

Development and Implementation of Atomically-Anisotropic First-Principles Force Fields: A Benzene Case Study

Tesia D. Janicki^{¶,†} Mary J. Van Vleet^{¶,‡} and J. R. Schmidt^{*,†}

[†]*Department of Chemistry, University of Wisconsin–Madison, 1101 University Avenue
Madison, WI 53706, United States*

[‡]*Department of Chemistry and Biochemistry, Spelman College, 350 Spelman Ln SW,
Atlanta, GA 30314, United States*

[¶]*These authors contributed equally to this Work*

E-mail: schmidt@chem.wisc.edu

Abstract

π -interactions are an important motif in chemical and biochemical systems. However, due to their anisotropic electron densities and complex balance of intermolecular interactions, aromatic molecules represent an ongoing challenge for accurate and transferable force field development. Historically, *ab initio* force fields for aromatics have not exhibited good accuracy with respect to bulk properties, or have only been used to study gas-phase dimers. Using benzene as a proof of concept, herein we show how our own *ab initio*, MASTIFF force field incorporates an atomically anisotropic description of intermolecular interactions to yield an accurate and robust model for aromatic interactions irrespective of phase. Compared to existing models, the MASTIFF benzene force field is not only accurate for liquid phase properties, but also offers transferability to the gas-

and solid-phases. Additionally, we introduce a computationally efficient OpenMM plugin that enables customizable anisotropic intermolecular functional forms, and which can be generically used in any MD simulation where a model for non-spherical atomic features is required. Overall, our results demonstrate the importance of atomic-level anisotropy in enabling next-generation *ab initio* force field development.

1 Introduction

π -contact interactions are ubiquitous in chemistry and play an instrumental role in applications ranging from protein folding and molecular recognition, to crystal engineering and drug design.^{1–11} Understanding the physics behind $\pi-\pi$ stacking and other related π -contact interactions is an active area of research for both theory^{6,12,13} and experiment.^{10,14–17} Force field models have significantly deepened our understanding of the structure and energetics of aromatic systems, and have overall been very successful in modeling environments such as neat liquids.^{8,18–21} However, these same force fields are not without shortcomings,^{22–24} especially when it comes to their transferability to more challenging π -contact environments such as with cation- π interactions^{25,26} or gas-phase clusters.^{27,28} The root cause(s) of these accuracy and transferability challenges are complex, but historically have been attributed to:

- a) difficulties obtaining accurate quantum mechanical (QM)^{5,29–32} data and certain types of molecular-resolution experimental benchmark data,^{32,33}
- b) the combination of several distinct components of the intermolecular interaction that all substantially contribute to structural and energetic trends in π -interactions,^{2,3,25,32,34}
- c) disparities between the minimum energy geometries of aromatic intermolecular contacts in the gas phase vs. in the bulk,^{1,2,35} and
- d) the neglect of essential physics in force fields, such as anisotropy, polarizability, and

charge penetration.^{3,22,28}

Regarding the first challenge, high accuracy QM calculations for small aromatic molecules have become computationally affordable within the last two decades,^{29,30,36,37} and have led to several notable efforts to develop force fields for aromatics *ab initio* (that is, from QM calculations alone).^{19,25,28,38,39} *Ab initio* development strategies are appealing for many reasons,⁴⁰ including the lack of dependence on experimental data, the ability to naturally separate the force field into meaningful physical components,^{28,41–43} and the possibility of using more complex functional forms and flexible parameters in order to better capture the essential physics of aromatic intermolecular interactions.^{38,39,44–47} However, while many of the existing *ab initio* force fields for aromatics have excellent accuracy in the gas-phase,^{19,28,38,39} the condensed phase is an active challenge. Force fields with simple functional forms and combination rules (which we refer to in this work as “general” force fields) typically have poor accuracy for bulk property predictions,^{19,25,28} while those with more advanced functional forms or explicit combination rules are challenging to implement within standard simulation software, and have thus not yet been applied to study the condensed phase.^{38,39}

The neglect of essential physics is also an important limitation of aromatic force fields. Standard force fields all utilize the popular isotropic (or "sum of spheres") approximation, in which atoms of a given pair interaction are modeled purely in terms of interatomic distances and any anisotropic (i.e. orientation-specific) effects are neglected.^{46,48,49} This sum-of-spheres approximation has been linked to accuracy issues in systems with highly non-spherical electron densities, such as with π -contact interactions, and we hypothesize that this explains many of the observed accuracy challenges with aromatic force fields.^{32,46} Select general force fields, such as AMOEBA, have explicitly incorporated anisotropy into the electrostatic term.^{50,51} Other models, such as OPLS-CS¹⁹ and the rigid-monomer “POT3” potential developed by Szalewicz and coworkers,^{38,52} have utilized off-atom sites to implicitly mimic the effects of anisotropy without introducing explicit orientation-dependent functional forms. However, few aromatic force fields (and, to our knowledge, no MD-compatible gen-

eral force fields) have explicitly included anisotropy in the non-electrostatic terms of the force field, despite the fact that both exchange and dispersion anisotropy are known to be important in π systems.^{39,46,47,53} We therefore see a need to incorporate atomic anisotropy explicitly into *each* term of the force field so as to simply but fully account for the physics of intermolecular interactions in aromatic molecules.

Taken together, the overarching goal of the present work is to alleviate two major hurdles associated with force field development for aromatics: first, the need for more sophisticated functional forms that better capture the essential physics of aromatic and π -type interactions, and second, the lack of compatibility between these sophisticated functional forms and standard simulation software packages. We have chosen to study benzene as a “proof of concept” prototype for aromatic interactions,^{4,7–10,54,55} and leave discussions of more complex molecules and aromatic interaction motifs for future work. To meet our first goal, we develop a benzene force field based on the MASTIFF (**M**ultipolar, **A**nisotropic, **S**later-**T**ype **I**ntermolecular **F**orce **F**ield) strategy introduced in prior work.^{45,46} MASTIFF incorporates both atomic-level anisotropy and polarization effects explicitly, and has shown notable improvement over isotropic non-polarizable models in reproducing DFT-SAPT energies for a diverse set of molecular geometries. To reach our second aim, we introduce a computationally efficient plugin to the popular OpenMM simulation software^{56,57} that enables customizable anisotropic intermolecular force expressions. Notably, this plugin is not specific to MASTIFF or aromatics, and is thus intended to be generally useful for any molecule and any phase state in which an orientation-dependent force field is needed. Our results compare our newly developed MASTIFF benzene model to existing empirical and *ab initio* benzene force fields for both dimer and bulk properties, and show how our force field and associated OpenMM plugin can be used as a general design strategy for accurate and transferable *ab initio* modeling of aromatic molecules.

2 Background

Table 1: Functional forms and parameterization methods for several popular benzene force fields. ^aBoth QM calculations and experimental bulk property data used in the training set.

Force Field	Refs	Year	Parameters	Electrostatics	Polarization	VdW Form
AMBER	⁵⁸	2003	Empirical	(PC)	Implicit	Lennard-Jones
AMOEBA	^{50,51,59}	2017	Mixed ^a	Multipoles	Explicit	Buffered 14–7
CHARMM27	^{60,61}	2004	Empirical	PC	Implicit	Lennard-Jones
OPLS-AA	^{62,63}	1990	Empirical	PC	Implicit	Lennard-Jones
OPLS-CS	¹⁹	2006	Mixed	Off-Site PC	Implicit	Lennard-Jones
OPT-FF	²⁸	2009	<i>Ab initio</i>	PC	Implicit	Lennard-Jones
QMD-FF	⁶⁴	2015	<i>Ab initio</i>	PC	Implicit	Lennard-Jones
POT3	^{38,52}	2010	<i>Ab initio</i>	Off-Site PC	Implicit	Generalized Exp6
MASTIFF	this work		<i>Ab initio</i>	Multipoles	Explicit	Slater-Type

A large number of molecular models (force fields)^{19,28,50,58,60–63,65,66} have been parameterized for benzene to study atomic scale interactions, each with varying success at reproducing gas and bulk-phase properties.¹⁸ A representative, non-exhaustive list of benzene force fields is given in Table 1 to highlight the diversity of parameterization methods and functional forms used. Of these force fields, AMBER,⁵⁸ CHARMM27,^{60,61} and OPLS-AA^{62,63} are popular general-purpose empirical force fields fit to reproduce bulk liquid properties such as the density and enthalpy of vaporization. Each of these “standard force field” models employs an isotropic, non-polarizable functional form (denoted LJ+ q) that treats van der Waals (vdW) forces via the common 12–6 Lennard-Jones potential and electrostatics via point charges (q). These simplistic empirical models are very accurate in predicting homogeneous bulk properties, particularly in the case of OPLS-AA.¹⁸ Such accuracy is to be expected given that these properties were included in their training sets. However, these force fields have also been shown to have limited accuracy with respect to mixtures and gas-phase clusters.^{18,24,28}

In order to improve accuracy and transferability of general force fields, several attempts have been made to incorporate advanced functional forms and/or *ab initio* data into parameterization. Early efforts in this area led to the OPLS-CS¹⁹ model (a mixed empirical/*ab-initio* model parameterized to both bulk properties and gas phase quantum

data, and which includes off-site point charges in an effort to better reproduce benzene’s multipole moments) and OPT-FF²⁸ (a purely *ab-initio* model that uses the standard LJ+q functional form and is parameterized to gas phase dimer energies). Unfortunately, Fu and Tian have shown that both approaches struggle to reproduce bulk properties.¹⁸ Later work on the QMD-FF force field,⁶⁴ which uses the same functional form as OPT-FF but a more extensive QM training set of dimers taken from MD simulations, displays much improved bulk property predictions, suggesting that larger training sets and judicious choice of QM reference data can overcome many of the hurdles associated with *ab initio* force field development. Lastly, the AMOEBA^{50,51,59,67} force field employs advanced functional forms and has been empirically parameterized for aromatic complexes including benzene. AMOEBA explicitly incorporates atomic anisotropy by replacing the traditional point charge model with electrostatic multipoles; however, the remaining VdW terms are isotropic. As demonstrated below, AMOEBA performs well for select liquid phase properties, but (arguably due to limitations in its VdW functional form,^{68,69} including the lack of non-electrostatic anisotropy^{53,69,70}) is of more limited accuracy for transport properties, gas-phase dimers, and the solid phase.

As an alternative to general force fields, multiple research groups have also modeled benzene using *ab initio* potential energy surfaces (PES).^{39,52} Similar to the present work, PES approaches are entirely *ab initio* and utilize high-quality QM benchmarks such as SAPT and/or CCSD(T) for parameterization.⁷¹ In other aspects, PES differ from general force fields (including the present work): by eliminating combination rules in favor of explicit cross-terms and their associated additional free parameters, these PES can achieve very high accuracy for a specific molecule, albeit with the disadvantages that a) the model cannot be used to study mixtures without substantial additional parameterization and b) explicit cross-terms are typically incompatible with standard MD software. For these reasons, PES have historically been restricted to studies of homogeneous dimer properties, although there has been recent effort to automate PES parameterization^{72,73} and simulate both neat liquids and binary mixtures.⁷⁴⁻⁷⁷ Insofar as benzene is concerned, the rigid-monomer “POT3” PES

developed by Podesszwa and co-workers^{38,52} is particularly noteworthy, and (as discussed in the Results section) predicts benzene dimer energies and geometries with nearly SAPT-level accuracy. The functional form for POT3 is based on point charges and a generalized Buckingham potential. While technically a potential with no angular dependence, POT3 uses 13 off-atom interaction sites (1 site at the center of mass, 1 site between each chemical bond) to model the non-sphericity of benzene and can thus be considered an “implicitly anisotropic” model.

Finally, the MASTIFF (**M**ultipolar, **A**nisotropic, **S**later-**T**ype **I**ntermolecular **F**orce **F**ield) approach, which is the focus of this work, attempts to overcome the limitations of the isotropic approximation via a physics-based approach that explicitly incorporates atomic-level anisotropy, not just into the electrostatics component, but into each term of the force field. MASTIFF is a general force field that has shown notable improvement over isotropic models in reproducing QM energies for a diverse set of molecular geometries,⁴⁶ and will be extended in the current work to model benzene as a prototype for aromatic interactions. Because the MASTIFF functional form and fitting methodology have been fully described in prior work, here we provide a background summary of the main equations but refer the interested reader to Ref. 46 for complete details.

MASTIFF describes the intermolecular interaction energy as a sum of physically-meaningful interaction components — exchange, electrostatics, induction, and dispersion — each of which is fit to a corresponding QM energy benchmark:

$$V_{\text{FF}} = \sum_{ij} V_{ij}^{\text{exch}} + V_{ij}^{\text{elst}} + V_{ij}^{\text{ind}(2)} + V_{ij}^{\delta\text{HF}} + V_{ij}^{\text{disp}} \quad (1)$$

Because DFT-SAPT subdivides induction into 2nd and higher order (“ δHF ”) components, the MASTIFF induction energy is also a sum of two components, $V^{\text{ind}(2)}$ and $V^{\delta\text{HF}}$.

The functional form for each component is defined by the following equations:

$$\begin{aligned}
V_{ij}^{\text{exch}} &= A_{ij}^{\text{exch}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) \\
V_{ij}^{\text{elst}} &= -A_{ij}^{\text{elst}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) + \sum_{tu} Q_t^i T_{tu} Q_u^j \\
V_{ij}^{\text{ind}(2)} &= -A_{ij}^{\text{ind}(2)} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) + V_{\text{pol}}^{(2)} \\
V_{ij}^{\delta\text{HF}} &= -A_{ij}^{\delta\text{HF}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) + V_{\text{pol}}^{(3-\infty)} \\
V_{ij}^{\text{disp}} &= -A_{ij}^{\text{disp}} \sum_{n=3}^6 f_{2n}(x) \frac{C_{ij,2n}}{r_{ij}^{2n}} \\
B_{ij} &= \sqrt{B_i B_j} \\
C_{ij,2n} &= \sqrt{C_{i,2n} C_{j,2n}} \\
P(B_{ij}, r_{ij}) &= \frac{1}{3} (B_{ij} r_{ij})^2 + B_{ij} r_{ij} + 1 \\
f_{2n}(x_{ij}) &= 1 - e^{-x} \sum_{k=0}^{2n} \frac{(x)^k}{k!} \\
x_{ij} &= B_{ij} r_{ij} - \frac{2B_{ij}^2 r_{ij} + 3B_{ij}}{B_{ij}^2 r_{ij}^2 + 3B_{ij} r_{ij} + 3} r_{ij}
\end{aligned} \tag{2}$$

The exchange potential, V^{exch} , describes short-range Pauli exclusion effects via a modified, ‘‘Slater-ISA’’⁴⁵ decaying exponential. This particular functional form is derived from the overlap^{78,79} of exponentially-decaying atom-in-molecule electron densities. Unlike the $1/r^{12}$ repulsion functional form used in standard force fields, the Slater-ISA form is physics-based and quantitatively matches the shape of the repulsive wall, even at small r and large repulsion energies (see Figure 3 and Ref. 45). The electrostatic potential, V^{elst} , models both short-range charge penetration and long-range Coloumbic interactions. Short-range effects are described by the same Slater-ISA functional form as is used for exchange, albeit with different parameters. The functional form for long-range electrostatics is identical to that used by AMOEBA: a distributed multipole expansion with multipole moments, Q , fit up to quadrupoles.^{51,80} Induction effects are modeled via Slater-ISA terms, which describe short-

range effects, and an Thole-damped induced dipole model, which describes the long-range and is identical to the AMOEBA functional form. Lastly, dispersion is described by a Tang-Toennies damped^{81,82} dispersion multipole expansion.

In order to account for anisotropic (i.e. angular-dependent) effects, the MASTIFF functional form treats each short-range prefactor, A , as an expansion in spherical harmonics, Y_{lm} :

$$\begin{aligned} A_{ij} &= A_i A_j \\ A_i &\equiv A_i(\theta_i, \varphi_i) = \textcolor{red}{A}_{i,\text{iso}}(1 + \xi_i), \\ \xi_i &= \sum_{l>0,m} \textcolor{red}{a}_{i,lm} \sqrt{\frac{4\pi}{2l+1}} Y_{lm}(\theta_i, \varphi_i) . \end{aligned} \tag{3}$$

This choice of combination rules and anisotropic functional form is derived from the first-principles overlap model for anisotropic repulsion.^{46,79,83} The orientation dependent variables θ_i and φ_i are defined via local coordinate systems assigned to each anisotropic atom. By aligning the local coordinate system to an atom’s local symmetry,⁴⁶ many $a_{i,lm}$ terms are reduced to zero. Thus, in practice, only a small number of parameterized $a_{i,lm}$ terms are required to converge to sub-kJ/mol accuracy for dimer interaction energies.

In order to train a MASTIFF model, parameters describing long-range effects are calculated from distributed atom-in-molecule monomer properties, whereas parameters for short-range effects must be fit directly to benchmark DFT-SAPT dimer calculations. For visual clarity, one instance of each parameter for the model (Equations (1) to (3)) has been shown in color, with calculated monomer parameters in blue and fitted dimer parameters in red. For a select few parameters, modest accuracy improvements have been found by first calculating the parameters from monomer calculations, and then using the calculated parameters as soft constraints when fitting to the DFT-SAPT dimer energies; these parameters are shown in yellow.

3 Software

OpenMM⁵⁷ is a popular GPU-accelerated molecular dynamics engine with customizable force types. OpenMM has been used with MASTIFF to simulate the condensed phase in prior work;⁴⁶ however, the previous software implementation was both computationally inefficient and limited to very small molecules (see SI for details). Here we have implemented a more general and more efficient open-source plugin that enables custom anisotropic functional forms in OpenMM. This CustomAnisotropicNonbonded (CAN) force type and associated plugin are available on GitHub,⁸⁴ and full implementation details and performance benchmarks are given in the SI. In the CAN plugin, anisotropy is incorporated by explicitly defining a local coordinate system at each atom site with respect to its surrounding molecular geometry. These local axes definitions can then be used to compute the polar and azimuthal coordinates (φ_i and θ_i , respectively, following the mathematics convention) needed for anisotropic functional forms. Similar to the typing syntax used in OpenMM’s Amoeba-MultipoleForce, the relevant atoms assigning each atomtype’s local geometry are added as per-particle parameters to a standard xml force field file. Five types of local geometries are currently supported by CAN: Z-then-X, Bisector, Z-Bisect, Threefold, and Z-Only (see Figure S3). Figure 1 shows an example of CAN’s atom-typing and local geometry syntax for the carbon and hydrogen atomtypes in benzene. It is important to stress that the CAN plugin is not limited to benzene or MASTIFF, but rather is intended to be broadly applicable for any molecule and any custom anisotropic force expression.

4 Methods

4.1 MASTIFF Force Field Parameterization

Parameterization of our new flexible-monomer MASTIFF benzene potential follows the same standard procedure as in prior work.⁴⁶ Permanent multipole moments (Q_i) and short-range

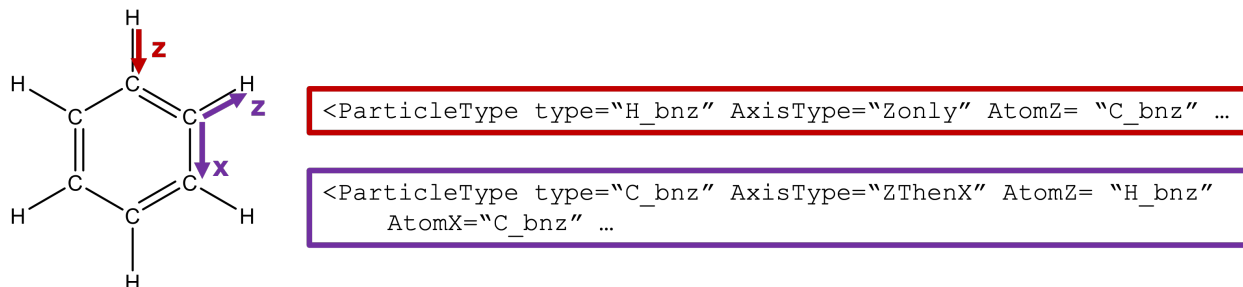


Figure 1: Local axes definitions for the hydrogen (red) and carbon (purple) atomtypes using the CAN plugin. For each atomtype, the local z- and (where required) x-axes vectors are defined with respect to neighboring atoms.

exponents (B_i) were calculated from atom-in-molecule monomer properties using the BS-ISA algorithm.^{45,80} Calculated B coefficients were treated as soft-constraints and allowed to relax slightly in fits to benchmark QM dimer energies (described in the next subsection). For this work, atomic polarizabilities were taken from Dang²⁶ and used without modification. In keeping with standard practice for the AMOEBA force field, we used a Thole-damping factor of 0.39. Dispersion parameters were taken from the sp^2 carbon and hydrogen (Csp2 and H_Csp2) atomtypes in Ref. 43 and used without modification. Intramolecular parameters for bonds, angles, and dihedrals were used without modification from Ref. 64.

Short-range parameters for MASTIFF ($A_{i,iso}$ and $a_{i,lm}$) were optimized to reproduce the benchmark QM energies according to the default procedure in Ref. 46. To best minimize root-mean-squared errors (RMSE, see Figure 2), we chose to treat both carbon and hydrogen atoms anisotropically. As in prior work,⁴⁶ only symmetry-allowed terms up to 2nd order ($l = 2$) were included in the expansion; introducing 3rd order terms led to only minor improvements in RMSE and thus were excluded. For hydrogen (treated as having quasi- $C_{\infty V}$ symmetry), only the Y_{10} and Y_{20} spherical harmonics are symmetry-allowed, and thus we only needed to optimize parameters for $a_{H,10}$ and $a_{H,20}$. Analogously for carbon (C_{2V} symmetry), only three anisotropic terms ($a_{C,10}$, $a_{C,20}$, $a_{C,22c}$) required fitting.

For dispersion-dominated complexes such as benzene, 3-body dispersion is known to be important for quantitative bulk property predictions.^{31,85,86} Thus a damped 3-body Axilrod-

Teller-Muto (ATM) term,^{46,85}

$$V_{ijk}^{\text{disp}} = C_{ijk,9} f_3(x_{ij}) f_3(x_{jk}) f_3(x_{ik}) \frac{1 + 3 \cos \phi_i \cos \phi_j \cos \phi_k}{r_{ij} r_{jk} r_{ik}}, \quad (4)$$

was added to the pair potential described above, where i, j, k refer to atomic indices. The f_3 damping functions are defined in Equation (2). In keeping with literature precedent,^{85,87} the C_9 coefficients were approximated using dispersion coefficients (C_6) and static dipole polarizabilities ($\alpha_1(0)$) taken from the pair potential:

$$C_{ijk,9} \approx \frac{2S^i S^j S^k (S^i + S^j + S^k)}{(S^i + S^j)(S^j + S^k)(S^k + S^i)} \quad (5)$$

$$S^a = C_{ii,6} \frac{\alpha_1^j(0) a_1^k(0)}{\alpha_1^i(0)} \quad (6)$$

Consequently, the 3-body potential ATM potential required no additional free parameters or fitting.

4.2 ISO-MASTIFF

In order to explore the extent to which the anisotropic terms (both $a_{i,lm}$ parameters and higher-order multipoles) improve fit quality, we also separately optimized a completely isotropic, “ISO-MASTIFF” version of the force field. ISO-MASTIFF is largely identical to MASTIFF, with the exception that all $a_{i,lm}$ and higher-order multipole terms are set to zero. Put differently, electrostatics for ISO-MASTIFF are computed at the point charge level, the short-range terms have no orientation dependence, and the final optimized A_{ij} and B_{ij} parameters for ISO-MASTIFF and MASTIFF differ numerically.

4.3 Benchmark Energies

Short-range parameters for MASTIFF require fitting to energy-decomposed benchmark QM dimer interaction energies. Similar to prior work,⁴⁵ we used the Molpro2009 software suite⁸⁸

and a DFT-SAPT/aVTZ+m level of theory to compute the interaction energy and associated energy decomposition for 1000 quasi-randomly generated benzene dimers. Here, +m refers to a set of 5s3p1d1f even-tempered midbond functions placed at the midpoint between the centers of mass of the two interacting benzene monomers. We also computed counterpoise-corrected dimer interaction energies using the PSI4 software at a DF-FNO-CCSD(T)/CBS(a[TQ]Z, δ aDZ) level of theory. With the exception of exchange, short-range parameters in MASTIFF (Equation (2)) were fit to the corresponding DFT-SAPT benchmark energy. In the case of exchange, however, we added the difference between the CCSD(T) and SAPT total interaction energies, $\delta(CC)$, to the SAPT exchange energy as follows:

$$E_{\text{exch}} = E_{\text{exch}}^{\text{DFT-SAPT}} + \delta(CC) \quad (7)$$

$$\delta(CC) = E_{\text{int}}^{\text{CCSD(T)}} - E_{\text{int}}^{\text{DFT-SAPT}} \quad (8)$$

This correction scheme ensured that the total benchmark energy corresponded to the total DF-FNO-CCSD(T)/CBS(a[TQ]Z, δ aDZ) energy, and was used to eliminate small inaccuracies due to method or basis set limitations with DFT-SAPT. All benchmark geometries and energies can be found in the SI.

4.4 Simulations

Condensed-phase benzene simulations were performed in a 600-monomer box. Unless otherwise stated, all simulations were run in an NPT ensemble at 1.0 atm and 298 K. Sample run scripts can be found in the SI. Initial configurations were generated using packmol⁸⁹ in a $50^3 \text{ \AA}^3$ box. Temperature was regulated using the stochastic Langevin integrator with a 2.0 ps^{-1} friction coefficient and 1.0 fs stepsize. Pressure was controlled via Monte Carlo Barostat with a 25 fs coupling time. The simulation box was allowed to equilibrate for 2 ns prior to 10 ns production, where statistics were collected every 1 ps for bulk properties. Liquid data

collection occurred at a minimum rate of 1.5 ns/day on a single GTX 1080 GPU card. Gas phase simulations were performed using the same protocol but with an initial density of 5.0 kg m⁻³. NVT simulations used the experimental room-temperature liquid density of 873.6 kg m⁻³.⁹⁰

Self-diffusion coefficients (D) were computed using the Einstein relation, as in Eq. 9,⁹¹ averaging over correlations in center-of-mass position $\mathbf{r}$.

$$D = \frac{1}{6} \frac{d}{dt} \langle r(0)r(t) \rangle^2 \quad (9)$$

To account for box size effects, the diffusion coefficient was calculated at constant volume for 4 system sizes containing 600, 450, 300, and 150 monomers. In the case of a higher-density model, 180 monomers were used instead of 150 to ensure cutoffs were less than half the box length. A linear fit of the diffusion coefficient vs. inverse box length was performed, and the bulk diffusion coefficient is reported as the y-intercept of this fit. Full details of the extrapolation procedure and the linear fits can be found in the SI.

System density was computed via thermodynamic average of the production NPT data set. Radial distribution functions of both center-of-mass and carbon-carbon distances were computed by histogram analysis.⁹¹

Enthalpy of vaporization was computed via Equation (10), assuming that gas and liquid intra-molecular energies were equal:

$$\Delta H_{vap} = -E_{intra,g} - (E_{inter,liq} + (E_{intra,liq})) = -E_{inter,liq} + RT \quad (10)$$

The heat capacity, C_P , was computed by separately accounting for inter- and intra-molecular contributions:

$$C_P = \left(\frac{\delta H_{inter}}{\delta T} \right)_P + C_P^{IG} - R \quad (11)$$

This strategy avoids spurious contributions from high-frequency intramolecular degrees of

freedom. The first term, representing intermolecular contributions to C_P , was computed by taking the slope of the intermolecular component of enthalpy, H_{inter} , across 5 evenly spaced temperatures from 300K to 450K. Intramolecular vibrational contributions were accounted for via the ideal gas heat capacity, here $C_P^{IG} = 19.5 \text{ cal mol}^{-1} \text{ K}^{-1}$.⁹² The gas constant, R , was subtracted so as not to double-count pressure-volume contributions.

Cohesive energies (E_{coh}) were computed by subtracting the gas-phase monomer energy from the 138K crystal lattice energy obtained by neutron diffraction.⁹³ E_{coh} was extrapolated to 0K by invoking the energetic difference for crystal lattice relaxation from 138K to 4K as calculated by local coupled cluster [OSV-LCCSD(T0)-F12] by Yang et al.³⁷.

Second virial coefficients (B_2) were calculated following the protocol described by McDaniel and Schmidt⁴³ as follows:

$$B_2 = 2\pi \int_0^\infty (1 - \langle \exp[-\beta U_{\text{inter}}(r)] \rangle) r^2 dr \quad (12)$$

where r is the center-of-mass distance. 10^5 random dimer orientations were used for each distance r at 0.001 nm intervals from 0.15-3.0nm, having confirmed B_2 values did not change significantly outside this range. To enforce an infinite repulsive wall at small distances, a step potential was added to the force expression as a function of van-der-Waals radii of interacting atoms.

5 Results & Discussion

5.1 Dimer Properties

Using the same strategy from prior work,^{45,46} the MASTIFF force field for benzene was fit on a component-by-component basis to a quasi-random sample⁴⁵ of 1000 benzene dimer configurations, encompassing a large range of energetically-relevant interatomic distances and orientations. The QM benchmark was taken to be FNO-CCSD(T)/CBS for the total

energy, and SAPT/aVTZ+m for the individual components. All benchmark geometries and associated energies can be found in the SI. The overall fit quality with respect to the QM benchmark is shown in Figure 2. Overall root-mean-squared (RMS) errors are 0.13 kcal/mol. If we only consider the RMSE for “net-attractive” configurations with $E_{CCSD(T)} < 0$ (labeled ‘aRMSE’), which serves as a rough proxy for the configurations expected to be energetically-relevant at room temperature, this error drops even further to 0.08 kcal/mol. As shown in the Supporting Information (Figure S4), this small overall RMS error originates from equally small (< 1 kcal/mol) errors across all energy components, with the exchange energy component perhaps being the largest contributor. The ability of MASTIFF to reproduce energetic components to this degree of accuracy is consistent with our prior work on smaller molecular dimers.⁴⁶ Moreover, such fidelity to each energy component stands in contrast to most standard force fields, which usually do not ensure a proper balance of exchange, electrostatics, dispersion, and induction.²⁸

Figure 3 compares the total energy and fit quality of MASTIFF to a variety of previously-published general benzene force fields: AMBER, AMOEBA, CHARMM, OPLS-AA, OPLS-CS, and OPT-FF. (This figure also shows results for ISO-MASTIFF and the POT3 potential energy surface; however, these comparisons will be discussed separately). RMSE and aRMSE for MASTIFF are typically an order of magnitude smaller than published general force fields. Aside from MASTIFF (RMSE=0.13 kcal/mol, aRMSE=0.08 kcal/mol), OPT-FF (RMSE=0.43 kcal/mol, aRMSE=0.29 kcal/mol) and AMOEBA (RMSE=0.82 kcal/mol, aRMSE=0.38 kcal/mol) have the best performance. AMOEBA includes some degree of anisotropy in its functional form, and OPT-FF was trained entirely against dimer energies, so these results are perhaps unsurprising. Standard empirical force fields (including OPLS-AA, arguably the best-performing model for bulk properties) all have RMSE larger than 1 kcal/mol and aRMSE larger than 0.6 kcal/mol. Additionally, these standard force fields are systematically over-repulsive compared to CCSD(T), making them overall poor models for predicting dimer energies. By contrast, the significantly reduced relative errors of MAS-

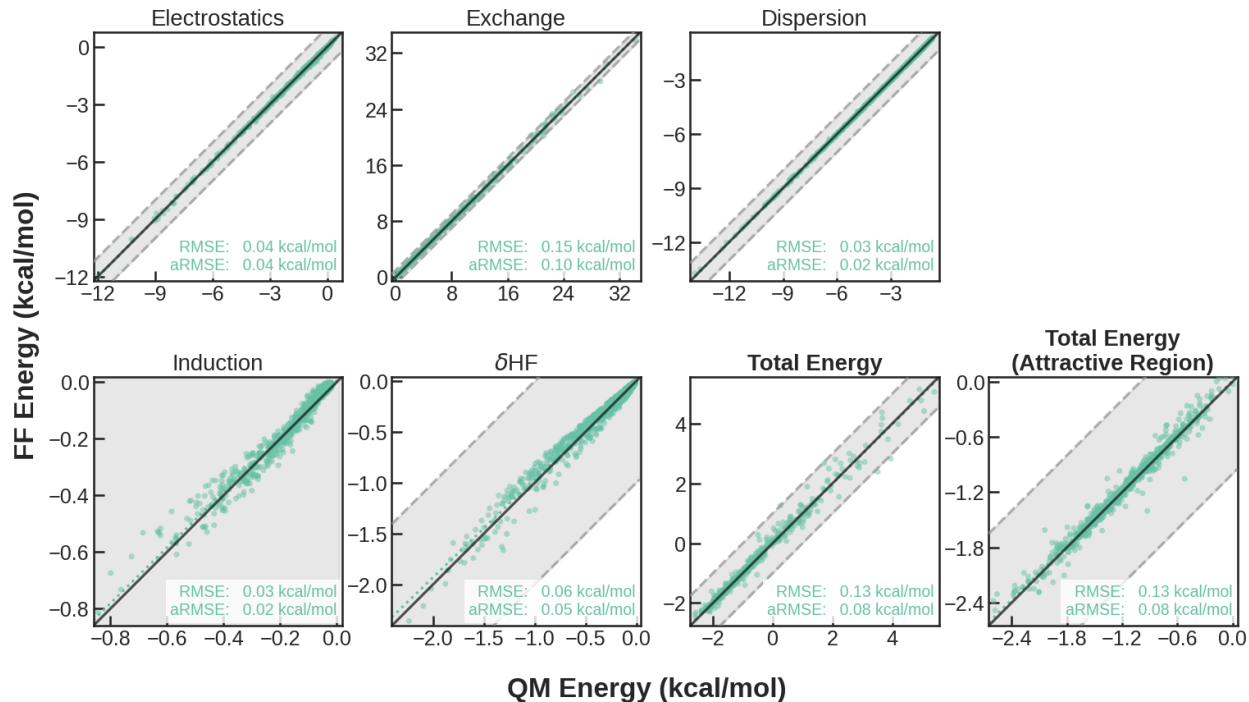


Figure 2: Force field fits for MASTIFF. Benchmark QM energies (x -axis) are FNO-CCSD(T)/CBS for the Total Energy, Equation (7) for the Exchange energy, and DFT-SAPT/aVTZ+m for all other energy components. The $y = x$ line represents perfect agreement between reference QM energies and MASTIFF, and gray regions represent points within ± 1 kcal/mol of the benchmark.

TIFF showcase the benefit of a fully-anisotropic *ab-initio* approach to predicting gas-phase dimer energies.

Since MASTIFF’s functional form contains several physics-based improvements — atomic-level anisotropy, charge penetration, advanced short-range repulsion functional forms, and explicit polarization — compared to standard LJ+ q force fields, it is worth taking a moment to analyze the extent to which atomic-level anisotropy by itself is responsible for MASTIFF’s accuracy. To this end, we have separately parameterized ISO-MASTIFF, an isotropic analogue of the MASTIFF force field that contains terms for explicit polarization, charge penetration, and Slater-ISA exchange, but that does NOT contain explicit terms related to atomic-level anisotropy (namely higher-order multipole moments and short-range orientation-dependent prefactors). Compared to standard literature force fields, ISO-MASTIFF performs quite well (RMSE=0.22 kcal/mol, aRMSE=0.14 kcal/mol), and is a fac-

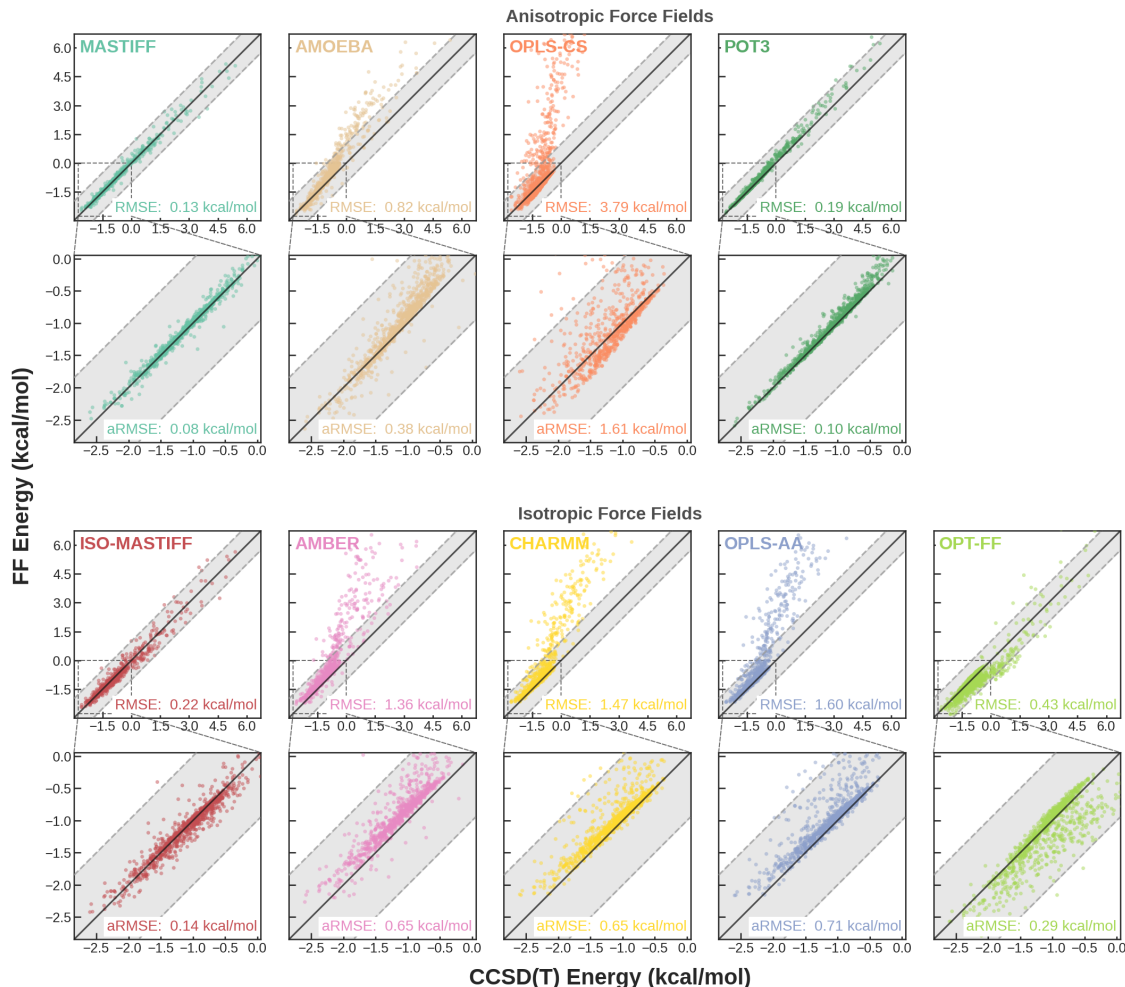


Figure 3: Comparison of force field fits vs. reference FNO-CCSD(T)/CBS Energies. Gray regions represent points within ± 1 kcal/mol of the benchmark. Root-mean squared errors (RMSE) are shown for each force field along with attractive root-mean squared errors (aRMSE) in which only points with $E_{\text{CCSD(T)}} < 0$ are included.

tor of two more accurate than OPT-FF, the best performing LJ+ q model our dimer training set. This result shows how accuracy improvements over standard LJ+ q models can be found even in the absence of anisotropy considerations. However, RMSE and aRMSE for ISO-MASTIFF are also a factor of two *worse* than MASTIFF. Taken together, these results highlight the idea that next-generation force fields require a number of sophisticated features in their functional forms to achieve sub-kJ/mol accuracy. We also argue that atomic-level anisotropy is one such feature that is of critical importance for molecules like benzene.

Figure 3 also compares MASTIFF to POT3, a high-accuracy benzene potential energy

surface developed by van der Avoird and co-workers. POT3 is entirely *ab-initio* and is fit to SAPT(DFT)/aVTZ+m data, which (disregarding minor differences in method and basis set) is very similar to the MASTIFF approach. Additionally, POT3 employs many of the same (or similar) asymptotically-correct functional forms to describe long-range interactions. Because POT3 uses off-atom interaction sites to achieve high-accuracy, we categorize POT3 as an *implicitly* anisotropic model, much like how popular 4- and 5-site water models use off-sites to implicitly mimic anisotropic electrostatic interactions.⁹⁴ The biggest difference between approaches is POT3’s use of explicit cross-terms in lieu of combination rules. This change significantly increases the number of free parameters (POT3 has 92 parameters for benzene compared to MASTIFF’s 49) and also makes more it challenging (vs. “general” force fields) to implement in standard MD software packages. Though not the focus of this work, we also note that POT3 and other explicit cross-term models, unlike MASTIFF,⁴⁶ are not directly transferable to mixed systems without substantial additional parameterization. However, the increased variational freedom that comes with explicit-cross terms can yield far more accurate fits to benchmark QM energies. For benzene, compared to the CCSD(T)/CBS benchmark, POT3 has very similar RMSE and aRMSE to MASTIFF. We point out (see ??) that some of POT3’s errors here may be due to slight differences in benchmark, and POT3 has closer to 0.02 kcal/mol precision compared to its own SAPT(DFT) training set. Still, given the greatly reduced number of parameters and the potential for transferability that comes with the MASTIFF approach, the fact that MASTIFF and POT3 are of similar accuracy is very promising.

In addition to quasi-randomly generated dimer configurations, geometries were optimized for four well-studied stationary points: sandwich (S), parallel-displaced (PD), T-shaped (T), and Y-shaped (Y). Minimum energies and associated geometry parameters are shown for each force field in Table 2 along with an overall RMSE comparison to a CCSD(T)/CBS benchmark. Single-point energy calculations for several additional stationary points can be found in the SI (??), however overall conclusions are the same as described below. Based on

the RMSE for the S, PD, T, and Y geometries, POT3 is the most accurate model, which is perhaps to be expected based on its highly-flexible parameterization approach. MASTIFF shows the best overall performance for both energies and intermolecular distances of all general force fields studied. This result emphasizes the robustness of the MASTIFF and its transferability to a diverse set of geometries outside of its training set. MASTIFF performs particularly well in capturing the perpendicular (T and Y) stationary points, in part due to its inclusion of anisotropic functional forms. This is evidenced by Figure 4, which shows by contrast how ISO-MASTIFF significantly under-predicts the binding energy curve for the T geometry.

The PD stationary point and associated potential energy scan along the R2 coordinate (shown in Figure 4 as the distance of parallel displacement at fixed perpendicular distance) are not well described by any of the models besides POT3. Of the general force fields studied, only MASTIFF and AMOEBA are able to accurately predict the location of the energy minima with respect to R2; however, MASTIFF underestimates the minimum binding energy by 0.35 kcal/mol, and AMOEBA underestimates the barrier height by more than 2 kcal/mol. A SAPT energy decomposition along the R2 coordinate (Figure 5) suggests that errors in the energy minima for MASTIFF primarily arise from the exchange and δ HF components. Higher-order anisotropic and/or induction contributions are expected to be more relevant in configurations where electron-rich regions are in close proximity, such as with the parallel configurations found in the S and PD stationary points. The isotropic induced-dipole model used by MASTIFF to account for induction does not explicitly account for these contributions; moreover, the exchange energy benchmark also accounts for the δ CCSD(T) term and thus could implicitly include higher-order induction effects. We hypothesize that these limitations (both the functional form and the ambiguity in the QM energy-decomposition) are responsible for the observed errors.

As a last comparison of dimer properties, the second virial coefficient, B_2 , was calculated as a function of temperature for each of the studied models and compared to experiment

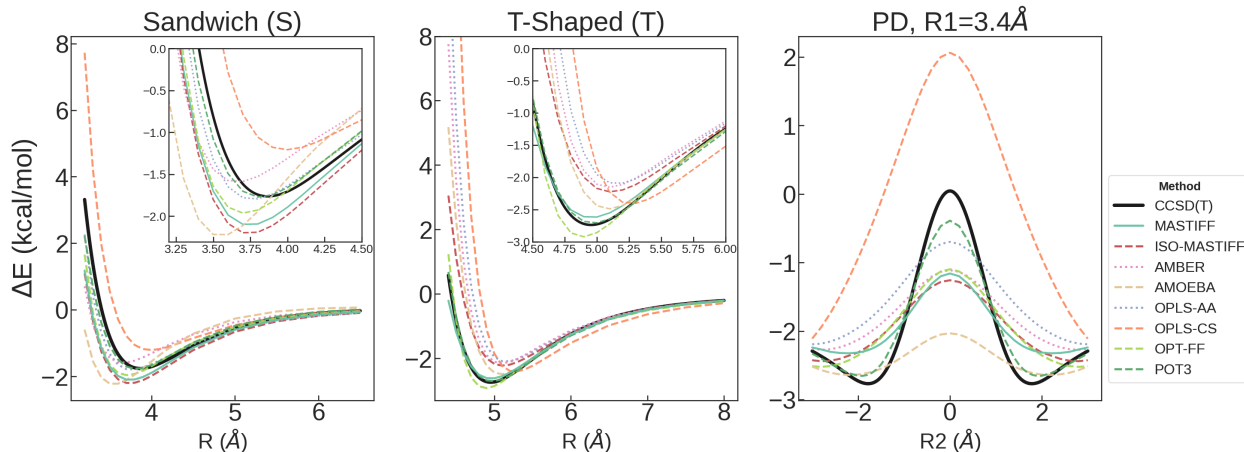


Figure 4: Potential energy curve for dimer geometries, scanned along R (or along R_2 for fixed R_1 for PD geometry). Data for CHARMM overlaps with OPLS-AA and is not shown for clarity.

(Figure 6). Several models, including MASTIFF, closely match the experimental value of B_2 across the full temperature range. MASTIFF underestimates the magnitude of B_2 by a modest $30\text{--}86\text{ cm}^3\text{mol}^{-1}$, while ISO-MASTIFF slightly overestimates this magnitude by $8\text{--}87\text{ cm}^3\text{mol}^{-1}$. POT1 (a pre-cursor to POT3 with similar parameters; data are not available for POT3) has the best performance by a small margin ($0.1\text{--}44\text{ cm}^3\text{mol}^{-1}$ error). Empirical models tend to underestimate the magnitude of B_2 , which is expected based on the observed underbinding for these models from Figure 3. OPT-FF and OPLS-CS perform well at higher temperatures but overestimate the virial at low T , which is again in keeping with the results of Figure 3.

5.2 Bulk Properties

In addition to gas-phase properties, condensed-phase simulations were carried out to compare and evaluate MASTIFF’s performance in the bulk. (We are unable to directly evaluate the bulk properties of POT3 due to its use of explicit cross terms and the implicit requirement of rigid body dynamics that comes with the extensive use of off-atom sites, neither of which are natively supported by OpenMM.) Table 3 reports results for a variety of room-temperature liquid properties. Of these properties, two — density, ρ , and enthalpy of vaporization,

Table 2: Optimized geometries and energies for the S, PD, T, and Y dimer configurations. Root mean squared errors (RMSE) for energies (in kcal/mol) and distances (in Å) are relative to the corresponding CCSD(T)/CBS benchmark. ^athis work. ^bRef. 18. ^cRef. 52. ^dRef. 95.



Model	ΔE_{RMSE}	R_{RMSE}	Sandwich		PD, R1=3.5Å		T-Shaped		Y-Shaped	
			ΔE	R	ΔE	$R2$	ΔE	R	ΔE	R
MASTIFF ^a	0.283	0.086	-2.11	3.74	-2.38	1.79	-2.63	4.95	-2.56	4.96
ISO-MASTIFF ^a	0.372	0.183	-2.22	3.74	-2.47	2.01	-2.23	5.11	-2.38	5.00
AMBER ^a	0.410	0.366	-1.74	3.64	-2.23	2.40	-2.09	5.10	-2.18	5.00
AMOEBA ^a	0.331	0.176	-2.24	3.55	-2.56	1.83	-2.49	5.09	-2.67	4.98
CHARMM27 ^b	0.404	0.463	-1.83	3.76	-2.22	2.62	-2.11	5.14	-2.23	5.04
OPLS-AA ^b	0.426	0.489	-1.69	3.78	-2.10	2.68	-2.15	5.10	-2.24	5.02
OPLS-CS ^b	—	—	repulsive		-2.89	4.00	-2.15	5.10	-2.24	5.02
OPT-FF ^b	0.325	0.417	-1.95	3.70	-2.48	2.50	-2.92	4.90	-2.90	4.84
POT3 ^c	0.037	0.082	-1.77	3.82	-2.69	1.84	-2.71	4.97	-2.45	5.01
CCSD(T)/CBS ^d	—	—	-1.71	3.87	-2.73	1.71	-2.70	5.01	-2.40	5.00

i

ΔH_{vap} — are explicitly included in the parameterization training sets for all empirical force fields. It is therefore unsurprising that the best performing empirical force fields (AMOEBA, OPLS-AA, and CHARMM) reproduce both ρ and ΔH_{vap} to within 1% of experiment. Encouragingly, and despite the fact that no bulk properties were included in its training set, MASTIFF predicts both the density and enthalpy of vaporization with only 3.5% error. These errors are comparable to a number of popular empirical force fields¹⁸ and are significantly smaller than the corresponding errors for OPT-FF (the only other MD-compatible force field exclusively fit to gas-phase calculations).

Regarding MASTIFF’s density and enthalpy of vaporization, several features are worth note. First, the observed density overestimation is consistent with MASTIFF’s slightly overly-attractive second virial (Figure 6). Second, errors in ΔH_{vap} are strongly correlated with the density overprediction: when recalculated in an NVT simulation at the experimental density, MASTIFF’s ΔH_{vap} prediction is accurate to within experimental error. Lastly,

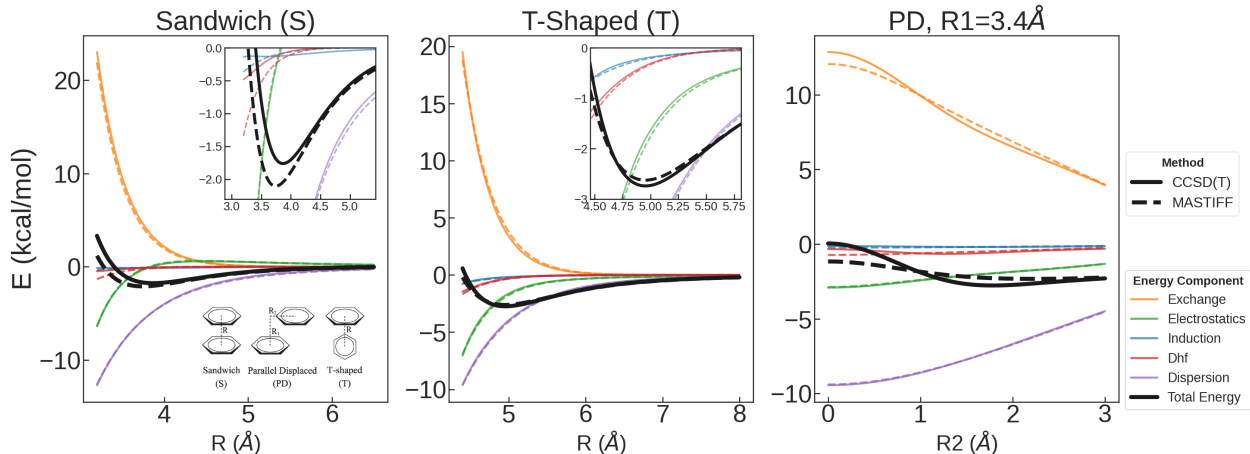


Figure 5: Energy decomposition of dimer geometries, scanned along R (or along R_2 for fixed R_1 for PD geometry).

MASTIFF’s density predictions are sensitive to 3-body dispersion contributions, but *not* to the inclusion of anisotropy. As expected,^{85,97} the ATM term (Equation (4)) has a net repulsive effect on the overall potential, and recomputing the density without 3-body dispersion results unfavorably in an additional 20 kg/m³ increase in the density prediction. However, predictions for ISO-MASTIFF (in which all anisotropic parameters, including electrostatic multipoles, are removed) are $\rho = 903.7$ kg/m³ and $\Delta H_{\text{vap}} = 8.47$ kcal/mol, essentially identical to that of MASTIFF. This result is somewhat surprising, both since ISO-MASTIFF predicts the second virial less accurately than MASTIFF, and because density predictions in our prior work on carbon dioxide⁴⁶ were found to be quite sensitive to the inclusion of anisotropy. However, experimental neutron diffraction has shown the neat benzene liquid to be “superficially” isotropic with a more complex underlying angular distribution function,^{19,35} and so it is possible that a combination of averaging and fortuitous error cancellation leads to the similarity in MASTIFF and ISO-MASTIFF’s densities despite differences in their underlying potentials.

In addition to the density and enthalpy of vaporization, we also used MASTIFF to simulate benzene’s radial distribution function, $g(r)$, self-diffusion coefficient, D , isobaric heat capacity, C_P , and cohesive energy, E_{coh} . These properties are generally not included in

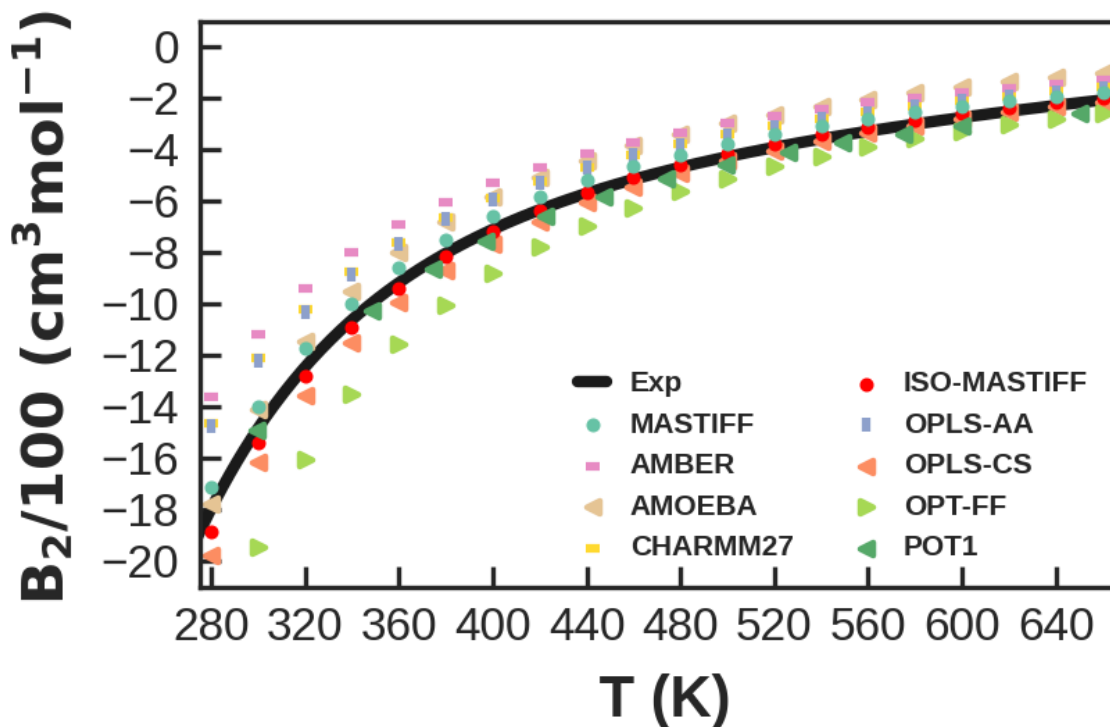


Figure 6: Second virial (B_2) coefficients. Experimental curve taken from Ref. 96, and POT1 data from Ref. 38. Points for CHARMM27 and OPLS-AA overlap and are virtually indistinguishable. Some points for OPT-FF extend below the graph.

the training sets for force fields (whether empirical or *ab initio*), and as out-of-fold predictions can thus serve as more rigorous “apples-to-apples” tests of each force field’s accuracy and transferability. Figure 7 shows $g(r)$ for both center-of-mass (COM) and carbon-carbon (CC) distances. Most models, including MASTIFF, reproduce the COM experimental benchmark very reliably. Somewhat larger errors can be seen in the CC $g(r)$. MASTIFF tends to be the most accurate force field at shorter length scales ($< 6\text{\AA}$), but under-predicts the correlation at larger length scales. The best-performing empirical force fields, by contrast, over-predict the 2nd coordination shell, but have better accuracy with respect to long-range order.

Regarding transport properties, the self-diffusion coefficient is underestimated by nearly all of the force fields studied. Best-performing models OPLS-AA and CHARMM27 have roughly 10% errors compared to experiment. MASTIFF’s self-diffusion coefficient performs

Table 3: Liquid Benzene Properties at 298K. For each property, the best performing model is shown in bold. ^aRef. 18 ^bRef. NIST 90 ^cRef. 98 ^dRef. 99 ^eMASTIFF NVT simulation at experimental density

Model	ρ (kg m ⁻³)	D (10 ⁻⁹ m ² s ⁻¹)	ΔH_{vap} (kcal mol ⁻¹)	C _p (cal mol ⁻¹ K ⁻¹)
MASTIFF	905.7 ± 5.6	1.86 ± 0.01	8.38 ± 0.15	30.0 ± 0.05
	—	(2.29 ± 0.08) ^e	(8.09) ^e	—
ISO-MASTIFF	903.7 ± 5.5	1.73 ± 0.06	8.47 ± 0.01	—
AMOEBA	874.8 ± 5.9	0.96 ± 0.14	8.11 ± 0.01	35.5 ± 0.2
AMBER ^a	835.9 ± 0.5	2.79 ± 0.22	7.24 ± 0.01	36.6 ± 1.3
OPLS-AA ^a	867.3 ± 0.4	1.97 ± 0.10	8.02 ± 0.01	33.3 ± 0.8
CHARMM27 ^a	870.4 ± 0.6	1.97 ± 0.17	8.17 ± 0.01	37.6 ± 1.2
OPLS-CS ^a	947.2 ± 0.03	not reported	14.86 ± 0.01	28.6 ± 0.05
OPT-FF ^a	1044.6 ± 0.6	0.35 ± 0.01	11.30 ± 0.02	35.2 ± 1.3
Exp	873.6 ^d	2.20 ^c	8.10 ± 0.02 ^b	32.4 ^d

only slightly worse, with an error of 15%. Notably, recalculating at the experimental density yields a much improved value of $2.29 \times 10^{-9} \text{m}^2 \text{s}^{-1}$, which is only a 4% error compared to experiment, and once again shows how most of our prediction error stems from our slightly overestimated density. MASTIFF’s results contrast with those for other *ab-initio* or mixed parameterization force fields, which have poor accuracy for self-diffusion predictions.

Most force fields adequately predict the experimental heat capacity, C_P . The performance of OPLS-AA is particularly excellent, reproducing experiment to within simulation uncertainty. MASTIFF is the second-best performing model. CHARMM27 had the poorest performance, giving an example of the limits to empirical methods when calculating properties to which they were not parameterized.

Because it lies well outside the temperature and phase region used in force field parameterization, and because it samples a very limited number of dimer orientations (in contrast to the liquid),^{19,102} benzene’s cohesive energy (see Table 4) is a particularly stringent test for a force field’s robustness and transferability. Furthermore, while the solid-phase is a known challenge for standard force fields and even QM methods, it is of central importance to applications such as crystal structure prediction and drug design.^{103,104} MASTIFF is the

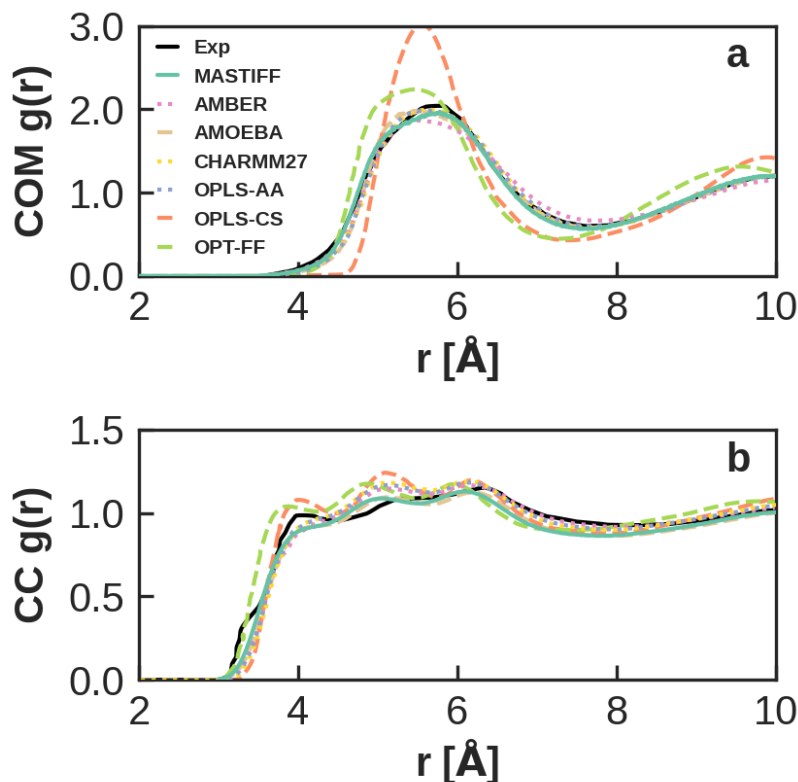


Figure 7: Pair distribution functions ($g(r)$) for (a) center-of-mass (COM) distances and (b) carbon-carbon (CC) distances. Experimental values taken for COM and CC distances from Refs. 35 and 100, respectively. OPT-FF, OPLS-AA, OPLS-CS, and CHARMM27 values were extracted from work by Fu and Tian¹⁸. ISO-MASTIFF is not shown but is virtually indistinguishable from MASTIFF.

best-performing approach for benzene’s cohesive energy, and with an error of 0.24 kcal/mol is well within experimental uncertainty. Energetic differences between solid-phase polymorphs can often be 2 kJ/mol or smaller,¹⁰⁵ and so the fact that MASTIFF (uniquely amongst force fields and QM methods compared in this work) can predict benzene’s cohesive energy within 1 kJ/mol (0.24 kcal/mol) accuracy shows great promise. By contrast, the hybrid QM/FF result from Podeszwa et al.,¹⁰¹ which used SAPT(DFT) to compute nearest-neighbor interactions and the POT3 potential to compute asymptotic dimer interactions, has errors in excess of 1 kcal/mol, demonstrating the difficulty for even high-accuracy QM methods to accurately predict lattice energies. As for the other force fields studied in this work,

Table 4: Cohesive Energy for Solid Benzene. The extrapolation procedure from 138K to 0K, described in the Methods section, is taken from Ref. 37. ^aData from Ref. 101, which uses POT3 for asymptotic dimer data and SAPT(DFT) for nearest-neighbor interactions. ^bRef. 37.

Model	$E_{\text{coh}}, 138 \text{ K}$ (kcal mol ⁻¹)	$E_{\text{coh}}, 0 \text{ K}$ (kcal mol ⁻¹)
MASTIFF	-12.64	-12.96
ISO-MASTIFF	-12.21	-12.53
AMOEBA	-10.90	-11.21
AMBER	-11.60	-11.91
OPLS-AA	-12.21	-12.52
CHARMM27	-12.20	-12.52
OPLS-CS	-13.39	-13.70
OPT-FF	-13.99	-14.30
SAPT(DFT)/POT3 ^a	-11.81	-12.13
Exp ^b	—	-13.2 $\pm$ 0.5

OPLS-CS is the only other method which falls within the experimental error bars, at their very upper limit of 0.50 kcal/mol. OPLS-AA, CHARMM, and ISO-MASTIFF perform adequately, each with errors of roughly 0.7 kcal/mol. The decreased accuracy of ISO-MASTIFF relative to MASTIFF matches results from our prior work on CO₂,⁴⁶ and highlights the increased importance of anisotropy in condensed-phase property predictions where only a limited distribution of dimer orientations are sampled. Like SAPT(DFT), AMOEBA, AMBER, OPLS-CS, and OPT-FF all have errors larger than 1 kcal/mol and thus cannot be considered quantitatively accurate.

6 Conclusion

In 2006, Baker and Grant (creators of OPLS-CS) eloquently summed up the long-standing challenge of first-principles approaches to general force field development for π -contact interactions: “models of increasing sophistication will be required before we can truly hope to model the full range of aromatic interactions with complete confidence”.¹⁹ Though benzene

is but one proof-of-concept example, we believe the results presented in this work represent significant progress towards this goal.

Our MASTIFF benzene model is simple enough to be used as a general force field while using sophisticated functional forms that account for the essential physics of aromatic interactions. In line with some other “next-generation” general force fields,^{51,53,68,106} MASTIFF accurately accounts for effects of charge penetration, explicit polarization, multipolar electrostatics, and short-range exchange-repulsion. Moreover, and is often neglected in the literature, MASTIFF explicitly accounts for atomic-level anisotropy of the electron density at both short- and long-range. Lastly, via the open-source CAN plugin for OpenMM introduced and described in this work, MASTIFF can be used to simulate not only clusters, but also condensed phases. The CAN plugin is completely general, and is intended to be broadly useful for the molecular modeling community to enable condensed-phase simulations using arbitrary anisotropic functional forms (not just MASTIFF).

Despite only using monomer and dimer properties in its training set, we have shown how our first-principles MASTIFF benzene force field can be used to accurately model the gas, liquid, *and* solid phases. Compared to standard force fields, MASTIFF offers strong performance (sub kJ/mol) accuracy for dimer energetics and geometries, while simultaneously capturing the emergent bulk liquid-phase properties (density, enthalpy of vaporization, self-diffusion, and heat capacity). Via the use of atomically anisotropic functional forms, the resulting force field can even capture challenging solid-phase cohesive energies, thus showing promise for use in crystal structure prediction and related fields.

It is interesting to contrast the *explicit* description of atomic anisotropy via the MASTIFF approach with an *implicit* approach taken via off-atom sites by potentials such as POT3. Both approaches certainly have competing merits and challenges. While the latter is somewhat conceptually simpler (and possibly easier to graft onto an existing rigid-body MD code), MASTIFF’s use of atom-centered parameters seems to suggest a higher possibility for transferability to mixtures and likely simplicity in parameterization.

Regardless of the approach, the results demonstrated herein clearly highlight the importance of atomic anisotropy to yield both high fidelity to the underlying dimer PES and to the resulting both properties. We thus hope that the results presented here may be a starting point for sophisticated *ab initio* force field development of general molecules and systems of interest, including, but not limited to, those containing π -contact interactions.

7 Author Information

Tesia D. Janicki and Mary J. Van Vleet contributed equally to this work.

Corresponding Author J.R. Schmidt (schmidt@chem.wisc.edu)

ORCID:

Tesia D Janicki: 0000-0001-6008-6414

J. R. Schmidt: 0000-0002-1438-117X

Mary Van Vleet 0000-0002-6526-3579

8 Acknowledgments

The authors wish to acknowledge Dr. Carlos H. Borca for contributing the CCSD(T) calculations performed in this work. Thanks are also given to Dr. Peter Eastman and Dr. David Sherrill for helpful discussions.

Computing Resources were provided through the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology (Atlanta, Georgia, USA), the UW–Madison Chemistry Department Phoenix cluster (under Grant No. CHE-0840494), the UW–Madison Starling cluster, and the UW–Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences (under Grant No. CHE-084049).

This material is based upon work supported by the National Science Foundation under Grant No. HRD-2101190 (MJVV). MJVV and TDJ acknowledges support from the Na-

tional Science Foundation Graduate Research Fellowship under Grants Nos. DGE-1747503 and DGE-1256259, and partial support by NSF through the University of Wisconsin Materials Research Science and Engineering Center (DMR-1720415). TDJ also acknowledges the support provided by the Graduate School and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation.

9 Supporting Information

OpenMM input files for benzene example. Benchmark energies and geometries for MASTIFF force field. CAN plugin description and usage. Additional analysis of benzene dimer energetics. Analysis of finite size effects in diffusion.

10 TOC Graphic

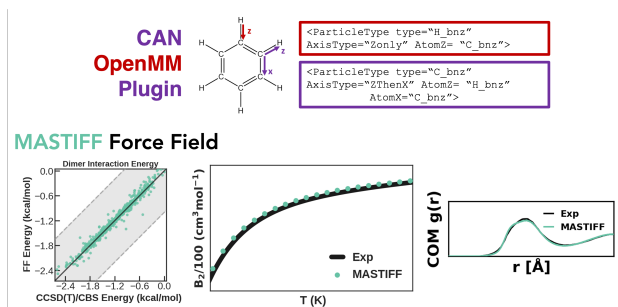


Figure 8: TOC Graphic

References

- (1) Thakuria, R.; Nath, N. K.; Saha, B. K. The Nature and Applications of π - π Interactions: A Perspective Published as part of a Crystal Growth and Design virtual special issue on π - π Stacking in Crystal Engineering: Fundamentals and Applications. *Growth Des* **2019**, *19*, 33.
- (2) Ninković, D. B.; Blagojević Filipović, J. P.; Hall, M. B.; Brothers, E. N.; Zarić, S. D. What Is Special about Aromatic-Aromatic Interactions? Significant Attraction at Large Horizontal Displacement. *ACS Central Science* **2020**, *6*, 420–425.
- (3) Šponer, J.; Šponer, J. E.; Mládek, A.; Jurečka, P.; Banáš, P.; Otyepka, M. Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics and experiment. *Biopolymers* **2013**, *99*, 978–988.
- (4) Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with aromatic rings in chemical and biological recognition. *Angewandte Chemie - International Edition* **2003**, *42*.
- (5) Riley, K. E.; Hobza, P. On the importance and origin of aromatic interactions in chemistry and biodisciplines. *Accounts of Chemical Research* **2013**, *46*, 927–936.
- (6) Tsuzuki, S.; Uchimaru, T.; Sugawara, K.-I. Energy profile of the interconversion path between T-shape and slipped-parallel benzene dimers. *J. Chem. Phys* **2002**, *117*, 11216.
- (7) Šponer, J.; Banáš, P.; Jurečka, P.; Zgarbová, M.; Kührová, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M. Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. *Journal of Physical Chemistry Letters* **2014**, *5*, 1771–1782.

- (8) Poletto, M. D.; Rusu, V. H.; Grisci, B. I.; Dorn, M.; Lins, R. D.; Verli, H. Aromatic rings commonly used in medicinal chemistry: Force fields comparison and interactions with water toward the design of New Chemical Entities. *Frontiers in Pharmacology* **2018**, *9*.
- (9) Burley, S. K.; Petsko, G. A. Aromatic-Aromatic Interaction: A Mechanism of Protein Structure Stabilization. *Science* **1985**, *229*, 23–28.
- (10) Kim, E. I.; Paliwal, S.; Wilcox, C. S. Measurements of Molecular Electrostatic Field Effects in Edge-to-Face Aromatic Interactions and CH- π Interactions with Implications for Protein Folding and Molecular Recognition. *Journal of the American Chemical Society* **1998**, *120*, 11192–11193.
- (11) Vernon, R. M.; Chong, P. A.; Tsang, B.; Kim, T. H.; Bah, A.; Farber, P.; Lin, H.; Forman-Kay, J. D. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. **2018**,
- (12) Hunter, C. A.; Sanders, J. K. M. The Nature of pi-pi Interactions. *J. Am. Chem. Soc.* **1990**, *112*, 5525–5534.
- (13) Gervasio, F. L.; Chelli, R.; Procacci, P.; Schettino, V. The Nature of Intermolecular Interactions Between Aromatic Amino Acid Residues. *PROTEINS* **2002**, *48*, 117–125.
- (14) Janda, K. C.; Hemminger, J. C.; Winn, J. S. Benzene dimer: A polar molecule. *J. Chem. Phys* **1975**, *63*, 1419.
- (15) Muehldorf, A. V.; Engen, D. V.; Warner, J. C.; Hamilton, A. D. Aromatic-aromatic interactions in molecular recognition: A family of artificial receptors for thymine that shows both face-to-face and edge-to-face orientations. *Journal of the American Chemical Society* **1988**, *110*, 6561–6562.

- (16) Shetty, A. S.; Zhang, J.; Moore, J. S. Aromatic π -stacking in solution as revealed through the aggregation of phenylacetylene macrocycles. *Journal of the American Chemical Society* **1996**, *118*, 1019–1027.
- (17) Headen, T. F.; Cullen, P. L.; Patel, R.; Taylor, A.; Skipper, N. T. The structures of liquid pyridine and naphthalene: The effects of heteroatoms and core size on aromatic interactions. *Physical Chemistry Chemical Physics* **2018**, *20*.
- (18) Fu, C.-F.; Tian, S. X. A Comparative Study for Molecular Dynamics Simulations of Liquid Benzene. *Journal of Chemical Theory and Computation* **2011**, *7*, 2240–2252.
- (19) Baker, C. M.; Grant, G. H. The structure of liquid benzene. *Journal of Chemical Theory and Computation* **2006**, *2*, 947–955.
- (20) Price, D. J.; Brooks, C. L. Detailed considerations for a balanced and broadly applicable force field: A study of substituted benzenes modeled with OPLS-AA. *Journal of Computational Chemistry* **2005**, *26*, 1529–1541.
- (21) Engkvist, O.; Hobza, P.; Selzle, H. L.; Schlag, E. W. Benzene trimer and benzene tetramer: Structures and properties determined by the nonempirical model (NEMO) potential calibrated from the CCSD(T) benzene dimer energies. *Journal of Chemical Physics* **1999**, *110*, 5758–5762.
- (22) Chessari, G.; Hunter, C. A.; Low, C. M. R.; Packer, M. J.; Vinter, J. G.; Zonta, C. An evaluation of force-field treatments of aromatic interactions. *Chemistry - A European Journal* **2002**, *8*, 2860–2867.
- (23) Garrido, N. M.; Jorge, M.; Queimada, A. J.; Gomes, J. R.; Economou, I. G.; MacEdo, E. A. Predicting hydration Gibbs energies of alkyl-aromatics using molecular simulation: A comparison of current force fields and the development of a new parameter set for accurate solvation data. *Physical Chemistry Chemical Physics* **2011**, *13*, 17384–17394.

- (24) Paton, R. S.; Goodman, J. M. Hydrogen Bonding and π -Stacking : How Reliable are Force Fields? A Critical Evaluation of Force Field Descriptions of Nonbonded Interactions Hydrogen Bonding and π -Stacking : How Reliable are Force Fields ? A Critical Evaluation of Force Field Desc. **2009**, 944–955.
- (25) Aschi, M.; Mazza, F.; Di Nola, A. Cation- π interactions between ammonium ion and aromatic rings: An energy decomposition study. *Journal of Molecular Structure: THEOCHEM* **2002**, 587, 177–188.
- (26) Dang, L. X. Molecular dynamics study of benzene–benzene and benzene–potassium ion interactions using polarizable potential models. *The Journal of Chemical Physics* **2000**, 113, 266–273.
- (27) Smith, G. D.; Jaffe, R. L. Comparative Study of Force Fields for Benzene. *The Journal of Physical Chemistry* **1996**, 100, 9624–9630.
- (28) Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. R. Assessment of standard force field models against high-quality ab initio potential curves for prototypes of π - π , CH/ π , and SH/ π interactions. *Journal of Computational Chemistry* **2009**, 30, NA–NA.
- (29) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H₂S $\dagger$. *The Journal of Physical Chemistry A* **2009**, 113, 10146–10159.
- (30) Sinnokrot, M. O.; Sherrill, C. D. Highly Accurate Coupled Cluster Potential Energy Curves for the Benzene Dimer: Sandwich, T-Shaped, and Parallel-Displaced Configurations. *The Journal of Physical Chemistry A* **2004**, 108, 10200–10207.

- (31) Kennedy, M. R.; McDonald, A. R.; DePrince, A. E.; Marshall, M. S.; Podeszwa, R.; Sherrill, C. D. Communication: Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory. *The Journal of Chemical Physics* **2014**, *140*, 121104.
- (32) wun Hwang, J.; Li, P.; Shimizu, K. D. Synergy between experimental and computational studies of aromatic stacking interactions. *Organic and Biomolecular Chemistry* **2017**, *15*, 1554–1564.
- (33) Řezáč, J.; Nachtigallová, D.; Mazzoni, F.; Pasquini, M.; Pietraperzia, G.; Becucci, M.; Müller-Dethlefs, K.; Hobza, P. Binding energies of the π -stacked anisole dimer: New molecular beam - Laser spectroscopy experiments and ccSD(t) calculations. *Chemistry - A European Journal* **2015**, *21*.
- (34) Carter-Fenk, K.; Herbert, J. M. Electrostatics does not dictate the slip-stacked arrangement of aromatic π - π interactions. *Chemical Science* **2020**, 6758–6765.
- (35) Headen, T. F.; Howard, C. A.; Skipper, N. T.; Wilkinson, M. A.; Bowron, D. T.; Soper, A. K. Structure of π - π Interactions in Aromatic Liquids. *J. Am. Chem. Soc* **2010**, *132*, 5735–5742.
- (36) Tauer, T. P.; Sherrill, C. D. Beyond the Benzene Dimer: An Investigation of the Additivity of π - π Interactions. *The Journal of Physical Chemistry A* **2005**, *109*, 10475–10478.
- (37) Yang, J.; Hu, W.; Usvyat, D.; Matthews, D.; Schütz, M.; Chan, G. K.-L. Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Science* **2014**, *345*, 640–643.
- (38) Podeszwa, R.; Bukowski, R.; Szalewicz, K. Potential energy surface for the benzene dimer and perturbational analysis of π - π interactions. *Journal of Physical Chemistry A* **2006**, *110*, 10345–10354.

- (39) Totton, T. S.; Misquitta, A. J.; Kraft, M. A First Principles Development of a General Anisotropic Potential for Polycyclic Aromatic Hydrocarbons. *Journal of Chemical Theory and Computation* **2010**, *6*, 683–695.
- (40) Liu, Y. P.; Kim, K.; Berne, B. J.; Friesner, R. A.; Rick, S. W. Constructing ab initio force fields for molecular dynamics simulations. *Journal of Chemical Physics* **1998**, *108*, 4739–4755.
- (41) Naseem-Khan, S.; Gresh, N.; Misquitta, A. J.; Piquemal, J.-P. Assessment of SAPT and Supermolecular EDA Approaches for the Development of Separable and Polarizable Force Fields. *Journal of Chemical Theory and Computation* **2021**, *17*, 2759–2774.
- (42) Stone, A. J.; Misquitta, A. J. Atom–atom potentials from ab initio calculations. *International Reviews in Physical Chemistry* **2007**, *26*, 193–222.
- (43) McDaniel, J. G.; Schmidt, J. Physically-Motivated Force Fields from Symmetry-Adapted Perturbation Theory. *The Journal of Physical Chemistry A* **2013**, *117*, 2053–2066.
- (44) Stone, A. J. *Non-Covalent Interactions in Quantum Chemistry and Physics*, 1st ed.; Elsevier Inc., 2017; pp 3–26.
- (45) 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. *Journal of Chemical Theory and Computation* **2016**, *12*, 3851–3870.
- (46) Van Vleet, M. J.; Misquitta, A. J.; Schmidt, J. New Angles on Standard Force Fields: Toward a General Approach for Treating Atomic-Level Anisotropy. *Journal of Chemical Theory and Computation* **2018**, *14*, 739–758.
- (47) Misquitta, A. J.; Stone, A. J. Ab Initio Atom-Atom Potentials Using CamCASP:

- Theory and Application to Many-Body Models for the Pyridine Dimer. *Journal of Chemical Theory and Computation* **2016**, *12*, 4184–4208.
- (48) Stone, A. J.; Price, S. L. Some new ideas in the theory of intermolecular forces: anisotropic atom-atom potentials. *The Journal of Physical Chemistry* **1988**, *92*, 3325–3335.
- (49) Price, S. L. Toward More Accurate Model Intermolecular Potentials for Organic Molecules. *Reviews in Computational Chemistry* **2000**, *14*, 225–289.
- (50) Zhang, C.; Bell, D.; Harger, M.; Ren, P. Polarizable Multipole-Based Force Field for Aromatic Molecules and Nucleobases. *Journal of Chemical Theory and Computation* **2017**, *13*, 666–678.
- (51) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *Journal of Physical Chemistry B* **2003**, *107*, 5933–5947.
- (52) Van der Avoird, A.; Podesszwa, R.; Szalewicz, K.; Leforestier, C.; van Harrevelt, R.; Bunker, P. R.; Schnell, M.; von Helden, G.; Meijer, G. Vibration–rotation–tunneling states of the benzene dimer: an ab initio study. *Physical Chemistry Chemical Physics* **2010**, *12*, 8219–8240.
- (53) Rackers, J. A.; Ponder, J. W. Classical Pauli repulsion: An anisotropic, atomic multipole model. *The Journal of Chemical Physics* **2019**, *150*, 084104.
- (54) Amos, R. D.; Andrews, J. S.; Handy, N. C.; Knowles, P. J. Open Shell $\{M\{\{\emptyset\}\}$ ller–Plesset} Perturbation Theory. *Chem. Phys. Letters* **1991**, *185*, 256–264.
- (55) Martinez, C. R.; Iverson, B. L. Rethinking the term ”pi-stacking”. *Chem. Sci* **2012**, *3*, 2191.
- (56) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.;

- et al., OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology* **2017**, *13*.
- (57) Eastman, P. et al. OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *Journal of Chemical Theory and Computation* **2013**, *9*, 461–469.
- (58) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry* **2003**, *24*, 1999–2012.
- (59) Ponder, J. W.; Wu, C.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; Distasio, R. A.; Head-gordon, M.; et al., Current Status of the AMOEBA Polarizable Force Field. *Journal of Physical Chemistry B* **2010**, *114*, 2549–2564.
- (60) Mackerell, A. D. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (61) Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry* **2004**, *25*, 1584–1604.
- (62) Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
- (63) Jorgensen, W. L.; Severance, D. L.; C Proc R Soc London A, J. A. Selected Values of Physical and Thermodynamic Properties of Hy-drocarbons and Related Compounds. *J. Am. Chem. Soc* **1990**, *112*, 1630.

- (64) Prampolini, G.; Livotto, P. R.; Cacelli, I. Accuracy of Quantum Mechanically Derived Force-Fields Parameterized from Dispersion-corrected DFT data: the Benzene Dimer as Prototype for Aromatic Interactions. *Journal of Chemical Theory and Computation* **2015**, 5182–5196.
- (65) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *Journal of Computational Chemistry* **2004**, 25, 1157–1174.
- (66) 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.
- (67) Ren, P.; Wu, C.; Ponder, J. W. Polarizable atomic multipole-based molecular mechanics for organic molecules. *Journal of Chemical Theory and Computation* **2011**, 7, 3143–3161.
- (68) Liu, C.; Piquemal, J. P.; Ren, P. AMOEBA+ Classical Potential for Modeling Molecular Interactions. *Journal of Chemical Theory and Computation* **2019**, 15, 4122–4139.
- (69) Das, A. K.; Urban, L.; Leven, I.; Loipersberger, M.; Aldossary, A.; Head-Gordon, M.; Head-Gordon, T. Development of an Advanced Force Field for Water Using Variational Energy Decomposition Analysis. *Journal of Chemical Theory and Computation* **2019**, 15, 5001–5013.
- (70) Das, A. K.; Demerdash, O. N.; Head-Gordon, T. Improvements to the AMOEBA Force Field by Introducing Anisotropic Atomic Polarizability of the Water Molecule. *Journal of Chemical Theory and Computation* **2018**, 14, 6722–6733.
- (71) McDaniel, J. G.; Schmidt, J. R. Next-Generation Force Fields from Symmetry-Adapted Perturbation Theory. *Annu. Rev. Phys. Chem.* **2016**, 67, 467–488.

- (72) Metz, M. P.; Szalewicz, K. Automatic Generation of Flexible-Monomer Intermolecular Potential Energy Surfaces. *Journal of Chemical Theory and Computation* **2020**, *16*, 2317–2339.
- (73) Metz, M. P.; Piszczatowski, K.; Szalewicz, K. Automatic Generation of Intermolecular Potential Energy Surfaces. *Journal of Chemical Theory and Computation* **2016**, acs.jctc.6b00913.
- (74) Akin-Ojo, O.; Szalewicz, K. Does a pair of methane molecules aggregate in water? *Journal of Chemical Physics* **2019**, *150*.
- (75) Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Götz, A. W.; Paesani, F. On the accuracy of the MB-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *The Journal of Chemical Physics* **2016**, *145*, 194504.
- (76) Riera, M.; Yeh, E. P.; Paesani, F. Data-Driven Many-Body Models for Molecular Fluids: CO₂/H₂O Mixtures as a Case Study. *Journal of Chemical Theory and Computation* **2020**, *16*, 2246–2257.
- (77) Bull-Vulpe, E. F.; Riera, M.; Götz, A. W.; Paesani, F. MB-Fit: Software infrastructure for data-driven many-body potential energy functions. *The Journal of Chemical Physics* **2021**, *155*, 124801.
- (78) Kim, Y. S.; Kim, S. K.; Lee, W. D. Dependence of the closed-shell repulsive interaction on the overlap of the electron densities. *Chemical Physics Letters* **1981**, *80*, 574–575.
- (79) Wheatley, R. J.; Price, S. L. An overlap model for estimating the anisotropy of repulsion. *Molecular Physics* **1990**, *69*, 507–533.

- (80) Misquitta, A. J.; Stone, A. J.; Fazeli, F. Distributed Multipoles from a Robust Basis-Space Implementation of the Iterated Stockholder Atoms Procedure. *Journal of Chemical Theory and Computation* **2014**, *10*, 5405–5418.
- (81) Tang, K. T.; Toennies, J. P. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *The Journal of Chemical Physics* **1984**, *80*, 3726–3741.
- (82) Tang, K. T.; Toennies, J. P. The damping function of the van der Waals attraction in the potential between rare gas atoms and metal surfaces. *Surface Science* **1992**, *279*, L203–L206.
- (83) Stone, A. J.; Tough, R. Spherical tensor theory of long-range intermolecular forces. *Chemical Physics Letters* **1984**, *110*, 123–129.
- (84) <https://github.com/jrschmidt2/OpenMM-CANplugin>.
- (85) McDaniel, J. G.; Schmidt, J. R. First-Principles Many-Body Force Fields from the Gas Phase to Liquid: A “Universal” Approach. *The Journal of Physical Chemistry B* **2014**, *118*, 8042–8053.
- (86) Anatole von Lilienfeld, O.; Tkatchenko, A. Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids. *The Journal of Chemical Physics* **2010**, *132*, 234109.
- (87) Wen, S.; Beran, G. J. O. Accurate molecular crystal lattice energies from a fragment QM/MM approach with on-the-fly Ab initio force field parametrization. *Journal of Chemical Theory and Computation* **2011**, *7*, 3733–3742.
- (88) Werner, H.-J. et al. MOLPRO, version 2012.1, a package of ab initio programs. 2012.
- (89) Martinez, L.; Andrade, R.; Bergin, E. G.; Martínez, J. M. Packmol: A package for

- building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.
- (90) NIST Chemistry WebBook. <https://webbook.nist.gov/cgi/cbook.cgi?ID=C71432&Mask=7#Thermo-Condensed>.
- (91) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: USA, 1987.
- (92) Riddick, J. A.; Bunger, W. B.; Sakano, T. K. *Techniques of Chemistry, Vol. II: Organic Solvents: Physical Properties and Methods of Purification*, 4th ed.; Wiley-Interscience: New York, 1986.
- (93) Bacon, G. E.; Curry, N. A.; Wilson, S. A. A crystallographic study of solid benzene by neutron diffraction. *Proc. R. Soc. Lond. A* **1964**, *279*, 98–110.
- (94) Tran, K. N.; Tan, M.-L.; Ichiye, T. A single-site multipole model for liquid water. *The Journal of Chemical Physics* **2016**, *145*, 034501.
- (95) Bludský, O.; Rubeš, M.; Soldán, P.; Nachtigall, P. Investigation of the benzene-dimer potential energy surface: DFT/CCSD(T) correction scheme. *Journal of Chemical Physics* **2008**, *128*.
- (96) Wormald, C. J. Second Virial Coefficient of Benzene and its Temperature Derivatives. *J. Chem. Soc., Faraday Trans. 1* **1975**, *71*, 726–735.
- (97) Yu, K.; Schmidt, J. R. Many-body effects are essential in a physically motivated CO₂ force field. *The Journal of Chemical Physics* **2012**, *136*, 034503.
- (98) Mccool, M. A.; Collings -F And, A. F.; Woolf, L. A. Pressure and Temperature Dependence of the Self-diffusion of Benzene. *J. Chem. Soc., Faraday Trans. 1* **1972**, *68*, 1489–1497.

- (99) Grolier, J. P.; Roux-Desgranges, G.; Berkane, M.; Jiménez, E.; Wilhelm, E. Heat capacities and densities of mixtures of very polar substances 2. Mixtures containing N, N-dimethylformamide. *The Journal of Chemical Thermodynamics* **1993**, *25*, 41–50.
- (100) Narten, A. H. X-ray diffraction pattern and models of liquid benzene. *J. Chem. Phys* **1977**, *67*, 2102.
- (101) Podeszwa, R.; Rice, B. M.; Szalewicz, K. Predicting structure of molecular crystals from first principles. *Physical Review Letters* **2008**, *101*, 1–4.
- (102) Jeffrey, G. A.; Ruble, J. R.; McMullan, R. K.; Pople, J. A.; Cox, E. G. The crystal structure of deuterated benzene. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **1987**, *414*, 47–57.
- (103) Price, S. L. Computational prediction of organic crystal structures and polymorphism. *International Reviews in Physical Chemistry* **2008**, *27*, 541–568.
- (104) Nguyen, A. L.; Mason, T. G.; Freeman, B. D.; Izgorodina, E. I. Prediction of lattice energy of benzene crystals: A robust theoretical approach. *Journal of Computational Chemistry* **2021**, *42*, 248–260.
- (105) Nyman, J.; Day, G. M. Static and lattice vibrational energy differences between polymorphs. *CrystEngComm* **2015**, *17*, 5154–5165.
- (106) Vandenbrande, S.; Waroquier, M.; Speybroeck, V. V.; Verstraelen, T. The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Noncovalent Interactions. *Journal of Chemical Theory and Computation* **2017**, *13*, 161–179.