

Transferable Atomic Multipole Machine Learning Models for Small Organic Molecules

Tristan Bereau,^{*,†} Denis Andrienko,[†] and O. Anatole von Lilienfeld^{*,‡,§}

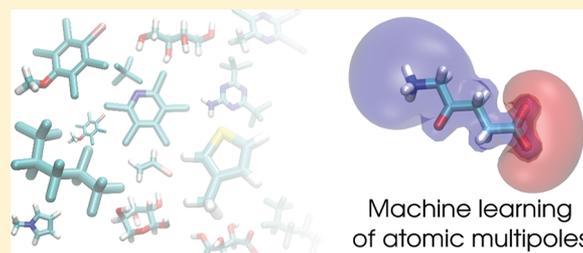
[†]Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

[‡]Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials, Department of Chemistry, University of Basel, 4056 Basel, Switzerland

[§]Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, Illinois 60439, United States

S Supporting Information

ABSTRACT: Accurate representation of the molecular electrostatic potential, which is often expanded in distributed multipole moments, is crucial for an efficient evaluation of intermolecular interactions. Here we introduce a machine learning model for multipole coefficients of atom types H, C, O, N, S, F, and Cl in any molecular conformation. The model is trained on quantum-chemical results for atoms in varying chemical environments drawn from thousands of organic molecules. Multipoles in systems with neutral, cationic, and anionic molecular charge states are treated with individual models. The models' predictive accuracy and applicability are illustrated by evaluating intermolecular interaction energies of nearly 1,000 dimers and the cohesive energy of the benzene crystal.



1. INTRODUCTION

The efficient evaluation of intermolecular interactions^{1–3} is an essential part of all classical molecular dynamics simulations. Electrostatic, induction, dispersion, and exchange repulsion are the most frequently encountered nonbonded contributions to the energy of interaction between molecules. In order to boost computational efficiency, these contributions are often projected on pairwise-additive functions, the sum of which then approximates the potential energy surface of a molecular assembly. Many-body effects (e.g., induction and dispersion) are accounted for effectively, by an appropriate parametrization of the potential energy surface. These parametrizations are, by construction, state-point-dependent and rely on either measured or first-principles-evaluated thermodynamic properties of a molecular assembly at a certain state point. For example, partial charges and Lennard-Jones parameters are often adjusted to fit the density, heat of vaporization, and other thermodynamic properties.⁴

Force field transferability and accuracy can of course be improved by retaining many-body contributions. The decisive advantage of this approach, which justifies extra computational effort, is that these terms can be evaluated perturbatively, i.e., by first calculating electronic properties of noninteracting molecules using first-principles methods and then accounting for electrostatic (first order), induction (second order), and dispersion (higher orders) contributions in a perturbative way.⁵ Such parametrizations do not require experimental input, are state-point-independent and, as such, can be used to prescreen chemical compounds *in silico*.

In this approach, however, even the molecular electrostatic potential must be evaluated for every single molecular

conformation, requiring electronic structure calculations at practically every molecular dynamics step. It has also been pointed out that the multipole-moment (MTP) description of the electrostatics must include not only atomic charges but also higher moments (e.g., dipoles and quadrupoles),^{6–9} improving free-energy calculations,^{10,11} spectroscopic signatures,^{12,13} and dynamics.¹⁴

Avoiding the need for frequent quantum-chemical calculations has motivated the development of fast prediction methods, such as machine learning (ML).^{15–17} With ML we refer to statistical algorithms that extract correlations by training on input/output data and that improve in predictive power as more training data are added.¹⁸ While ML models for the fitting of potential energies have been in use for decades,¹⁹ the possibility to infer point charges, MTPs, and polarizabilities has been investigated only recently.^{20–23} These approaches interpolate between a large number of conformations to accurately describe the effects of changes in the geometry. The accuracy that is reached comes at a price: The specificity of the learning procedure limits its applicability to the given molecule of interest. Instead of training electrostatic models for every new molecule, here we construct a *transferable* MTP model which can be applied not only to different molecular conformers but also atom types.

This work is organized as follows. We first describe how to build a machine learning MTP model that predicts static, atomic point charges, dipoles, and quadrupoles for H, C, O, N, S, F, and Cl atom types in specific chemical environments.

Received: March 31, 2015

Next, the resulting electrostatic interactions are combined with a classical many-body dispersion (MBD)²⁴ in order to validate the model by estimating intermolecular energies of nearly 1,000 molecular dimers as well as the cohesive binding energy of the benzene crystal. We find that the machine learning model retains an accuracy similar to the same model parametrized from individual quantum-chemical calculations.

2. METHODS

The following describes the ML model, the baseline property used in the Δ -learning procedure, the data set, and the description of the reference MTPs.

2.1. Machine Learning Model. We rely on supervised learning to construct a kernel-ridge regression which generalizes the linear-ridge regression model (i.e., linear regression with regularizer λ) by mapping the input space x into a higher dimensional “feature space”, $\phi(x)$, thereby casting the problem in a linear way.^{16,25} The strength of the method comes from avoiding the actual determination of ϕ thanks to the so-called kernel trick.²⁶ Since the ML algorithm only requires the inner product between data vectors in feature space, one can apply a kernel function $k(x, x')$ to compute dot products within input space, thereby leaving the feature space entirely implicit. As a result, the problem is reformulated from a v -dimensional input space (i.e., the dimensionality of each data vector) into an n -dimensional space spanned by the number of samples in the training set. This characteristic implies that the larger n is, the better the prediction ought to be—thus the denomination of a *supervised learning* method.

Here, we build on the Δ -ML approach,²⁷ which estimates the difference between the desired property and an inexpensive baseline model that accounts for the most relevant physics. More specifically, a refined target property $p(x)$ is predicted from baseline property p^{Vor} (see section 2.3) plus the ML-model Δ

$$p(x) = p^{\text{Vor}}(x) + \Delta(x, p^{\text{Vor}}) \quad (1)$$

where x corresponds to the representation vector—or descriptor—of the input sample (e.g., query molecule). Δ corresponds to the standard kernel-ridge regression model of the difference between baseline and target property constructed for n training samples,

$$\Delta(x, p^{\text{Vor}}) = \sum_{i=1}^n \alpha_i [k(x, x_i) + k'(p^{\text{Vor}}, p_i^{\text{Vor}})] \quad (2)$$

where α_i is the weight given to training molecule i . These weights are determined by best reproducing the reference property $p^{\text{ref}}(x)$ for each sample in the training set according to the closed-form solution $\alpha = (\mathbf{K} + \mathbf{K}' + \lambda \mathbf{I})^{-1}(\mathbf{p}^{\text{ref}} - \mathbf{p}^{\text{Vor}})$, where $\mathbf{p}^{\text{ref}} - \mathbf{p}^{\text{Vor}}$ is the vector of training properties, i.e., the difference between reference and baseline, and \mathbf{K} and \mathbf{K}' are the two kernel matrices. Note that, in eq 2, we have included representation *and* baseline property in the kernel, each having a different width in their respective kernel functions.

ML maps an input representation vector \mathbf{x} into a scalar value of similarity. Thus, before applying ML to predict atomic MTPs, the information contained in the three-dimensional structure of a molecule must be encoded in a vector of numbers, i.e., its representation or descriptor. Ideally, this information should reflect symmetries of molecular structures with respect to rotations, translations, reflections, and atom

index permutations, etc. Here, we rely on the Coulomb-matrix descriptor,²⁸

$$C_{ij} = \begin{cases} \frac{1}{2} Z_i^{2.4} & \forall i = j \\ \frac{Z_i Z_j}{|\mathbf{R}_i - \mathbf{R}_j|} & \forall i \neq j \end{cases} \quad (3)$$

where i and j are index atoms in the molecule, Z_i is atom i 's atomic number, and \mathbf{R}_i represents its Cartesian coordinates. Note that the Coulomb matrix not only encodes inverse pairwise distances between atoms but also the chemical elements involved. As different molecules have different numbers of atoms, their Coulomb matrices will vary in size. Distant neighbors are expected to have a comparatively small impact on a prediction, such that the inclusion of all neighbors can prove inefficient for large molecules. Given a set of molecules, we pad matrices with zeros such that their size amounts to $n \times n$, where n is the number of closest neighboring atoms considered.²⁸ In the following, we set $n = 4$. Given a molecule's d atoms, there are d individual atomic MTP samples for the ML to learn from. For each, an individual Coulomb matrix is built in which the atom of interest fills up the first row/column, while the indices of the surrounding n atoms are sorted according to the atoms' Euclidean distances to the query atom. As such, we coarsen our descriptor to contain at least the first shell of n covalently bound neighbors, and atoms that only differ in their environment at larger distances will be assigned the same MTP. We have found $n = 4$ to correspond to a reasonable compromise between computational efficiency and performance. Note, however, that while such choices of descriptor typically do affect the model's performance for given training sets, other descriptor choices could work just as well—as long as they meet the requirements and invariances necessary for the ML of quantum properties.²⁹

In the context of applying ML to the prediction of tensorial quantities, such as MTPs, properties $p^{\text{Vor}}(x)$ and $p(x)$ will be expressed as *vectors* of size m —the number of independent coefficients of the tensor of interest (e.g., 1 for a scalar charge, 3 for a vector dipole moment, 5 for a traceless second-rank tensor quadrupole). We express MTP moments with their minimal number of independent coefficients by using the spherical-coordinate representation. We recognize that the kernel matrices, \mathbf{K} and \mathbf{K}' , will remain unmodified when learning/predicting different tensor components of the same input data vector. Finally, the weights α are expressed as a matrix of size $m \times n$, which naturally reduces to a vector when predicting a scalar quantity.

For this work, we have used the Laplacian kernels,

$$k(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|}{\sigma N_t}\right) \quad (4)$$

$$k'(p_i^{\text{Vor}}, p_j^{\text{Vor}}) = \exp\left(-\frac{|p_i^{\text{Vor}} - p_j^{\text{Vor}}|}{\zeta N_t}\right) \quad (5)$$

where σ and ζ are free parameters and $|\dots|$ corresponds to the Manhattan, or city block, L_1 norm. This combination of kernel functions and distance measure has previously been shown to yield the best performance for the modeling of molecular atomization energies and other electronic properties using the Coulomb-matrix representation.^{30,31} N_t is the number of

occurrences of the chemical element type to which atom i belongs. As a result N_i normalizes the width to be consistent with the training set size of a given chemical element. We report later (Table 2) the strong variance of occurrence numbers of chemical elements in the employed training set. Hyperparameter optimization on 85% of the elements encompassing the training set (see later) yielded $\sigma = 0.005$, $\zeta = 0.002$, and $\lambda = 10^{-9}$. We have subsequently used these parameters for element-specific models throughout. Combining eqs 2, 4, and 5, the Δ -learning ML model predicts the deviation from the Voronoi baseline for a new query atom x of element t with Voronoi property p^{Vor} according to

$$\Delta(x, p^{\text{Vor}}) = \sum_{i=1}^n \alpha_i (e^{-|x-x_i|/\sigma N_i} + e^{-|p^{\text{Vor}}-p_i^{\text{Vor}}|/\zeta N_i}) \quad (6)$$

For the modeling of MTP in positively and negatively charged molecules ($\pm 1 e$), we have trained respectively different ML models for the same set of molecules, systematically assigning the corresponding global molecular charge and assuming a doublet state.

2.2. Multipole Moments. Molecular electron densities were computed using density functional theory calculations at the M06-2X level of theory³² and cc-pVDZ basis set.³³ All *ab initio* calculations were performed using the Gaussian09 program.³⁴

The generalized distributed multipole analysis (GDMA)⁵ allowed us to partition the density into atomic MTPs up to quadrupoles, where we relied on grid-based quadrature (i.e., switch value of 4). The same protocol was applied to train the ML models for positively and negatively charged molecules, after reassigning the global charge of each molecule.

The reference MTPs, obtained from the distributed multipole analysis, were rotated into a molecular reference frame, which was constructed from the (sorted) eigenvectors of the molecule's moment of inertia tensor centered at the atom in question. To ensure uniqueness, we set the positive axis of each vector such that its scalar product with the vector pointing from the atom of interest to the molecule's center of mass is positive. For linear (e.g., diatomic) molecules, we assign the interatomic direction as the first axis and arbitrarily construct two orthogonal axes. After the ML prediction, we rotated back the MTPs in the original, global frame.

All MTP interactions were computed in CHARMM³⁵ using the MTPL module,^{11,36} while our in-house code used for the many-body dispersion energies²⁴ is freely available online.³⁷

2.3. Voronoi Partitioning of the Charge Density. The Voronoi baseline model relies on a systematic, geometry-dependent estimation of a system's underlying charge density. Reference atomic MTP coefficients are extracted from the partitioning of said density (see later text for details), where monopole, dipole, and quadrupole contributions are given by

$$q^{(i)} = \int d\mathbf{r} n_i(\mathbf{r}) \quad (7)$$

$$\mu_\alpha^{(i)} = \int d\mathbf{r} n_i(\mathbf{r}) r_\alpha \quad (8)$$

$$Q_{\alpha\beta}^{(i)} = \int d\mathbf{r} n_i(\mathbf{r}) r_\alpha r_\beta \quad (9)$$

respectively, where $n_i(\mathbf{r})$ denotes the *partitioned* density attributed to atom i , as a function of spatial coordinate \mathbf{r} , and $\alpha, \beta \in \{x, y, z\}$. Rather than being derived from quantum-

chemical calculations, $n_i(\mathbf{r})$ is constructed as a Gaussian-based atomic density

$$n_i(\mathbf{r}) = \frac{1}{(2\pi r_i^2)^{3/2}} \exp\left(-\frac{|\mathbf{r} - \mathbf{R}_i|^2}{2r_i^2}\right) \quad (10)$$

where \mathbf{R}_i is the position of nucleus i and r_i is the chemical element's free-atom radius which is fixed independent of molecular environment or geometry. For this, we have used parameters reported elsewhere.^{38,39} Atomic densities $\{n_i(\mathbf{r})\}$ are partitioned according to a Voronoi scheme,⁴⁰ whereby only the closest atom contributes to a given spatial coordinate. The Euclidean distance provides the distance metric to identify a region \mathcal{R}_p associated with atom p

$$\mathcal{R}_p = \{\mathbf{r} \in \mathbb{R}^3 | d(\mathbf{r}, \mathbf{r}_p) \leq d(\mathbf{r}, \mathbf{r}_j) \text{ for all } j \neq p\} \quad (11)$$

Figure 1 illustrates the Voronoi-based density partitioning between the atoms of water. Each color corresponds to the

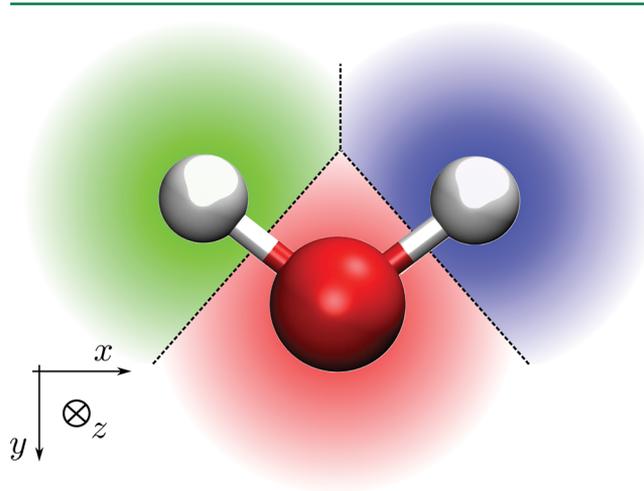


Figure 1. Diagram of Voronoi-based density partitioning for a water molecule (rendered with VMD⁴¹). Dashed lines delineate the partition boundaries. The axis system illustrates the orientation of the water molecule's global frame for the calculations presented in section 3.1: aligned in the xy plane with the oxygen atom pointing toward the y axis.

atomic density of the corresponding atom. We recently introduced a similar protocol to effectively estimate atomic polarizabilities which serve as input for many-body dispersion interactions.²⁴

Note that the Voronoi model contains no free parameter—the free-atom radii being applied without prior fitting. Though the model hardly reproduces any of the reference MTP coefficients, it provides a *qualitative* evaluation of the coefficients. In particular, the baseline model reproduces elementary symmetries of the system that are entirely determined by the geometry; e.g., a planar molecule cannot generate an orthogonal dipole moment.

While we compute eqs 7, 8, and 9 in Cartesian coordinates, we subsequently convert them to their spherical counterparts $Q_{\kappa m}$ where κ denotes the rank (e.g., $\kappa = 0$ for monopoles) and m indexes the (real) component of the MTP (see Stone⁵ for more details). Given a molecular structure, we estimate for each atom the baseline property $p^{\text{Vor}} = \{Q_{00}, Q_{1m}, Q_{2m}\}$, where m runs over the $2\kappa + 1$ elements of an MTP of order κ .

Table 1. MTP Coefficients of Oxygen in Water (See Figure 1) and in the Carbonyl Bond of Formic Acid, as Well as the MTPs of Carbon in Benzene^a

	Q_{00}	Q_{10}	Q_{11c}	Q_{11s}	Q_{20}	Q_{21c}	Q_{21s}	Q_{22c}	Q_{22s}
Water Oxygen									
Vor	0.04	0	0	-0.12	0.43	0	0	-0.21	0.01
Δ -ML	-0.13	-0.01	-0.02	0	0.16	-0.03	0.04	-0.18	0
ref	-0.39	0	0.01	-0.40	-0.92	0	0	0.45	0
Formic Acid Carbonyl Oxygen									
Vor	0.01	0	0	0	0.08	0	0	0	0
Δ -ML	-0.35	-0.30	-0.03	0.03	0.55	0.10	-0.04	-0.10	0.15
ref	-0.45	-0.10	-0.15	0	0.38	0.13	0	-0.32	0
Benzene Carbon									
Vor	0.01	-0.01	0.01	0	-0.01	0	0	0.07	0.01
Δ -ML	-0.10	-0.04	-0.10	-0.09	-0.73	-0.24	0.19	0.17	-0.05
ref	-0.03	0	0	0	-0.65	0	0	-0.14	0

^a“Vor”, “ Δ -ML”, and “ref” correspond to the Voronoi-based property evaluation, the Δ -learned prediction across chemical compound space (see the Δ -ML-85 model in text), and the ab initio data, respectively. All coefficients are expressed in units of $e \text{ \AA}^l$, where l is the MTP’s rank. The comparatively low accuracy of the Δ -ML model for water is rationalized in the main text.

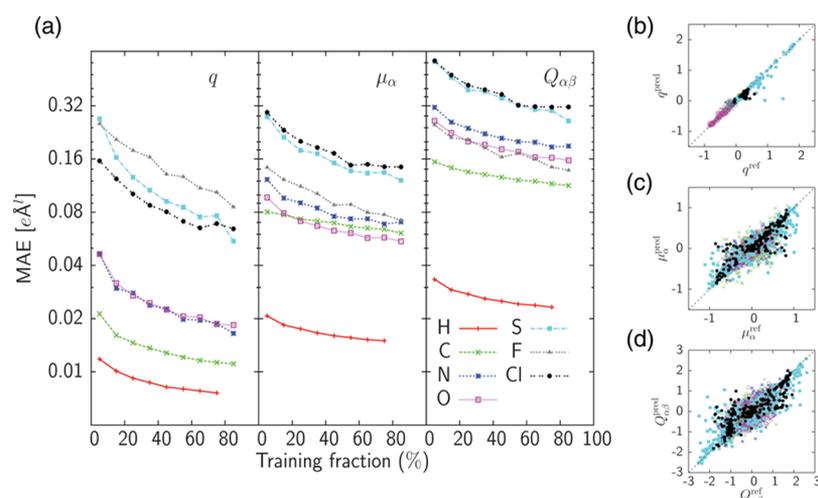


Figure 2. (a) Mean absolute prediction error (MAE) as a function of training set percentage of an 82k-atom database obtained from neutral molecules. MAEs are given in $e \text{ \AA}^l$ units with l being the MTP rank. Errors correspond to MTP Δ -ML model predictions for individual components of atomic monopole, dipole, and quadrupole moments shown in left, middle, and right panels, respectively, for all elements present. Scatter correlation plots for all components of (b) monopoles, (c) dipoles, and (d) quadrupoles, as predicted with the Δ -ML-85 model. Colors correspond to the atom legend in panel a. The outliers in the monopole correlation curve correspond to sulfur oxide groups (see main text).

2.4. Molecular Data Set. To refine atomic properties beyond the baseline prediction, we train the transferable ML algorithm on 3,167 neutral molecules obtained from the Ligand.Info database,⁴² totaling 82.1 kiloatoms. Atoms have been segregated between training and prediction pools randomly. The database provides three-dimensional coordinates of small, organic molecules. We focus exclusively on a subset of neutral molecules that include elements H, C, N, O, S, F, and Cl. The atomic coordinates of all molecules are provided in the Supporting Information.

3. RESULTS

3.1. Voronoi-Based Baseline Evaluation. To illustrate the applicability of the Voronoi-based baseline evaluation of MTP coefficients, we compare its prediction with the reference MTP coefficients obtained from *ab initio* methods. Given that MTPs are inherently axis-system-dependent (apart from the monopole), we first describe the global frame used for the water molecule in Figure 1. Inherent symmetries of the geometry impose some coefficients to be zero; e.g., there can be

no dipole moment along the x or z directions due to the molecule’s C_{2v} symmetry. While the Voronoi baseline does not even qualitatively reproduce the nonzero coefficients—due to the method not entailing *any* prior parametrization—its ability to impose the right symmetry is very desirable. The same kind of behavior is also shown for a carbon atom on benzene or the carbonyl oxygen of formic acid in Table 1. For comparison, Table 1 also shows already the corresponding ML augmented MTP result. As such, the baseline recovers an important aspect of the underlying symmetry, which the augmenting ML model no longer needs to account for.

3.2. Δ -ML MTP Model Trained and Tested across Chemical Space. In principle, the previously mentioned Coulomb matrix encodes enough chemistry to train all chemical elements. Memory limitations of the kernel-ridge regression, however, make atom-type specific models better tractable. We now investigate the Δ -ML model’s capabilities to predict MTP coefficients *across* chemical space, one for each of those chemical elements that are most frequent in small, organic molecules (see earlier text). The ML model has been

Table 2. MAE for Each Chemical Element and MTP Rank of the ML-85 and Δ -ML-85 Transferable Models (Neutral Molecules), Corresponding to an 85% Training-Set Size for All Elements but H (only 75%)^a

	no. of atoms		MAE [$e \text{ \AA}^l$]					
			q		μ_α		$Q_{\alpha\beta}$	
	training (N_t)	prediction	ML-85	Δ -ML-85	ML-85	Δ -ML-85	ML-85	Δ -ML-85
H	28,822	9,607	0.01	0.01	0.03	0.01	0.04	0.02
C	24,356	4,297	0.01	0.01	0.05	0.05	0.18	0.09
N	4,054	715	0.02	0.02	0.09	0.05	0.26	0.15
O	6,134	1,082	0.02	0.02	0.08	0.04	0.22	0.12
F	363	63	0.03	0.09	0.04	0.05	0.14	0.11
S	1,542	272	0.05	0.05	0.12	0.09	0.31	0.20
Cl	739	130	0.03	0.06	0.11	0.10	0.28	0.26
	66,010	16,166	0.02	0.04	0.08	0.06	0.20	0.13

^aThe last lines are averages over all chemical elements. The second column denotes the number of molecules for which the MTP moments have been predicted.

trained on various fractions of the considered data set's 82 kiloatoms.

Figure 2a displays error saturation curves for individual chemical elements. These monotonically decaying learning curves are presented as a function of training size of the data set, where the predicted mean absolute error (MAE) is calculated across the *remaining* atoms not included in the training set. The finding of monotonically decreasing error with training set size represents numerical evidence for a crucial feature of the supervised learning working hypothesis: The accuracy of the Δ -ML model of MTPs improves as more data are being added. Monopole coefficients have the fastest learning rate, quickly reaching MAE between $0.1 e$ and $0.01 e$ for all elements. Differences between elements are simply due to their relative frequency in the database. More specifically, since hydrogens and carbons predominate in small organic molecules, they provide much larger training sets, and consequently more accurate ML models when measured in terms of the percentage of training data being used. The scatter correlation plot between predicted and reference monopoles for each element is given in Figure 2b for the largest training fraction used: 85% (denoted " Δ -ML-85"), the exception being the hydrogen Δ -ML model trained on only 75% of the data set given its converged accuracy. The monopoles modeled by Δ -ML-85 reach high Pearson correlation coefficients: $R^2 \approx 97\%$, except F and Cl, for which $R^2 = 17\%$ and 68% , respectively. Such poor performance is explained by the small size training data set available for these elements. The outliers in the monopole scatter plot (Figure 2b) correspond to sulfur oxide groups. Also here, the few samples of these groups in the training set result in significant prediction errors.

Predicted dipoles show a MAE across elements between $0.02 e\text{\AA}$ and $0.15 e\text{\AA}$, depending on training set size. The heterogeneity of the chemical environments of the elements is reflected in the ML-model's performance. The Δ -ML dipole moments of hydrogens are extremely accurate—most likely due to hydrogens showing weak overall MTP moments and due to their repeating saturating bonding pattern. By contrast, carbon atoms, albeit being nearly as frequent as hydrogens in the database, have ML MTP models with significantly larger MAE. We rationalize the Δ -ML's relative difficulty to predict this element by the large chemical variety carbon exhibits, i.e., strongly varying hybridization states and possible bonding with all other elements. Also note the reversal of the relative offset of the F and Cl learning curves as one proceeds from monopole to

dipole moments, despite the fact that there are roughly twice as many Cl as F atoms in the database. This effect is possibly due to chlorine's larger polarizability, which implies that the chemical environment of the atom plays a more important role for the dipole moment, turning the ML-based modeling into a higher dimensional and thereby more challenging statistical learning problem. Such effects, however, can only fully be explained through an in-depth study with significantly larger data sets. The scatter plot for predicted versus reference dipole moments is shown for all elements in Figure 2c for Δ -ML-85. Clearly, the correlation is worse in comparison to monopoles ($R^2 \approx 50\%$), which is in line with what one would expect for a more complex vectorial quantity.

Of all MTPs considered, quadrupoles represent the most complex and challenging property. Not surprisingly, the resulting ML models yield the largest MAE for our training set: between $0.02 e\text{\AA}^2$ and $0.30 e\text{\AA}^2$ depending on training set size. The spread of MAEs across elements is strikingly more pronounced. Nevertheless, we find a larger correlation coefficient compared to dipoles: $R^2 \approx 65\%$; see Figure 2c,d. Note that the Cl/F accuracy reversal with respect to the monopole model is also manifested for the Δ -ML MTP model of this rank.

3.3. ML MTP vs Δ -ML MTP Model. We have compared the relative improvement gained when combining the ML with the baseline evaluation from the Voronoi scheme. Table 2 compares MAEs decomposed by chemical element for the prediction with 85% training fraction both with (i.e., " Δ -ML-85") and without (i.e., "ML-85") prior Voronoi baseline evaluation. The table also specifies the number of atoms involved in the prediction pool (i.e., outside the training fraction). While the Voronoi scheme does nothing to improve monopoles (for F and Cl it even worsens the prediction), it is increasingly helpful as we move to dipoles (with negligible change for F and Cl) and quadrupoles (with small improvement for F and Cl). We stress that the observed trends for F and Cl should be interpreted with utmost caution since their frequency in the database is very small (363 and 739). The lack of improvement for monopoles stems directly from the Voronoi scheme's strategy: Merely encoding symmetries, only higher MTP moments can benefit from the absence of a number of components that are forbidden by the underlying geometry. For fixed training size, the MAE is roughly halved for quadrupoles when using the Δ -learning procedure, compared to the standard ML methodology.

All results discussed so far refer to ML models of atomic MTPs in neutral molecules. For positively and negatively charged compounds we have found ML models to yield very similar trends and accuracy (see Figure S1 and S2 of the Supporting Information). The individual comparisons between ML or Δ -ML MTPs with QM MTPs are provided in the Supporting Information for all molecules that were not part of the training set.

4. VALIDATION

To assess the introduced MTP model's applicability, we have used predicted electrostatic coefficients to evaluate intermolecular interaction energies in molecular dimers and organic crystals. To do this, we accounted only for static MTP electrostatics and many-body dispersion (MBD) interactions,

$$E_{\text{inter}} \approx E_{\text{MTP}} + E_{\text{MBD}} \quad (12)$$

neglecting induction, penetration, and repulsion terms. A short description of the MBD formalism is provided in the next paragraph. As discussed earlier, our MTP and MBD models also represent approximations in the form of the Δ -ML model and dipole–dipole–many-body expansion, respectively. To better gauge the effect of the introduced ML MTP model, we also compare to intermolecular energy predictions using quantum-mechanically (QM) derived MTPs.

Common approximations in the exchange-correlation potential used in density functional theory lead to inadequate predictions of dispersion interactions. This has motivated the development of a number of dispersion-corrected methods. We hereby rely on the method developed by Tkatchenko and co-workers,⁴³ in which free-atom polarizabilities are first scaled according to their close environment following a partitioning of the electron density. The many-body dispersion up to infinite order (in the dipole approximation) is then obtained by diagonalizing the Hamiltonian of a system of coupled quantum harmonic oscillators, thereby coupling the scaled atomic polarizabilities at long range. The importance of many-body effects and accuracy of the method has been demonstrated on a large variety of systems.⁴³ Later, a classical approximation relaxed the requirement for an electron density, using instead a physics-based approach to estimate how atomic polarizabilities needed to be scaled based on a Voronoi partitioning.²⁴

4.1. Molecular Dimers. To gauge the accuracy of the electrostatics alone, we compare the electrostatic component of reference symmetry adapted perturbation theory (SAPT) results^{44,45} to the corresponding intermolecular components derived either from QM MTPs or from the Δ -ML MTP model. Figure 3 displays the correlation plot between the two model MTP electrostatics calculations and SAPT for the S22 dimers⁴⁶ at different intermolecular distances.^{44,45,47} The plot confirms that both MTP models generally underestimate the electrostatic SAPT component of the interaction energies, presumably due to a lack of penetration effects. A recent study, which observed a similar deviation, was able to correct it by introducing a model for charge penetration.⁴⁸ We expect their method to be applicable to the present MTPs.

Not all errors are distributed uniformly across compounds. Figure 4 compares the MTP energy contributions $E_{\text{MTP}}^{\text{QM}}$ and $E_{\text{MTP}}^{\text{ML}}$, as well as reference SAPT data for each molecular dimer of the S22 data set (i.e., at their equilibrium distance). We find good correlations between the QM and ML MTP models: correlation $R^2 = 78\%$ and MAE = 2.65 kcal/mol. Both QM and ML MTP models correlate virtually identically to the SAPT

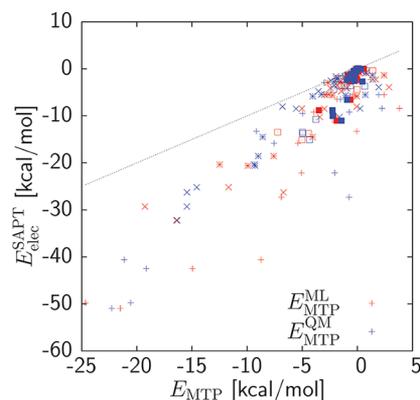


Figure 3. Scatter plot between reference energy components from SAPT and computed electrostatics from either quantum-mechanical coefficients, $E_{\text{MTP}}^{\text{QM}}$, or the universal Δ -ML-85 MTP model, $E_{\text{MTP}}^{\text{ML}}$, for intermolecular dimers in the S22 \times 5 data set.⁴⁷ Dimers closer than the equilibrium distance (i.e., factor 0.9) were systematically excluded. Symbols correspond to different distance factors that multiply the equilibrium distance: 0.9 (plus sign), 1.0 (cross), 1.2 (star), 1.5 (open square), and 2.0 (filled square). The straight line corresponds to ideal correlation.

results: correlation $R^2 = 87\%$ and 91% and MAE = 5.46 and 5.88 kcal/mol, respectively. We note a number of qualitative discrepancies between QM and ML MTP models: In particular, the ML model fails to reproduce the attractive nature of the electrostatic interaction for the water and ammonia dimers. Presumably, the model fails to predict their coefficients due to the molecules' unique chemical composition: the ML model relies on interpolations across the trained molecules, of which some must be chemically similar to the new compound. Larger molecules are less problematic because similarities between chemical fragments occur far more frequently. We do find systematic deviations between the MTP energies and the SAPT reference electrostatic data for strongly hydrogen-bonding compounds, for which penetration effects^{5,48} become significant.

We have calculated molecular dimer energies corresponding to various data sets for which high-level quantum-chemistry numbers have previously been published. We have considered the following databases: S22,⁴⁶ S22 \times 5,⁴⁷ S66 and S66 \times 8,⁴⁹ SCAI,⁵⁰ and X40 and X40 \times 10.⁵¹ All MTP coefficients have been predicted using the Δ -ML-85 model (see Figure 2 and Table 2). We only considered dimers made up of the chemical elements H, C, O, N, S, F, and Cl, keeping 992 out of over 1,300 intermolecular dimers.

Figure 5 contrasts the scatter correlation between reference intermolecular energies, E_{ref} , and the sum of many-body dispersion and ML-predicted MTP electrostatics, $E_{\text{MTP}}^{\text{ML}} + E_{\text{MBD}}$. The mean absolute error of all intermolecular estimates using the MTPs from individual quantum-chemistry calculations²⁴ and the ML predictions amount to 2.36 and 2.19 kcal/mol, respectively. In other words, the ML MTP prediction is on par with MTPs derived from explicit electron densities generated by computationally demanding quantum-chemistry calculations. Interactions between charged–charged amino acids of the SCAI database are reasonably well reproduced (see insets of Figure 5), pointing to the robustness of the method not only for neutral compounds but also charged species. The most challenging compounds in the X40 database are mostly composed of dimers that include a water or a

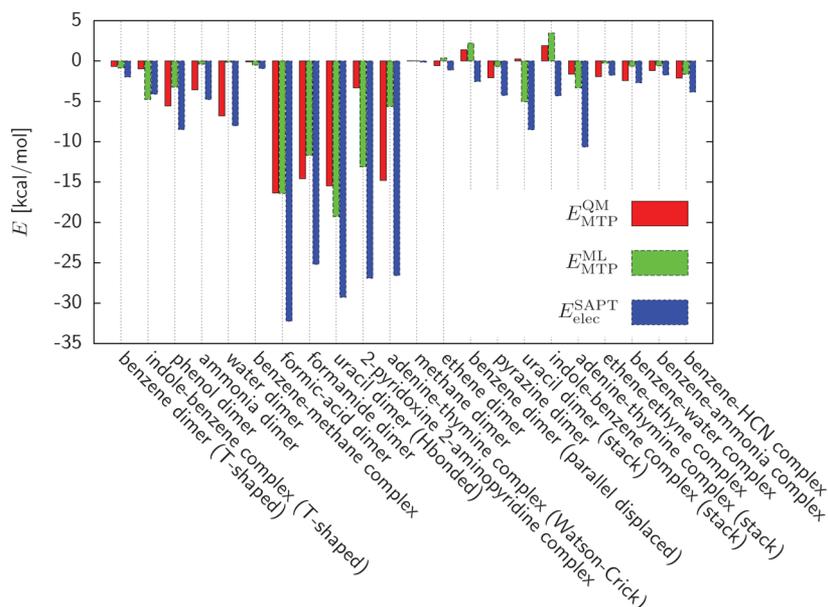


Figure 4. Intermolecular static electrostatic energy contribution for each dimer of the S22 data set, for coefficients parametrized from quantum-mechanical calculations, E_{MTP}^{QM} , and the transferable ML model Δ -ML-85, E_{MTP}^{ML} . Electrostatic energies from reference SAPT calculations are also provided.⁴⁵

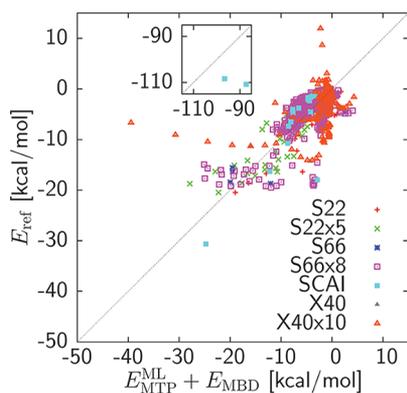


Figure 5. Correlation plot between reference intermolecular energies, E_{ref} , and energies predicted from the classical many-body dispersion method with static electrostatics, the latter parametrized from the Δ -ML-85 model, $E_{MTP}^{ML} + E_{MBD}$. The inset corresponds to the two charged–charged side-chain interactions of the SCAI database. The present lack of repulsion, induction, and penetration effects is a cause for the outliers.

diatomic molecule (e.g., HF and HCl). As argued earlier, the smallest molecules are challenging due to the absence of such small compounds in the database, which affects the quality of the interpolation. For a more detailed account of the performance of ML MTPs, we provide decompositions of the electrostatic and dispersion energies for ML and QM MTPs, as compared to reference energies (see Figures S3–S26 in the Supporting Information).

4.2. Benzene Crystal. Increasingly accurate and fast methods provide the means for crystal structure prediction of organic compounds,⁵³ to the point of ranking polymorphs of molecular crystals.^{54,55} Moving toward a condensed-phase system, we have also evaluated the cohesive binding energy predictions of a molecular benzene crystal. Following previous work,^{56,57} we have computed the binding energy for different ratios of the unit-cell density with respect to the experimental density,⁵² ρ/ρ_{exp} . Isotropic density scalings were performed

without further optimization, and the binding energy included interactions with the neighboring unit cells, as discussed in our previous publication.²⁴ Figure 6 features the MTP- and MBD-

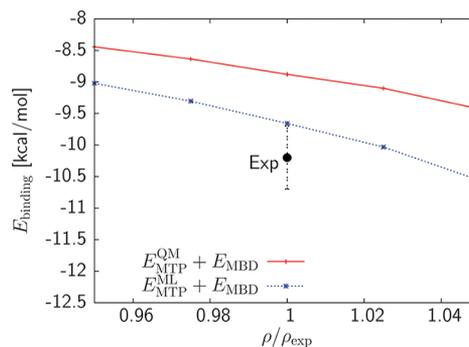


Figure 6. Cohesive binding energy of the benzene crystal as a function of the ratio of densities, ρ/ρ_{exp} . Both models for intermolecular energies are presented: $E_{MTP}^{QM} + E_{MBD}$ and $E_{MTP}^{ML} + E_{MBD}$. The experimental value is shown explicitly (black dot), including the error bar.⁵²

based intermolecular estimates of cohesive energy as a function of density. Again, we compare the resulting numbers once using the Δ -ML-predicted MTP electrostatics, combined with MBD, and once using the QM MTP model combined with MBD. We find very good agreement between the two methods, validating here as well the ML model. Coincidentally, the two methods yield cohesive binding energies in good agreement with the experimental value. The lack of repulsive interactions prohibits a further increase in energy as ρ gets larger.²⁴

5. CONCLUSION

We have introduced machine learning models for electrostatic multipoles (MTPs) of H, C, O, N, S, F, and Cl atom types. The models have been trained on atomic multipole coefficients of small organic molecules evaluated using first-principles calculations. Neutral, cationic, and anionic molecular states

were treated with separate models. The model yields highly accurate MTPs for H, reasonable performance for C, N, and O, and significant errors for S, F, and Cl due to their sparsity in the training set.

Focusing on the intermolecular S22 dimer data set, MTP energies show good correlation between the coefficients parametrized from ML and individual quantum-chemistry calculations. A comparison with reference electrostatic interactions from symmetry-adapted perturbation theory (SAPT) is satisfactory for large intermolecular separations and impaired by the lack of penetration effects at short distances.

Furthermore, MTPs from the ML model have been combined with a classical many-body dispersion potential to estimate intermolecular energies of nearly 1,000 molecular dimers as well as the cohesive energy of the benzene crystal. The results show that the ML model retains overall a similar accuracy compared to calculations with the MTPs parametrized from individual quantum-chemical calculations.

The Δ -ML approach, which augments a physics-based baseline model by a ML model, has proven to be useful to more efficiently train vector and tensor quantities. Incorporation of molecular symmetries via the Voronoi partitioning of the charge density, included in the baseline model, is at the heart of this improvement.

The proposed models alleviate the need for quantum-chemistry calculations for every single molecule/molecular conformation in a perturbative evaluation of intermolecular interactions, bringing us one step forward toward the task of automated parametrizations of accurate state-independent and transferable force fields.

■ ASSOCIATED CONTENT

● Supporting Information

Atomic coordinates of all chemical compounds included in the training and test sets, comparison between ML and QM multipoles for each sample in the test set, and electrostatic and dispersion components of the intermolecular energies as well as reference energy for all molecular dimers. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00301.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: bereau@mmpip-mainz.mpg.de.

*E-mail: anatole.vonlilienfeld@unibas.ch.

Funding

O.A.v.L. acknowledges funding from the Swiss National Science Foundation (Grant No. PP00P2_138932). This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy (DOE) under Contract DE-AC02-06CH11357.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Andreas Hesselmann and John M. Herbert for kindly providing the SAPT reference energies, Matthias Rupp for useful discussions, and Aoife Fogarty and Markus Meuwly for critical reading of the manuscript.

■ REFERENCES

- (1) Atkins, P. W.; de Paula, J. *Physical Chemistry*; Wiley-VCH: Weinheim, Germany, 2013.
- (2) Israelachvili, J. N. *Intermolecular and Surface Forces*; Academic Press: San Diego, CA, USA, 1985.
- (3) Wedler, G. *Lehrbuch der Physikalischen Chemie*; Wiley-VCH: Weinheim, Germany, 1997.
- (4) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (5) Stone, A. *The theory of intermolecular forces*; Oxford University Press: Oxford, United Kingdom, 2013.
- (6) Kramer, C.; Spinn, A.; Liedl, K. R. *J. Chem. Theory Comput.* **2014**, *10*, 4488–4496.
- (7) Piquemal, J.-P.; Chevreau, H.; Gresh, N. *J. Chem. Theory Comput.* **2007**, *3*, 824–837.
- (8) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *J. Chem. Theory Comput.* **2007**, *3*, 1960–1986.
- (9) Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. V. *Annu. Rev. Phys. Chem.* **2013**, *64*, 553–578.
- (10) 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., Jr.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- (11) Bereau, T.; Kramer, C.; Meuwly, M. *J. Chem. Theory Comput.* **2013**, *9*, 5450–5459.
- (12) Lee, M. W.; Carr, J. K.; Göllner, M.; Hamm, P.; Meuwly, M. *J. Chem. Phys.* **2013**, *139*, No. 054506.
- (13) Cazade, P.-A.; Bereau, T.; Meuwly, M. *J. Phys. Chem. B* **2014**, *118*, 8135–8147.
- (14) Jakobsen, S.; Bereau, T.; Meuwly, M. *J. Phys. Chem. B* **2015**, *119*, 3034–3045.
- (15) Cortes, C.; Vapnik, V. *Mach. Learn.* **1995**, *20*, 273–297.
- (16) Müller, K.-R.; Mika, S.; Rätsch, G.; Tsuda, K.; Schölkopf, B. *IEEE Trans. Neural Networks* **2001**, *12*, 181–201.
- (17) Schölkopf, B.; Smola, A. J. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*; MIT Press: Cambridge, MA, USA, 2002.
- (18) Witten, I. H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2005.
- (19) Sumpter, B. G.; Noid, D. W. *Chem. Phys. Lett.* **1992**, *192*, 455–462.
- (20) Rai, B. K.; Bakken, G. A. *J. Comput. Chem.* **2013**, *34*, 1661–1671.
- (21) Ivanov, M. V.; Talipov, M. R.; Timerghazin, Q. K. *J. Phys. Chem. A* **2015**, *119*, 1422–1434.
- (22) Handley, C. M.; Hawe, G. I.; Kell, D. B.; Popelier, P. L. *Phys. Chem. Chem. Phys.* **2009**, *11*, 6365–6376.
- (23) Mills, M. J.; Popelier, P. L. *Theor. Chem. Acc.* **2012**, *131*, 1–16.
- (24) Bereau, T.; von Lilienfeld, O. A. *J. Chem. Phys.* **2014**, *141*, No. 034101.
- (25) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*; Springer Series in Statistics; Springer: New York, NY, USA, 2001.
- (26) Schölkopf, B.; Smola, A.; Müller, K.-R. *Neural Comput.* **1998**, *10*, 1299–1319.
- (27) Ramakrishnan, R.; Dral, P.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ -Machine Learning Approach, <http://arxiv.org/abs/1503.04987>, 2015.
- (28) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. *Phys. Rev. Lett.* **2012**, *108*, No. 058301.
- (29) von Lilienfeld, O. A.; Ramakrishnan, R.; Rupp, M.; Knoll, A. *Int. J. Quantum Chem.*, **2015**, DOI: 10.1002/qua.24912.
- (30) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419.
- (31) Ramakrishnan, R.; von Lilienfeld, O. A. *Chimia* **2015**, *69*, 182–186.
- (32) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1849–1868.

- (33) Dunning, T. H., Jr *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian: Wallingford, CT, USA, 2009.
- (35) Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caffisch, A.; Caves, L.; Cui, Q.; Dinner, A.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R.; Post, C.; Pu, J.; Schaefer, M.; Tidor, B.; Venable, R.; Woodcock, H.; Wu, X.; Yang, W.; York, D.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (36) Bereau, T.; Kramer, C.; Monnard, F. W.; Nogueira, E. S.; Ward, T. R.; Meuwly, M. *J. Phys. Chem. B* **2013**, *117*, 5460–5471.
- (37) <https://github.com/tbureau/mbvdw>.
- (38) Chu, X.; Dalgarno, A. *J. Chem. Phys.* **2004**, *121*, No. 4083.
- (39) von Lilienfeld, O. A.; Tkatchenko, A. *J. Chem. Phys.* **2010**, *132*, No. 234109.
- (40) Okabe, A.; Boots, B.; Sugihara, K.; Chiu, S. N. *Spatial tessellations: Concepts and applications of Voronoi diagrams*, Vol. 501; John Wiley & Sons: Hoboken, NJ, USA, 2009.
- (41) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics Modell.* **1996**, *14*, 33–38.
- (42) Grotthuss, M. v.; Koczyk, G.; Pas, J.; Wyrwicz, L. S.; Rychlewski, L. *Comb. Chem. High Throughput Screening* **2004**, *7*, 757–761.
- (43) Tkatchenko, A.; DiStasio, R. A., Jr.; Car, R.; Scheffler, M. *Phys. Rev. Lett.* **2012**, *108*, No. 236402.
- (44) Hesselmann, A. *J. Phys. Chem. A* **2011**, *115*, 11321–11330.
- (45) Lao, K. U.; Herbert, J. M. *J. Chem. Phys.* **2014**, *140*, No. 044108.
- (46) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
- (47) Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 2365–2376.
- (48) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardère, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. *J. Chem. Theory Comput.* **2015**, DOI: 10.1021/acs.jctc.5b00267.
- (49) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.
- (50) Berka, K.; Laskowski, R.; Riley, K. E.; Hobza, P.; Vondrášek, J. *J. Chem. Theory Comput.* **2009**, *5*, 982–992.
- (51) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.
- (52) Schweizer, W. B.; Dunitz, J. D. *J. Chem. Theory Comput.* **2006**, *2*, 288–291.
- (53) Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S. X.; Braun, D. E.; Cruz-Cabeza, A. J.; Day, G. M.; della Valle, R. G.; Desiraju, G. R.; van Eijck, B. P.; Facelli, J. C.; Ferraro, M. B.; Grillo, D.; Habgood, M.; Hofmann, D. W. M.; Hofmann, F.; Jose, K. V. J.; Karamertzanis, P. G.; Kazantsev, A. V.; Kendrick, J.; Kuleshova, L. N.; Leusen, F. J. J.; Maleev, A. V.; Misquitta, A. J.; Mohamed, S.; Needs, R. J.; Neumann, M. A.; Nikylov, D.; Orendt, A. M.; Pal, R.; Pantelides, C. C.; Pickard, C. J.; Price, L. S.; Price, S. L.; Scheraga, H. A.; van de Streek, J.; Thakur, T. S.; Tiwari, S.; Venuti, E.; Zhitkov, I. K. *Acta Crystallogr., Sect. B: Struct. Sci.* **2011**, *67*, 535–551.
- (54) Marom, N.; DiStasio, R. A.; Atalla, V.; Levchenko, S.; Reilly, A. M.; Chelikowsky, J. R.; Leiserowitz, L.; Tkatchenko, A. *Angew. Chem., Int. Ed.* **2013**, *52*, 6629–6632.
- (55) Reilly, A. M.; Tkatchenko, A. *Phys. Rev. Lett.* **2014**, *113*, No. 055701.
- (56) Meijer, E. J.; Sprik, M. *J. Chem. Phys.* **1996**, *105*, 8684–8689.
- (57) Tapavicza, E.; Lin, I.-C.; von Lilienfeld, O. A.; Tavernelli, I.; Coutinho-Neto, M. D.; Rothlisberger, U. *J. Chem. Theory Comput.* **2007**, *3*, 1673–1679.