) Jinnouchi, R.; Karsai, F.; Kresse, G. On-the-fly Machine Learning Force Field Generation: Application to Melting Points. *Phys. Rev. B* **2019**, *100*, 014105.

(41) Oliveira, M. P.; Andrey, M.; Rieder, S. R.; Kern, L.; Hahn, D. F.; Riniker, S.; Horta, B. A. C.; Hünenberger, P. H. Systematic Optimization of a Fragment-Based Force Field against Experimental Pure-Liquid Properties Considering Large Compound Families: Application to Saturated Haloalkanes. *J. Chem. Theory Comput.* **2020**, *16*, 7525−7555.

(42) Li, Y.; Li, H.; Pickard, F. C.; Narayanan, B.; Sen, F. G.; Chan, M. K. Y.; Sankaranarayanan, S. K. R. S.; Brooks, B. R.; Roux, B. Machine Learning Force Field Parameters from Ab Initio Data. *J. Chem. Theory Comput.* **2017**, *13*, 4492−4503.

(43) Zanette, C.; Bannan, C. C.; Bayly, C. I.; Fass, J.; Gilson, M. K.; Shirts, M. R.; Chodera, J. D.; Mobley, D. L. Toward Learned Chemical Perception of Force Field Typing Rules. *J. Chem. Theory Comput.* **2019**, *15*, 402−423.

(44) 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.; Gilson, M. K.; Eastman, P. K. Escaping Atom Types in Force Fields Using Direct Chemical Perception. *J. Chem. Theory Comput.* **2018**, *14*, 6076−6092.

(45) Qiu, Y.; Smith, D. G. A.; Stern, C. D.; Feng, M.; Jang, H.; Wang, L.-P. Driving Torsion Scans with Wavefront Propagation. *J. Chem. Phys.* **2020**, *152*, 244116.

(46) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0 − The Parsley Small-Molecule Force Field. *J. Chem. Theory Comput.* **2021**, *17*, 6262−6280.

(47) Beauchamp, K. A.; Behr, J. M.; Rustenburg, A. S.; Bayly, C. I.; Kroenlein, K.; Chodera, J. D. Toward Automated Benchmarking of Atomistic Force Fields: Neat Liquid Densities and Static Dielectric Constants from the ThermoML Data Archive. *J. Phys. Chem. B* **2015**, *119*, 12912−12920.

(48) Horton, J. T.; Allen, A. E. A.; Dodda, L. S.; Cole, D. J. QUBEKit: Automating the Derivation of Force Field Parameters from Quantum Mechanics. *J. Chem. Inf. Model.* **2019**, *59*, 1366−1381.

(49) Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning. *J. Chem. Theory Comput.* **2016**, *12*, 2312−2323.

(50) Vandenbrande, S.; Waroquier, M.; Speybroeck, V. V.; Verstraelen, T. The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Noncovalent Interactions. *J. Chem. Theory Comput.* **2017**, *13*, 161−179.

(51) Visscher, K. M.; Geerke, D. P. Deriving Force-Field Parameters from First Principles Using a Polarizable and Higher Order Dispersion Model. *J. Chem. Theory Comput.* **2019**, *15*, 1875−1883.

(52) McDaniel, J. G.; Schmidt, J. Physically-Motivated Force Fields from Symmetry-Adapted Perturbation Theory. *J. Phys. Chem. A* **2013**, *117*, 2053−2066.

(53) Visscher, K. M.; Geerke, D. P. Deriving a Polarizable Force Field for Biomolecular Building Blocks with Minimal Empirical Calibration. *J. Phys. Chem. B* **2020**, *124*, 1628−1636.

(54) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497−4514.

(55) van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R. Beyond Born-Mayer: Improved Models for Short-Range Repulsion in ab Initio Force Fields. *J. Chem. Theory Comput.* **2016**, *12*, 3851−3870.

(56) Bereau, T.; DiStasio, R. A.; Tkatchenko, A.; von Lilienfeld, O. A. Non-Covalent Interactions Across Organic and Biological Subsets of Chemical Space: Physics-Based Potentials Parametrized from Machine Learning. *J. Chem. Phys.* **2018**, *148*, 241706.

(57) Schriber, J. B.; Nascimento, D. R.; Koutsoukas, A.; Spronk, S. A.; Cheney, D. L.; Sherrill, C. D. CLIFF: A Component-Based, Machine-Learned, Intermolecular Force Field. *J. Chem. Phys.* **2021**, *154*, 184110.

(58) Chatterjee, P.; Sengul, M. Y.; Kumar, A.; MacKerell, A. D. Harnessing Deep Learning for Optimization of Lennard-Jones Parameters for the Polarizable Classical Drude Oscillator Force Field. *J. Chem. Theory and Comput.* **2022**, *18*, 2388−2407.

(59) Wang, Y.; Fass, J.; Chodera, J. D. End-to-End Differentiable Molecular Mechanics Force Field Construction. *arXiv* **2021**, No. arXiv.2010.01196, DOI: 10.1039/D2SC02739A.

(60) Harris, W. H. *Machine Learning Transferable Physics-Based Force Fields Using Graph Convolutional Neural Networks.* MSc. Eng. Thesis, MIT, 2020,

(61) Li, L.; Hoyer, S.; Pederson, R.; Sun, R.; Cubuk, E. D.; Riley, P.; Burke, K. Kohn-Sham equations as regularizer: Building prior knowledge into machine-learned physics. *Physical review letters* **2021**, *126*, 036401.

(62) Iten, R.; Metger, T.; Wilming, H.; del Rio, L.; Renner, R. Discovering Physical Concepts with Neural Networks. *Phys. Rev. Lett.* **2020**, *124*, 010508.

(63) Hernandez, A.; Balasubramanian, A.; Yuan, F.; Mason, S. A. M.; Mueller, T. Fast, Accurate, and Transferable Many-Body Interatomic Potentials by Symbolic Regression. *Npj Comput. Mater.* **2019**, *5*, 112.

(64) Quade, M.; Abel, M.; Shafi, K.; Niven, R. K.; Noack, B. R. Prediction of Dynamical Systems by Symbolic Segression. *Phys. Rev. E* **2016**, *94*, 012214.

(65) Udrescu, S.-M.; Tegmark, M. AI Feynman: A Physics-Inspired Method for Symbolic Regression. *Sci. Adv.* **2020**, *6*, No. eaay2631.

(66) Ma, H.; Narayanaswamy, A.; Riley, P.; Li, L. Evolving symbolic density functionals. *arXiv* **2022**, No. arXiv:2203.02540, DOI: 10.1126/sciadv.abq0279.

(67) Thürlemann, M.; Böselt, L.; Riniker, S. Learning Atomic Multipoles: Prediction of the Electrostatic Potential with Equivariant Graph Neural Networks. *J. Chem. Theory Comput.* **2022**, *18*, 1701−1710.

(68) Donchev, A. G.; Taube, A. G.; Decolvenaere, E.; Hargus, C.; McGibbon, R. T.; Law, K.-H.; Gregersen, B. A.; Li, J.-L.; Palmo, K.; Siva, K.; Bergdorf, M.; Klepeis, J. L.; Shaw, D. E. Quantum Chemical Benchmark Databases of Gold-Standard Dimer Interaction Energies. *Sci. Data* **2021**, *8*, 55.

(69) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436−444.

(70) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. *arXiv* **2015**, No. arXiv:1603.04467, DOI: 10.48550/arXiv.1603.04467.

(71) Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. *Automatic Differentiation in Pytorch*; Linux Foundations, 2017,

(72) Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; Monfardini, G. The Graph Neural Network Model. *IEEE Trans. Neural Netw. Learn. Syst.* **2009**, *20*, 61−80.

(73) Battaglia, P. W.; Hamrick, J. B.; Bapst, V.; Sanchez-Gonzalez, A.; Zambaldi, V. F.; Malinowski, M.; Tacchetti, A.; Raposo, D.; Santoro, A.; Faulkner, R.; Gülçehre, Ç.; Song, H. F.; Ballard, A. J.; Gilmer, J.; Dahl, G. E.; Vaswani, A.; Allen, K. R.; Nash, C.; Langston, V.; Dyer, C.; Heess, N.; Wierstra, D.; Kohli, P.; Botvinick, M.; Vinyals, O.; Li, Y.; Pascanu, R. Relational Inductive Biases, Deep Learning, and Graph Networks. *arXiv* **2018**, No. arXiv:1806.01261, DOI: 10.48550/arXiv.1806.01261.

(74) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. International Conference on Machine Learning. 2017; pp 1263−1272.

(75) Battaglia, P. W.; Pascanu, R.; Lai, M.; Rezende, D. J.; Kavukcuoglu, K. Interaction Networks for Learning about Objects, Relations and Physics. *Adv. Neural Inf. Process. Syst.* **2016**, *29*.

(76) Mie, G. Zur kinetischen Theorie der einatomigen Körper. *Ann. Phys.* **1903**, *316*, 657−697.

(77) 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.

(78) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712−725.

(79) 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.

(80) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.

(81) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. 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.

(82) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.

(83) Liu, C.; Piquemal, J.-P.; Ren, P. AMOEBA+ Classical Potential for Modeling Molecular Interactions. *J. Chem. Theory Comput.* **2019**, *15*, 4122−4139.

(84) Rackers, J. A.; Wang, Q.; Liu, C.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. An Optimized Charge Penetration Model for Use With the AMOEBA Force Field. *Phys. Chem. Chem. Phys.* **2017**, *19*, 276−291.

(85) Rackers, J. A.; Silva, R. R.; Wang, Z.; Ponder, J. W. Polarizable Water Potential Derived from a Model Electron Density. *J. Chem. Theory Comput.* **2021**, *17*, 7056−7084.

(86) Rackers, J. A.; Ponder, J. W. Classical Pauli Repulsion: An Anisotropic, Atomic Multipole Model. *J. Chem. Phys.* **2019**, *150*, 084104.

(87) Johnson, E. R.; Becke, A. D. A Post-Hartree-Fock Model of Intermolecular Interactions. *J. Chem. Phys.* **2005**, *123*, 024101.

(88) McGibbon, R. T.; Taube, A. G.; Donchev, A. G.; Siva, K.; Hernández, F.; Hargus, C.; Law, K.-H.; Klepeis, J. L.; Shaw, D. E. Improving the Accuracy of Møller-Plesset Perturbation Theory with Neural Networks. *J. Chem. Phys.* **2017**, *147*, 161725.

(89) Otero-de-la Roza, A.; LeBlanc, L. M.; Johnson, E. R. Dispersion XDM with Hybrid Functionals: Delocalization Error and Halogen Bonding in Molecular Crystals. *J. Chem. Theory Comput.* **2019**, *15*, 4933−4944.

(90) Otero-de-la Roza, A.; Johnson, E. R. Van der Waals Interactions in Solids Using the Exchange-Hole Dipole Moment Model. *J. Chem. Phys.* **2012**, *136*, 174109.

(91) Burns, L. A.; Faver, J. C.; Zheng, Z.; Marshall, M. S.; Smith, D. G. A.; Vanommeslaeghe, K.; MacKerell, A. D.; Merz, K. M.; Sherrill, C. D. The BioFragment Database (BFDb): An Open-Data Platform for Computational Chemistry Analysis of Noncovalent Interactions. *J. Chem. Phys.* **2017**, *147*, 161727.

(92) Burns, L. A.; Mayagoitia, A. V.; Sumpter, B. G.; Sherrill, C. D. Density-Functional Approaches to Noncovalent Interactions: A Comparison of Dispersion Corrections (DFT-D), Exchange-Hole Dipole Moment (XDM) Theory, and Specialized Functionals. *J. Chem. Phys.* **2011**, *134*, 084107.

(93) Smith, D. G. A.; Burns, L. A.; Patkowski, K.; Sherrill, C. D. Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory. *J. Phys. Chem. Lett.* **2016**, *7*, 2197−2203.

(94) Faver, J. C.; Benson, M. L.; He, X.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Kennedy, M. R.; Sherrill, C. D.; Merz, K. M. Formal Estimation of Errors in Computed Absolute Interaction Energies of Protein-Ligand Complexes. *J. Chem. Theory Comput.* **2011**, *7*, 790−797.

(95) Faver, J. C.; Benson, M. L.; He, X.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Sherrill, C. D.; Merz, K. M., Jr. The Energy Computation Paradox and ab Initio Protein Folding. *PLoS ONE* **2011**, *6*, No. e18868.

(96) Jurecka, P.; Sponer, J.; Cerný, J.; Hobza, P. Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.

(97) Parker, T. M.; Sherrill, C. D. Assessment of Empirical Models versus High-Accuracy Ab Initio Methods for Nucleobase Stacking: Evaluating the Importance of Charge Penetration. *J. Chem. Theory Comput.* **2015**, *11*, 4197−4204.

(98) Rezàc, J.; Riley, K. E.; Hobza, P. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466−3470.

(99) Rezác, J. Non-Covalent Interactions Atlas Benchmark Data Sets 5: London Dispersion in an Extended Chemical Space. *ChemRxiv* **2022**, DOI: 10.26434/chemrxiv−2022−pl3r8.

(100) Kríz, K.; Nováček, M.; Rezác, J. Non-Covalent Interactions Atlas Benchmark Data Sets 3: Repulsive Contacts. *J. Chem. Theory Comput.* **2021**, *17*, 1548−1561.

(101) Rezác, J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.* **2020**, *16*, 2355−2368.

(102) Rezác, J. Non-Covalent Interactions Atlas Benchmark Data Sets 2: Hydrogen Bonding in an Extended Chemical Space. *J. Chem. Theory Comput.* **2020**, *16*, 6305−6316.

(103) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of Quantum Chemical Methods for Large Noncovalent Complexes. *J. Chem. Theory and Comput.* **2013**, *9*, 3364−3374.

(104) Al-Hamdani, Y. S.; Nagy, P. R.; Zen, A.; Barton, D.; Kállay, M.; Brandenburg, J. G.; Tkatchenko, A. Interactions Between Large Molecules Pose a Puzzle for Reference Quantum Mechanical Methods. *Nat. Commun.* **2021**, *12*, 3927.

(105) Li, A.; Muddana, H. S.; Gilson, M. K. Quantum Mechanical Calculation of Noncovalent Interactions: A Large-Scale Evaluation of PMx, DFT, and SAPT Approaches. *J. Chem. Theory Comput.* **2014**, *10*, 1563−1575.

(106) Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M. L. The S66 × 8 Benchmark for Noncovalent Interactions Revisited: Explicitly Correlated Ab Initio Methods and Density Functional Theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20905−20925.

(107) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; D, M. A. 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.

(108) Korth, M.; Pitoňák, M.; Řezáč, J.; Hobza, P. A Transferable H-Bonding Correction for Semiempirical Quantum-Chemical Methods. *J. Chem. Theory Comput.* **2010**, *6*, 344−352.

(109) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.

(110) 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.

(111) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

(112) Otero-De-La-Roza, A.; Johnson, E. R. A Benchmark for Non-Covalent Interactions in Solids. *J. Chem. Phys.* **2012**, *137*, 054103.

(113) Reilly, A. M.; Tkatchenko, A. Understanding the Role of Vibrations, Exact Exchange, and Many-Body van der Waals Interactions in the Cohesive Properties of Molecular Crystals. *J. Chem. Phys.* **2013**, *139*, 024705.

(114) Dolgonos, G. A.; Hoja, J.; Boese, A. D. Revised Values for the X23 Benchmark Set of Molecular Crystals. *Phys. Chem. Chem. Phys.* **2019**, *21*, 24333−24344.

(115) Cutini, M.; Civalleri, B.; Corno, M.; Orlando, R.; Brandenburg, J. G.; Maschio, L.; Ugliengo, P. Assessment of Different Quantum Mechanical Methods for the Prediction of Structure and Cohesive Energy of Molecular Crystals. *J. Chem. Theory Comput.* **2016**, *12*, 3340−3352.

(116) Brandenburg, J. G.; Grimme, S. Accurate Modeling of Organic Molecular Crystals by Dispersion-Corrected Density Functional Tight Binding (DFTB). *J. Phys. Chem. Lett.* **2014**, *5*, 1785−1789.

(117) Horta, B. A. C.; Merz, P. T.; Fuchs, P. F. J.; Dolenc, J.; Riniker, S.; Hünenberger, P. H. A GROMOS-Compatible Force Field for Small Organic Molecules in the Condensed Phase: The 2016H66 Parameter Set. *J. Chem. Theory Comput.* **2016**, *12*, 3825−3850.

(118) Boothroyd, S.; Madin, O.; Mobley, D.; Wang, L.-P.; Chodera, J.; Shirts, M. Improving Force Field Accuracy by Training Against Condensed Phase Mixture Properties. *J. Chem. Theory Comput.* **2022**, *18*, 3577 DOI: 10.1021/acs.jctc.1c01268.

(119) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17*, 4291−4300.

(120) Wagner, J.; Thompson, M.; Dotson, D.; Jang, H.; Boothroyd, S.; Rodríguez-Guerra, J. *openforcefield/openff-forcefields*, ver. 2.0.0 Sage; 2021.

(121) *CRC Handbook of Chemistry and Physics*; CRC Press, 2021,

(122) Thakkar, A. J.; Wu, T. How Well Do static Electronic Dipole Polarizabilities from Gas-Phase Experiments Compare with Density Functional and MP2 Computations? *J. Chem. Phys.* **2015**, *143*, 144302.

(123) Gussoni, M.; Rui, M.; Zerbi, G. Electronic and Relaxation Contribution to Linear Molecular Polarizability. An Analysis of the Experimental Values. *J. Mol. Struct.* **1998**, *447*, 163−215.

(124) Miller, K. J. Additivity Methods in Molecular Polarizability. *J. Am. Chem. Soc.* **1990**, *112*, 8533−8542.

(125) Olney, T. N.; Cann, N. M.; Cooper, G.; Brion, C. E. Absolute Scale Determination for Photoabsorption Spectra and the Calculation of Molecular Properties Using Dipole Sum-rules. *Chem. Phys.* **1997**, *223*, 59−98.

(126) Calais, J.-L. Theory of Molecular Fluids. Volume 1: Fundamentals. *Int. J. Quantum Chem.* **1990**, *38*, 497−497.

(127) Thakkar, A. J. Dipole Oscillator Strength Distributions, Sum Rules, Mean Excitation Energies, and Isotropic van der Waals Coefficients for Benzene, Pyridazine, Pyrimidine, Pyrazine, s-Triazine, Toluene, Hexafluorobenzene, and Nitrobenzene. *J. Chem. Phys.* **2020**, *153*, 124307.

(128) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.

(129) Johnson, E. R.; Becke, A. D. A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections. *J. Chem. Phys.* **2006**, *124*, 174104.

(130) Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.

(131) Gould, T. How Polarizabilities and C6 Coefficients Actually Vary with Atomic Volume. *J. Chem. Phys.* **2016**, *145*, 084308.

(132) Kannemann, F. O.; Becke, A. D. Atomic Volumes and Polarizabilities in Density-Functional Theory. *J. Chem. Phys.* **2012**, *136*, 034109.

(133) Manz, T. A.; Chen, T.; Cole, D. J.; Limas, N. G.; Fiszbein, B. New Scaling Relations to Compute Atom-In-Material Polarizabilities and Dispersion Coefficients: Part 1. Theory and Accuracy. *RSC Adv* **2019**, *9*, 19297−19324.

(134) Jones, J. E.; Chapman, S. On the Determination of Molecular Fields. I. From the Variation of the Viscosity of a Gas With Temperature. *Proc. R. Soc. Lond.* **1924**, *106*, 441−462.

(135) Sun, H. COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase Applications Overview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B* **1998**, *102*, 7338−7364.

(136) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development. *J. Chem. Theory Comput.* **2016**, *12*, 3894−3912.

(137) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of Symmetry Adapted Perturbation Theory (SAPT). I. Efficiency and Performance for Interaction Energies. *J. Chem. Phys.* **2014**, *140*, 094106.

(138) Mueller, T.; Hernandez, A.; Wang, C. Machine Learning for Interatomic Potential Models. *J. Chem. Phys.* **2020**, *152*, 050902.

(139) Thole, B. Molecular Polarizabilities Calculated With a Modified Dipole Interaction. *Chem. Phys.* **1981**, *59*, 341−350.

(140) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14*, 5273−5289.

(141) Burnham, C. J.; English, N. J. A New Relatively Simple Approach to Multipole Interactions in Either Spherical Harmonics or Cartesians, Suitable for Implementation into Ewald Sums. *Int. J. Mol. Sci.* **2020**, *21*, 277.

(142) Smith, W. Point Multipoles in the Ewald Summation (Revisited). *Information Newsletter for Computer Simulation of Condensed Phases* **1998**, 15−25.

(143) Lin, D. Generalized and Efficient Algorithm for Computing Multipole Energies and Gradients Based on Cartesian Tensors. *J. Chem. Phys.* **2015**, *143*, 114115.

(144) Applequist, J.; Carl, J. R.; Fung, K.-K. Atom Dipole Interaction Model for Molecular Polarizability. Application to Polyatomic Molecules and Determination of Atom Polarizabilities. *J. Am. Chem. Soc.* **1972**, *94*, 2952−2960.

(145) Salem, L.; Longuet-Higgins, H. C. The Forces Between Polyatomic Molecules. II. Short-Range Repulsive Forces. *Proc. R. Soc. Lond.* **1961**, *264*, 379−391.

(146) *TensorFlow*; TensorFlow.org, DOI: 10.5281/zenodo.4724125, 2021.

(147) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(148) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(149) 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.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.

(150) Van Rossum, G.; Drake, F. L. *Python 3 Reference Manual*; Python Software Foundation, 2009.

(151) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array Programming with NumPy. *Nature* **2020**, *585*, 357−362.

(152) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **2007**, *9*, 90−95.

(153) Waskom, M. L. Seaborn: Statistical Data Visualization. *J. Open Source Softw.* **2021**, *6*, 3021.

(154) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528−1532.

(155) Landrum, G.; Tosco, P.; Kelley, B. et al.. *rdkit/rdkit*, 2020_09_5 (Q3 2020) Release; npm.Inc., 2021.

(156) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55*, 2562−2574.

(157) Misra, D. Mish: A Self Regularized Non-Monotonic Neural Activation Function. *arXiv* **2019**, No. arXiv:1908.08681.

(158) He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* **2015**, 123.

(159) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, No. arXiv:1412.6980.

(160) Pascanu, R.; Mikolov, T.; Bengio, Y. Understanding the Exploding Gradient Problem (On the difficulty of training Recurrent Neural Networks). *arXiv* **2012**, No. arXiv:1211.5063v2.

(161) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887−1930.

(162) Grimme, S. Improved Second-Order Møller-Plesset Perturbation Theory by Separate Scaling of Parallel- and Antiparallel-Spin Pair Correlation Energies. *J. Chem. Phys.* **2003**, *118*, 9095−9102.

(163) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Nardelli, M. B.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M.; Colonna, N.; Carnimeo, I.; Corso, A. D.; de Girondoli, S.; Delugas, P.; DiStasio, R. A.; Ferretti, A.; Floris, A.; Fratesi, G.; Fugallo, G.; Gebauer, R.; Gerstmann, U.; Giustino, F.; Gorni, T.; Jia, J.; Kawamura, M.; Ko, H.-Y.; Kokalj, A.; Küçükbenli, E.; Lazzeri, M.; Marsili, M.; Marzari, N.; Mauri, F.; Nguyen, N. L.; Nguyen, H.-V.; de-la Roza, A. O.; Paulatto, L.; Poncé, S.; Rocca, D.; Sabatini, R.; Santra, B.; Schlipf, M.; Seitsonen, A. P.; Smogunov, A.; Timrov, I.; Thonhauser, T.; Umari, P.; Vast, N.; Wu, X.; Baroni, S. Advanced Capabilities for Materials Modelling with Quantum ESPRESSO. *J. Phys.: Condens. Matter* **2017**, *29*, 465901.

(164) Giannozzi, P.; Baseggio, O.; Bonfâ, P.; Brunato, D.; Car, R.; Carnimeo, I.; Cavazzoni, C.; de Girondoli, S.; Delugas, P.; Ferrari Ruffino, F.; Ferretti, A.; Marzari, N.; Timrov, I.; Urru, A.; Baroni, S. Quantum ESPRESSO Toward the Exascale. *J. Chem. Phys.* **2020**, *152*, 154105.

(165) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Corso, A. D.; de Girondoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M. QUANTUM ESPRESSO: A Modular and Open-Source Software Project for Quantum Simulations of Materials. *J. Phys.: Condens. Matter* **2009**, *21*, 395502.

(166) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(167) Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew-Burke-Ernzerhof Exchange-Correlation Functional. *J. Chem. Phys.* **1999**, *110*, 5029−5036.

(168) Dal Corso, A. Pseudopotentials Periodic Table: From H to Pu. *Comput. Mater. Sci.* **2014**, *95*, 337−350.

(169) Becke, A. D. On the Large-Gradient Behavior of the Density Functional Exchange Energy. *J. Chem. Phys.* **1986**, *85*, 7184−7187.

(170) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Cryst. B* **2016**, *72*, 171−179.

(171) 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.

(172) Izaguirre, J. A.; Sweet, C. R.; Pande, V. S. Multiscale Dynamics of Macromolecules Using Normal Mode Langevin. *Pac. Symp. Biocomput.* **2009**, 240−251.

(173) Chow, K.-H.; Ferguson, D. M. Isothermal-Isobaric Molecular Dynamics Simulations with Monte Carlo Volume Sampling. *Comput. Phys. Commun.* **1995**, *91*, 283−289.

(174) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(175) Fennell, C. J.; Li, L.; Dill, K. A. Simple Liquid Models with Corrected Dielectric Constants. *J. Phys. Chem. B* **2012**, *116*, 6936−6944.