The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Non-Covalent Interactions

Steven Vandenbrande, Michel Waroquier, Veronique Van Speybroeck, and Toon Verstraelen[*]

Center for Molecular Modeling (CMM), Ghent University, Technologiepark 903, B9000 Ghent, Belgium, (Member of the QCMM Ghent–Brussels Alliance)

E-mail: toon.verstraelen@ugent.be

Abstract

We propose a methodology to derive pairwise-additive non-covalent force fields from monomer electron densities without any empirical input. Energy expressions are based on the Symmetry-Adapted Perturbation Theory (SAPT) decomposition of interaction energies. This ensures a physically motivated force field featuring an electrostatic, exchange-repulsion, dispersion and induction contribution, which contain two types of parameters. First, each contribution depends on several fixed atomic parameters, resulting from a partitioning of the monomer electron density. Second, each of the last three contributions (exchange-repulsion, dispersion and induction) contains exactly one linear fitting parameter. These three so-called interaction parameters in the model, are initially estimated separately using SAPT reference calculations for the S66x8 database of non-covalent dimers. In a second step, the three interaction parameters are further refined simultaneously to reproduce CCSD(T)/CBS interaction energies for the same database. The limited number of parameters that are fitted to

1

dimer interaction energies (only three), avoids ill-conditioned fits that plague conventional parameter optimizations.

For the exchange-repulsion and dispersion component, good results are obtained for all dimers in the S66x8 database using one single value for the associated interaction parameters. The values of those parameters can be considered *universal* and can also be used for dimers not present in the original database used for fitting. For the induction component such an approach is only viable for the dispersion-dominated dimers in the S66x8 database. For other dimers (such as hydrogen-bonded complexes), we show that our methodology remains applicable. However the interaction parameter needs to be determined on a case-specific basis.

As an external validation, the force field predicts interaction energies in good agreement with CCSD(T)/CBS values for dispersion-dominated dimers extracted from an HIV-II protease crystal structure with a bound ligand (indinavir). Furthermore, experimental second virial coefficients of small alkanes and alkenes are well reproduced.

1   Introduction

Classical force-field simulations are an attractive tool to study molecular systems because the phase space can be sampled at a very low computational cost (compared to ab initio methods). This computational efficiency allows one to study phenomena at the nanoscale such as the competitive adsorption of H_2O and CO_2 in Zeolite 13X,[1] the mechanical energy absorption ability of the MIL-53(Al) metal-organic framework[2] or the composition and behavior of disordered protein states.[3] These are only a few examples of succesful applications of force fields. The non-covalent interactions are most often based on experimental data.[4–8] This is of course a limitation, as it is potentially very interesting to study systems for which no experimental data is available. Additionally, in most cases it is impossible to assess a priori the accuracy of such force fields as the final results critically depend on the specific terms included in the force-field expression and on the parametrization procedure. Force

fields derived from ab initio calculations offer in a way the best of two worlds. On the one hand the analytical energy expression can be evaluated efficiently, allowing simulations of large systems on a long time scale. On the other hand, these force fields succeed reasonably well in reproducing ab initio data and can be derived for any molecular system for which reliable ab initio data can be generated. In this work, we will therefore focus on describing non-covalent interactions with only first-principles calculations as input. Constructing such non-covalent force fields exclusively based on ab initio data is a long-standing problem. A *conventional* and straightforward approach is to generate relevant configurations of the system(s) of interest, determine the intermolecular interaction energy and/or forces for these configurations using an ab initio method and then fit (many) parameters of the non-covalent force field by minimizing a cost function with ab initio results as reference data. This procedure has been applied successfully numerous times.[9–11] There are however also some drawbacks of this straightforward matching of ab initio and force-field results, such as the introduction of many parameters and the difficulty in giving a physical interpretation to each of the contributions. In this work we will try to circumvent some of these problems associated with deriving non-covalent force fields. Typically atom types are introduced leading to a large number of parameters that need to be fitted. Even though the number of unknowns in the cost function can be considerably reduced by using mixing rules, the fitting in a high-dimensional parameter space will often be ill-conditioned.[12] A second drawback is that the physical origin of interactions is not properly incorporated. For example, most non-covalent force fields omit the penetration effect, which has been shown to be substantial,[13] when describing electrostatics. This contribution is then erroneously included in the van der Waals energy. Indeed it has been shown that most force fields rely to a large extent on compensation of errors.[14] A third and final deficiency of the conventional approach, is that scanning the PES with an ab initio method is computationally very costly. Therefore one is usually limited to Density Functional Theory (DFT) or second-order Møller-Plesset perturbation theory (MP2), which both have some known limitations (mostly relevant for

3

intermolecular interactions): DFT cannot properly describe electron correlation, while the perturbation term in MP2 is rather large, leading to overestimated dispersion interactions in for example stacked aromatic-aromatic systems.[15,16] It would therefore be very interesting to base non-covalent force fields on coupled-cluster theory with single and double excitations and a perturbation correction for the triples contribution (CCSD(T)) in the complete basis set (CBS) limit, the current gold standard in quantum chemistry.[17]

The issues mentioned above have been partly addressed in earlier work and these efforts are briefly summarized hereafter. The Quantum Mechanically Derived Force Field[18] (QMDFF) is based on a limited number of universal parameters (28 in total for the non-covalent interactions) combined with atomic charges from a partitioned molecular electron density. QMDFF succeeds well in reproducing B3LYP-D3/def2-QZVP reference interaction energies for gas-phase dimers. However, effects such as charge penetration are not included and the parameters are obtained, at least partially, by manual inspection thus rendering the physical interpretation of the different terms unclear. To correctly estimate parameters, it is desirable to match the force field term by term with an intermolecular perturbation theory that provides a decomposition of the interaction energy into physically recognizable components, such as Symmetry-Adapted Perturbation Theory[19] (SAPT). It has been shown that this approach leads to physically-motivated non-covalent force fields that can predict second virial coefficients and gas adsorption in metal-organic frameworks in quantitative agreement with experiment.[20–23] Although the resulting parameters have been shown to be transferable to other chemical environments, the distribution of molecular properties over atomic contributions remains a tedious work. The study of SAPT exchange energies at small intermolecular distances showed that improved models (compared to traditional Lennard-Jones or Born-Mayer functional forms) lead to a more accurate, transferable and robust description of short-range interactions.[24] The SAPT decomposition has also been used for the construction of intermolecular potentials for small organic molecules using the CamCASP suite of programs and this methodology is proven to be succesful for the pyridine system.[25]

A related approach yielded similar success for the water dimer.[26] Recent advances in the AMOEBA force field use SAPT to improve the description of electrostatics[27] and the sum of echange-repulsion and dispersion,[28] aiming to improve the accuracy and transferability of the potential. Another method to ensure physical parameters, is to use the results from a density-based energy decomposition analyis[29] (DEDA). This has been applied in a force-field context only to a few selected systems such as rare-gas dimers[30] or directional dependence of hydrogen bonds.[31] Also the SIBFA[32] (sum of interactions between fragments ab initio computed) force field should be mentioned, as it successfully applies a decomposition of the total interaction energy similar to the one used in this work.

We believe the added value of the current work is mainly that several beneficial properties of the efforts discussed above are combined. In this way we take a step forward to tackle the three deficiencies mentioned earlier. At most three linear parameters are calibrated at a time, thus avoiding ill-conditioned fits. Furthermore all terms are physically inspired (for instance including in a natural way the charge penetration effect using distributed charges) and a limited number of high-level ab initio calculations suffices to confidently determine all parameters. Our results show that the description of induction is the weakest point of the model and can be used with one universal value for the interaction parameter only for dispersion-dominated complexes. For hydrogen-bonded complexes, the interaction parameter needs to be determined on a case-specific basis. We thus gain insight into apparent limitations of the proposed methodology for the description of the SAPT induction for systems that feature hydrogen bonds. We note that many-body effects are very important in the induction energy for such systems and these need to be implicitly captured in our, by construction, pairwise-additive force field. Concerning the exchange and dispersion contributions however, we show that good results are obtained for all dimers using a single universal value for the corresponding interaction parameters. Moreover our model succeeds in the explicit description of the electrostatic penetration effect using distributed charges. The combination of SAPT (to ensure physical values) and CCSD(T)/CBS (to ensure accurate

values) reference data in the fitting process, through the introduction of prior knowledge about the parameters in the cost function, is a novel aspect of the presented work.

To conclude this introduction, we discuss the structure of the remainder of this paper. In section 2, we present the analytical expression for the non-covalent contributions to the energy of our force field. All terms are based on the partitioning of the ab initio monomer density, resulting in a force field containing only a very limited number of free parameters to be fitted, the so called interaction parameters. In section 3, we will discuss how these interaction parameters are determined. In first instance this will be done by fitting to SAPT components of the interaction energy of a database of dimer configurations. In a second step the parameters are refined by fitting to CCSD(T)/CBS interaction energies for the same database of dimers. In section 4, we will perform an external validation of the constructed force field. First of all this will be done by comparing with ab initio interaction energies for another database of dimers, next by comparing with experimental second virial coefficients for small alkanes and alkenes. Finally we will also investigate whether the force field is suitable to predict properties of condensed-phase systems. Section 5 summarizes our main findings.

2   Force-field components

The primary goal is to reproduce ab initio interaction energies of dimers with a force field. We are not concerned with intramolecular interactions, as the force field proposed in this work only describes non-covalent interactions. We therefore start by presenting the analytical expression for the interaction energy *between* two molecules, for which only intermolecular interactions are present.

To construct a force field for a dimer, we require the ab initio molecular ground-state electron density of the separate, isolated molecules where the nuclear positions are taken from the monomer geometries in the dimer. In other words, the nuclear positions are not relaxed

when going from the dimer configuration to the monomer configurations. As DFT provides reliable ground-state electron densities for most molecules, it will be used in this paper to compute the monomer properties that serve as input for the force field. More accurate wavefunction-based methods, SAPT and CCSD(T), are only used in this work to determine the three interaction parameters related to intermolecular interactions. Using DFT for the monomer calculations makes it possible to apply the proposed method to relatively large molecular systems. The ab initio molecular electron densities of both isolated molecules will be represented by $\rho_1(\mathbf{r})$ and $\rho_2(\mathbf{r})$ respectively. The Minimal Basis Iterative Stockholder (MBIS) scheme[33] is used to partition the ab initio molecular electron density of the isolated molecules into atomic contributions. This partitioning scheme allows one to approximate the molecular electron density as a sum of s-type Slater functions:

$$\rho_1(\mathbf{r}) \approx \sum_{A=1}^{N_{\text{atoms}}^1} \sum_{i=1}^{m_A} \frac{N_{Ai}}{8\pi\sigma_{Ai}^3} \exp\left(-\frac{|\mathbf{r} - \mathbf{R}_A|}{\sigma_{Ai}}\right) \tag{1}$$

where the sum over A runs over all N_{atoms}^1 atoms in molecule 1 and i runs over all m_A shells of atom A. The position of atom A is represented by $\mathbf{R}_A$ and the parameters N_{Ai} and σ_{Ai} give the population and *width* of the s-type Slater function of shell i on atom A. The width of the s-type Slater function is the characteristic length over which the density decays appreciably. In other words the width σ_{Ai} is inversely proportional to the rate of decay of the Slater function. The parameters N_{Ai} and σ_{Ai} are determined in such a way that the information loss (expressed as the Kullback-Leibler divergence) from the ab initio reference density to the pro-density is minimized, under constraints on the total charge and atomic populations. This is the most simple and concise model for an electron density that is physically reasonable, inspired by the piece-wise exponential ansatz from statistical models for atomic densities.[34,35] The MBIS charges and widths are the only molecule-specific information used in the proposed force field and are independent of the other monomer present in a dimer. These *atomic* parameters are not fitted to dimer interaction energies,

but obtained directly from the monomer electron density. This is in sharp contrast with the *interaction* parameters (introduced later on in this section), which are the fitting parameters of the model.

For force-field applications, we further simplify the approximation of the molecular electron density by condensing the core Slater functions onto the nuclear charge. The force-field approximation of the valence molecular electron density $\tilde{\rho}_1(\mathbf{r})$ is then given by:

$$\tilde{\rho}_1(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}^1} \frac{N_{A,v}}{8\pi\sigma_{A,v}^3} \exp\left(-\frac{|\mathbf{r} - \mathbf{R}_A|}{\sigma_{A,v}}\right) \tag{2}$$

where $N_{A,v}$ and $\sigma_{A,v}$ are the population and width of the valence Slater function of atom A. We therefore call $N_{A,v}$ and $\sigma_{A,v}$ the valence population and valence width respectively. The effective core charge $q_{A,c}$ of atom A is given by

$$q_{A,c} = Z_A - \sum_{i=1}^{m_A-1} N_{Ai} \tag{3}$$

where Z_A is simply the atomic number or nuclear charge of atom A. To clarify this approach and to get a sense for the magnitude of populations and widths, we present the parameters of a water molecule in Table 1. The oxygen atom has two shells of electrons and thus the MBIS partitioning leads to a model density with two Slater functions. The first Slater function contains 1.66 electrons ($N_{O1} = 1.66$) and decays exponentially with a characteristic length of $\sigma_{O1} = 0.03$Å. The second shell, containing valence electrons, contains 7.20 electrons ($N_{O2} = 7.20$) and decays exponentially with a characteristic length of $\sigma_{O2} = 0.22$Å. As expected, the first Slater function decays a lot faster than the second one. This means that the first Slater function can be approximated accurately by a point charge in force-field simulations. The charge of this first Slater function ($N_{O1} = 1.66$) is subtracted from the nuclear charge ($Z_O = 8$), resulting in an effective core charge $q_{O,c} = 6.34$. The oxygen atom is modeled in the force field as a point charge $q_{O,c} = 6.34$ and a Slater function with $N_{O,v} = 7.20$ and $\sigma_{O,v} = 0.22$Å. For the hydrogen atoms, the situation is less complicated.

Because a hydrogen atom only has one electron shell, the MBIS partitioning leads to a model density with only one Slater function. This same model is used for the force field. We represent a hydrogen atom by a point charge of $q_{H,c} = 1.00$ (equal to the nuclear charge) and a Slater function with $N_{H,v} = 0.57$ and $\sigma_{H,v} = 0.19\text{Å}$.

Table 1: MBIS parameters and resulting force-field parameters for a model of the electron density of a water molecule (DFT with B3LYP[36]/aug-cc-pVTZ[37]).

Element	N_{A1}	σ_{A1} [Å]	N_{A2}	σ_{A2} [Å]	Z_A	$N_{A,v}$	$\sigma_{A,v}$ [Å]	$q_{A,c}$
O	1.66	0.03	7.20	0.22	8	7.20	0.22	6.34
H	0.57	0.19	/	/	1	0.57	0.19	1.00
H	0.57	0.19	/	/	1	0.57	0.19	1.00

The total force-field interaction energy for the molecular dimer is written as a sum of 4 terms, each one with a clear physical interpretation: electrostatics, exchange-repulsion, dispersion and induction. The intermolecular energy can thus be expressed as:

$$E_{\text{inter}} = E_{\text{elst}} + E_{\text{exch-rep}} + E_{\text{disp}} + E_{\text{ind,ct}} \tag{4}$$

This corresponds to the decomposition of the total interaction energy used in SAPT. Every term will be discussed separately in the following subsections. The last three terms, the exchange-repulsion, dispersion and induction component, all feature a separate linear parameter. These three parameters are called *interaction* parameters because they are all used to scale a specific interaction contribution to the total energy. The meaning of the exchange-repulsion interaction parameter will be discussed in Section 2.2, the meaning of the dispersion interacton parameter in Section 2.3 and the meaning of the induction interaction parameter in Section 2.4. Note that the electrostatic contribution, discussed in Section 2.1, is constructed without any interaction parameter.

We thus obtain a force field with two different *types* of parameters. On the one hand there are atomic parameters (such as the valence population N_v and valence width σ_v) obtained from the MBIS partitioning of the monomer density. These parameters are derived from monomer properties and not fitted to dimer interaction energies. On the other hand, there

9

are three interaction parameters which are fitted to reference data for dimer interaction energies. These three interaction parameters are the only fitting parameters present in the model.

2.1 Electrostatics

The frozen-density electrostatic interaction energy between two molecules is given by the Coulomb integral

$$
E_{\text{elst}} = \sum_{A=1}^{N_{\text{atoms}}^1} \sum_{B=1}^{N_{\text{atoms}}^2} \frac{q_{A,c} q_{B,c}}{|\mathbf{R}_A - \mathbf{R}_B|} - \sum_{A=1}^{N_{\text{atoms}}^1} q_{A,c} \int d\mathbf{r} \frac{\tilde{\rho}_2(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_A|}
$$
$$
- \sum_{B=1}^{N_{\text{atoms}}^2} q_{B,c} \int d\mathbf{r} \frac{\tilde{\rho}_1(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_B|} + \int d\mathbf{r}_1 \int d\mathbf{r}_2 \frac{\tilde{\rho}_1(\mathbf{r}_1)\tilde{\rho}_2(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \tag{5}
$$

where N_{atoms}^1 is the number of atoms in molecule 1 and N_{atoms}^2 is the number of atoms in molecule 2. The first term describes repulsion between effective core charges, the second and third term describe attraction between effective core charges and valences electrons and finally the fourth term describes repulsion between valence electrons. The Coulomb integral can be evaluated analytically for our force-field model of the density from Eq. (2) to give:

$$
E_{\text{elst}} = \sum_{A=1}^{N_{\text{atoms}}^1} \sum_{B=1}^{N_{\text{atoms}}^2} E_{\text{elst}}^{AB} \tag{6}
$$

$$
E_{\text{elst}}^{AB} = \frac{q_{A,c} q_{B,c}}{R_{AB}} - \frac{q_{A,c} N_{B,v}}{R_{AB}} \left[1 - g(\sigma_{B,v}, R_{AB}) \right] - \frac{N_{A,v} q_{B,c}}{R_{AB}} \left[1 - g(\sigma_{A,v}, R_{AB}) \right]
$$
$$
+ \frac{N_{A,v} N_{B,v}}{R_{AB}} \left[1 - f(\sigma_{A,v}, \sigma_{B,v}, R_{AB}) - f(\sigma_{B,v}, \sigma_{A,v}, R_{AB}) \right] \tag{7}
$$

$$
g(\sigma, r) = \left(1 + \frac{r}{2\sigma} \right) \exp\left(-\frac{r}{\sigma} \right) \tag{8}
$$

$$
f(\sigma_i, \sigma_j, r) = \frac{\sigma_i^4}{\left(\sigma_i^2 - \sigma_j^2\right)^2} \left[1 + \frac{r}{2\sigma_i} - \frac{2\sigma_j^2}{\sigma_i^2 - \sigma_j^2} \right] \exp\left(-\frac{r}{\sigma_i} \right) \tag{9}
$$

The expression for the Coulomb interaction of s-type Slater densities has been presented before.[31] For convenience we provide a derivation of the above formula in the Supporting

Information. By using the net atomic charges of atoms A and B, $q_A = q_{A,c} - N_{A,v}$ and $q_B = q_{B,c} - N_{B,v}$, we can rewrite the electrostatic interaction energy between those atoms as:

$$E_{\text{elst}}^{AB} = \frac{q_A q_B}{R_{AB}} + E_{\text{penetration}}^{AB} \tag{10}$$

The first term is the well-known electrostatic interaction between point charges. The second term decays exponentially as the atoms are separated and describes the penetration effect of the electrostatic interaction energy:

$$
\begin{aligned}
E_{\text{penetration}}^{AB} = {} & \frac{q_{A,c} N_{B,v}}{R_{AB}} g\left(\sigma_{B,v}, R_{AB}\right) + \frac{N_{A,v} q_{B,c}}{R_{AB}} g\left(\sigma_{A,v}, R_{AB}\right) \\
& - \frac{N_{A,v} N_{B,v}}{R_{AB}} \left[f\left(\sigma_{A,v}, \sigma_{B,v}, R_{AB}\right) + f\left(\sigma_{B,v}, \sigma_{A,v}, R_{AB}\right)\right]
\end{aligned}
\tag{11}
$$

We remark that existing implementations to evaluate long-range electrostatics in periodic systems (such as the Ewald summation,[38,39] PPPM[40] or FMM[41]) do not need to be modified to calculate the first term in Equation 10. The second term, containing the penetration correction to the electrostatic energy, clearly decays exponentially and can be evaluated similar to other short-ranged non-covalent terms. By tabulating the values of this second term, our force field remains compatible with existing force-field codes.

The importance of the penetration effect has been discussed before.[42] Distributed charges have already been employed successfully in force fields. Most often Gaussian functions are used,[43] some force fields feature Slater densities[30,31] and one force field even uses contracted Gaussian functions to approximate a Slater function.[44] Distributed charges offer a natural way to describe the penetration effect and can give accurate results on condition that the model density is a good approximation to the ab initio density. It has been shown that this is the case for model densities resulting from MBIS partitioning.[33]

2.2 Exchange-Repulsion

The wavefunction of a dimer has to be antisymmetric when two electrons are exchanged. This antisymmetrized wavefunction leads to a higher energy expectation value than if a simple product of the two monomer wavefunctions would be used. The resulting repulsive energy is only present when the electron densities of the two molecules in the dimer overlap. This repulsive energy dominates the so called exchange-repulsion energy.[45] The energy due to the exchange-repulsion effect has no classical analogue and cannot be expressed simply as a functional of the isolated molecular densities. We therefore base our model on the observed proportionality between exchange-repulsion energy and overlap of electron densities:[25,46–48]

$$E_{\text{exch-rep}} \approx U_{\text{exch-rep}} S_{12} \tag{12}$$

$$= U_{\text{exch-rep}} \int d\mathbf{r} \tilde{\rho}_1(\mathbf{r}) \tilde{\rho}_2(\mathbf{r}) \tag{13}$$

If we plug in our force-field model of the electron densities from Eq. (2), we can rewrite this as a sum of atomic pairwise interactions:

$$E_{\text{exch-rep}} = \sum_{A=1}^{N_{\text{atoms}}^1} \sum_{B=1}^{N_{\text{atoms}}^2} E_{\text{exch-rep}}^{AB} \tag{14}$$

$$E_{\text{exch-rep}}^{AB} = U_{\text{exch-rep}} S^{AB} \tag{15}$$

$$S^{AB} = \frac{N_{A,v} N_{B,v}}{8\pi R_{AB}} \left[h(\sigma_{A,v}, \sigma_{B,v}, R_{AB}) + h(\sigma_{B,v}, \sigma_{A,v}, R_{AB}) \right] \tag{16}$$

$$h(\sigma_i, \sigma_j, r) = \left[\frac{4\sigma_i^2 \sigma_j^2}{\left(\sigma_j^2 - \sigma_i^2\right)^3} + \frac{\sigma_i}{\left(\sigma_j^2 - \sigma_i^2\right)^2} r \right] \exp\left(-\frac{r}{\sigma_i}\right) \tag{17}$$

A derivation of this expression is provided in the Supporting Information.

$U_{\text{exch-rep}}$ is the first *interaction* parameter in our model. We use the term interaction parameter for this proportionality constant to avoid confusion with atomic or atom-pair parameters. Here, the *total* exchange-repulsion interaction energy is proportional to $U_{\text{exch-rep}}$, which justifies the name interaction parameter. The value of the interaction parameter

$U_{\text{exch-rep}}$ will be determined in section 3 by a simple linear regression of the overlap of the force-field densities with exchange-repulsion energies from SAPT calculations for a large database of dimer configurations. It will be shown that the interaction parameter $U_{\text{exch-rep}}$ is, to some extent, system independent. Using a universal value for all dimers in the database leads to accurate results. Notice the important difference with some earlier works[24,49] where the prefactor of the overlap is atom-type specific (i.e. $E^{AB}_{\text{exch-rep}} = U^{AB}_{\text{exch-rep}} S^{AB}$). Such an approach of course involves more paramaters to be fitted to intermolecular reference data than the approach presented here.

2.3 Dispersion

The correlated movement of electrons in interacting molecules causes an attractive force. This effect is called dispersion. The dispersion energy between two molecules, which are far apart, can be written in second-order perturbation theory using a power expansion in terms of the intermolecular distance: $E_{\text{disp}} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} - \dots$ From this expansion we only retain the two leading terms. For our force field, we write the energy as a sum of atomic pairwise contributions to the dispersion interaction. Finally we use a damping function to control the behavior for small distances. Our force-field model of the dispersion energy then has the following form:

$$E_{\text{disp}} = \sum_{A=1}^{N^1_{\text{atoms}}} \sum_{B=1}^{N^2_{\text{atoms}}} E^{AB}_{\text{disp}} \tag{18}$$

$$E^{AB}_{\text{disp}} = -f_6\left(x_{AB}\right) \frac{C^6_{AB}}{R^6_{AB}} - f_8\left(x_{AB}\right) \frac{U_{s8} C^8_{AB}}{R^8_{AB}} \tag{19}$$

$$f_n\left(x_{AB}\right) = 1 - \left(\sum_{k=0}^{n} \frac{x^k_{AB}}{k!}\right) \exp\left(-x_{AB}\right) \tag{20}$$

with U_{s8} another interaction parameter that is used to linearly scale all interatomic C^8_{AB} coefficients and thus the total C^8 dispersion interaction energy. The dispersion energy is damped at short interatomic distances using the Tang-Toennies damping function[50] with as

13

argument the interatomic distance divided by the average of the Slater widths of the atoms involved: $x_{AB} = \frac{R_{AB}}{(\sigma_{A,v} + \sigma_{B,v})/2}$

To determine the interatomic C_{AB}^6 coefficients, we use the Tkatchenko-Scheffler method.[51] In this method, the C_{AB}^6 coefficients of atoms in a molecule are computed using free-atom reference values that are modified based on the partitioning of the molecular electron density into atomic contributions. In the original paper, the Hirshfeld partitioning scheme[52] was used. To maintain consistency with the other force-field terms, we will use the MBIS partitioning scheme. It has been shown that the combination of the Tkatchenko-Scheffler method and MBIS partitioning leads to reliable C_{AB}^6 coefficients.[33] For more information about the computation of C_{AB}^6 coefficients we refer to the Supporting Information.

To compute the C_{AB}^8 coefficients, we use a recursion relation proposed by Starkschall:[53]

$$C_{AB}^8 = \frac{3}{2} C_{AB}^6 \sqrt{\frac{\langle r_A^4 \rangle}{\langle r_A^2 \rangle} + \frac{\langle r_B^4 \rangle}{\langle r_B^2 \rangle}} \tag{21}$$

The expectation values of powers of r_A and r_B are evaluated using the free-atom densities.[54] In the energy expression Eq. (19) the C_{AB}^8 coefficients are uniformly scaled with the interaction parameter U_{s8}, to account for the higher order contributions to the dispersion energy and possible deficiencies in the calculation of the C_{AB}^8 coefficients and damping function.

2.4 Induction

The electrostatic interaction given by Eq. (5) assumes a frozen-density approach, where the molecular densities are kept frozen when the two molecules approach each other. In reality, the molecular densities will deform as both molecules approach each other, an effect called polarization. The resulting stabilization is called the induction energy. The construction of polarizable force fields that succeed in quantitatively reproducing this induction energy is not trivial and is an issue that has received a lot of attention.[55,56] The most common approach is

to model each atom as an inducible dipole, either using a Drude oscillator (charge-on-spring) model[15,57–62] or as a Thole- or Gausian-damped point dipole.[60,63–65] Each inducible dipole interacts with the atomic point charges and all other dipoles. For a given geometry, the energy of the inducible dipoles is minimized, in analogy to the Born-Oppenheimer approximation. In such models, at least two parameters per atom are needed: an atomic dipole polarizability and a damping parameter. We have tested models with inducible point dipoles, where realistic atomic polarizabilities are obtained with the Tkatchenko-Scheffler method.[51] However, the resulting induction energies were too small (in absolute value) when compared to SAPT reference values, especially at short intermolecular distances. The same conclusion is obtained for the induction term of the AMOEBA[66] force field. The results of those two induction models are presented in Section 5 of the Supporting Information. This investigation reveals that another phenomenon plays a dominant role in the SAPT induction component. As an alternative, some of the authors recently proposed a polarizable force field where all parameters can be evaluated as expectation values from a quantum-mechanical wavefunction.[67,68] This method is however not readily applicable to classical molecular dynamics simulations, as it would require evaluating the quantum-mechanical wavefunction at every timestep.

We therefore resort to an approximation with a simpler analytical expression: the induction energy is modeled as a pairwise-additive term with one interaction parameter $U_{\mathrm{ind,ct}}$.

$$E_{\mathrm{ind,ct}} = \sum_{A=1}^{N^1_{\mathrm{atoms}}} \sum_{B=1}^{N^2_{\mathrm{atoms}}} E^{AB}_{\mathrm{ind,ct}} \tag{22}$$

$$E^{AB}_{\mathrm{ind,ct}} = -U_{\mathrm{ind,ct}} S^{AB} \tag{23}$$

The expression for the overlap of 1s-type Slater densities S^{AB} is given in Eq. (16). The proposed proportionality of the intermolecular induction energy is inspired by the exponentially decaying nature of charge transfer.[23,32,69–73] More specifically, the proportionality between exchange-repulsion and the charge-transfer interaction has been demonstrated in the context

of SAPT.[74] We expect that this assumption will work fairly well for dispersion-dominated complexes, where the induction energy is dominated by contributions from charge-transfer effects because classical polarization plays a minor role in such complexes. The induction energy is in general attractive, so we expect a positive value for $U_{\text{ind,ct}}$. The pairwise-additive expression for the induction energy is a serious approximation and its applicability will be assessed in section 3. For complexes where the proposed method is not viable we will discuss possible causes of the observed deviations and extensions of the model.

In summary, our force field for non-covalent interactions contains three interaction parameters: $U_{\text{exch-rep}}$ gives the proportionality between the exchange-repulsion energy and monomer density overlap, U_{s8} is a uniform scaling factor for the C^8 coefficients used to compute the dispersion energy and finally $U_{\text{ind,ct}}$ gives the proportionality between the induction energy and monomer density overlap. Note that the total interaction energy is a linear function of these three interaction parameters.

3   Parameter calibration

In the previous section, analytical expressions have been proposed for the four terms composing the total intermolecular interaction energy: $E_{\text{inter}} = E_{\text{elst}} + E_{\text{exch-rep}} + E_{\text{disp}} + E_{\text{ind,ct}}$, representing electrostatic, exchange-repulsion, induction and dispersion energy, containing three interaction parameters in total. These interaction parameters together with the MBIS partitioned density of the molecules at hand form a complete force field to describe the interaction between those two molecules. High-level ab initio interaction energies obtained for a large database of dimers will be used to determine the three interaction parameters. At the same time, the degree of reproduction of the high-level ab initio intermolecular interaction energies by the force field serves as a first validation that our force-field model is appropriate and that the interaction parameters for the exchange-repulsion and dispersion component are to a large extent system independent and their values can be considered universal.

We used the S66x8[75] database as a training set to determine the values of the interaction parameters. This database covers the most common types of non-covalent interactions in biomolecules, keeping a balanced representation of electrostatic-dominated, dispersion-dominated and mixed-influence complexes. Based on a SAPT decomposition of the interaction energies, all 66 complexes in the database are classified either as hydrogen-bonded (23 complexes), dispersion-dominated (23 complexes) or "others" (20 complexes). In order to construct a total of 66 dimers, various combinations of 14 monomers are considered. The 14 selected monomers are acetic acid, acetamide, benzene, cyclopentane, ethene, ethyne, neopentane, n-pentane, methylamine, methanol, N-methylacetamide, pyridine, uracil and water. All these monomers are composed exclusively of hydrogen, carbon, nitrogen and oxygen atoms. All complexes are relatively small (6 to 34 atoms), such that CCSD(T)/CBS and SAPT calculations are feasible for the entire dataset. The S66x8 database does not only contain equilibrium geometries of the 66 dimers but also scans the dissociation curve with displacements relative to the equilibrium distance of 0.90, 0.95, 1.00, 1.05, 1.10, 1.25, 1.50 and 2.00, resulting in $528 = 66 \times 8$ data points. We note that a relative displacement of 1.00 corresponds approximately with the equilibrium separation distance. Finally we note that the monomer geometries remain fixed along the dissociation curve, which means that there are no changes in the covalent energy along the dissociation curve. This allows us to assess the accuracy of the non-covalent force field.

3.1 SAPT energies for the S66x8 database

One of the ultimate goals of the force field is the capability to accurately reproduce dissociation curves obtained using the CCSD(T)/CBS method. We proposed analytical expressions for four contributions composing the total intermolecular interaction energy: electrostatic, exchange, dispersion and induction. A decomposition of the CCSD(T)/CBS energy into these four categories is not straightforward and not unambigious. In the context of force-field development it is however instructive to also study methods that provide more chemical

insight into the interactions, by decomposing the energy into meaningful contributions. It should be stressed that a decomposition of the total interaction energy is not unique and somewhat arbitrary. Indeed, several decomposition schemes proposed in literature, such as Kitaura-Morokuma[76] (based on the Hartree-Fock approximation), constrained space orbital variation[77] or density-based energy decomposition analysis[29] will all lead to different results. We will use SAPT[19] as a benchmark, as it directly computes electrostatics, exchange-repulsion, dispersion and induction contributions separately. SAPT is an ab initio method and gives total interaction energies in reasonable agreement with CCSD(T)/CBS values.[78] Unfortunately, due to its double perturbative nature (the intermolecular interaction and the residual monomer correlation interactions), it is not always clear how to classify higher-order terms, as they show a mixture of exchange, dispersion and induction. It is also clear that the perturbation expansion is not converged in third order, the highest order currently available in SAPT implementations. This is why we start by studying different SAPT approximations and compare them to CCSD(T)/CBS values. Our results are in line with conclusions from a similar analysis by Li et al.[78]

We used PSI4[79] for all SAPT and coupled-cluster calculations. For SAPT, the aug-cc-pVTZ basis set[37] was used with density fitting[80,81] to avoid evaluation of four-center integrals. Core orbitals were frozen and MP2 natural orbitals[82] were used to speed up the evaluation of second-order T_2 amplitudes and the triples contribution to the dispersion energy. All exchange-type terms are scaled with the factor $\frac{E_{\text{exch}}^{(10)}}{E_{\text{exch}}^{(10)}(S^2)}$, as recommended in literature.[83] The CCSD(T)/CBS energies were computed as the sum of the Hartree-Fock energy in the aug-cc-pVQZ basis set,[37] an MP2 correction extrapolated to the complete basis set limit using Helgaker's formula[84] applied to aug-cc-pVTZ and aug-cc-pVQZ basis sets,[37] and finally a CCSD(T) correction computed with the aug-cc-pVDZ basis set.[37] This procedure corresponds to the one used originally for the S66 dataset.[75]

In Table 2 we compare several SAPT levels-of-theory with CCSD(T)/CBS values. As reported before for the S66 dataset[83] (only for the equilibrium separation distance), SAPT2+(3)

performs better than SAPT2+3. This is an indication that the SAPT expansion is not converged, because an exact treatment of third-order perturbations (SAPT2+3) performs worse than an approximate treatment (SAPT2+(3)). Convergence properties of SAPT have been discussed extensively, see Refs.[19,85,86] and references therein. We note that although SAPT2+(3)δMP2/aug-cc-pVTZ[37] is considered as the optimal SAPT level-of-theory,[83] it cannot be used for the present application because it is no longer possible to clearly separate induction from dispersion contributions. We therefore use SAPT2+(3) in the remainder of this work, with the caveat that it may show some deficiencies, especially at short intermolecular distances.

Table 2: RMSD of CCSD(T)/CBS and SAPT interaction energies in $kJ\,mol^{-1}$ for the S66x8 dataset.

Relative displacement	SAPT0	SAPT2+(3)	SAPT2+3
0.90	10.4	2.6	4.6
0.95	8.2	1.8	3.2
1.00	6.6	1.4	2.4
1.05	5.4	1.1	1.8
1.10	4.4	0.9	1.4
1.25	2.7	0.6	0.8
1.50	1.5	0.4	0.4
2.00	0.7	0.2	0.2

3.2 Parameter calibration by comparing with SAPT energy components

We will now investigate each term of the force field separately and compare it with its SAPT counterpart, in order to determine the introduced interaction parameters. All the interaction parameters are fitted to the 23 dispersion-dominated complexes present in the S66x8 dataset, except if noted otherwise. Although this selection is only really needed for the induction term, as will be explained below, other parameters were treated consistently. The electron densities of the isolated monomers in the S66x8 database were calculated using DFT with a B3LYP[36] functional and aug-cc-pVTZ basis set[37] using Gaussian.[87] The molecular electron

density was then partitioned with the MBIS scheme using HORTON,[88] with the built-in ultrafine grid.

Electrostatics

No interaction parameters are included in the electrostatic term E_{elst} (Eq.(5)) of the force field. It simply consists of the Coulomb interaction between the total charge densities of the two monomers determined by the effective core point charges and the Slater-type charge distributions representing the valence electrons. Figure 1 shows a correlation diagram of the force-field electrostatic energies and the reference SAPT values for the 528 structures of the S66x8 dataset. As expected, most of the electrostatic interaction energies are negative (510 out of 528 structures). For practical reasons, we only display those cases where the SAPT electrostatic energy is below $-1\,\text{kJ}\,\text{mol}^{-1}$. It is clear that the force-field electrostatic model is able to capture the important penetration effect accurately. In figure 2 we provide a similar plot to investigate the performance of RESP[89] fitted point charges (obtained using antechamber[90]) from a HF/6-31G* density, a method often used in force fields. It is clear that the agreement is unsatisfactory. We note that point charges obtained with other methods (partitioning schemes such as Hirshfeld,[52] Iterative-Hirshfeld[91] or ESP fits with other levels-of-theory) result in similar deviations because of their intrinsic incapability to model the penetration effect.[33] Also the electrostatic part of the AMOEBA force field,[64,66,92,93] calculated using point multipoles up to quadrupoles ($\ell = 2$), is significantly improved when including an empirical penetration model.[27]

Exchange-Repulsion

The exchange-repulsion term of our force field contains one interaction parameter, $U_{\text{exch-rep}}$. We determine this parameter by minimizing a cost function χ^2 that expresses the mean-
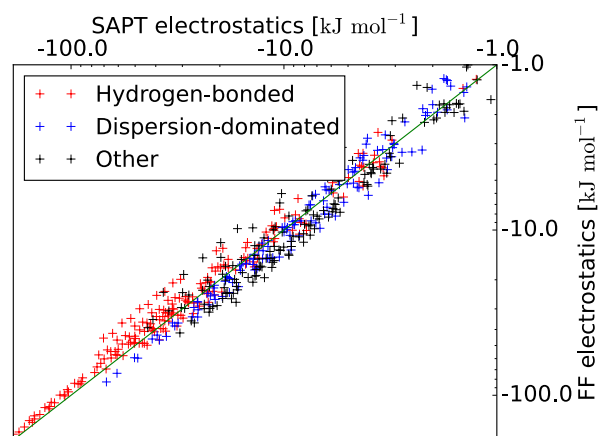
Figure 1: Correlation plot of SAPT and force-field electrostatic energies for the S66x8 database. Note the use of a logarithmic scale.
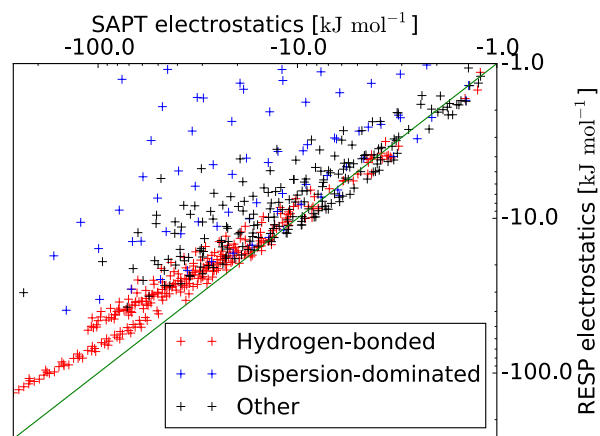

Figure 2: Correlation plot of SAPT and RESP fitted point charges (HF/6-31G*) electrostatic energies for the S66x8 database. Note the use of a logarithmic scale.

square-deviation between force-field and SAPT exchange-repulsion energy:

$$\chi^2 = \frac{1}{N_d} \sum_{n=1}^{N_d} \left[E_{\text{exch-rep}}^{\text{FF}}(n) - E_{\text{exch-rep}}^{\text{SAPT}}(n) \right]^2 \tag{24}$$

The summation over n runs over all N_d data points that we consider. Because our force-field exchange-repulsion energy $E_{\text{exch-rep}}^{\text{FF}}(n) = U_{\text{exch-rep}} S(n)$ is a linear function of $U_{\text{exch-rep}}$ (with $S(n)$ the overlap of our molecular force-field densities), minimizing the cost function χ^2 effectively amounts to a linear regression with solution:

$$U_{\text{exch-rep}} = \frac{\sum_n S(n) E_{\text{exch-rep}}^{\text{SAPT}}(n)}{\sum_n S(n) S(n)} \tag{25}$$

We first determine the value of $U_{\text{exch-rep}}$ for every dimer separately and Figure 3 justifies our approximation that the value of the interaction parameter $U_{\text{exch-rep}}$ is indeed universal, as it takes on about the same value for all dimers independent of their classification (hydrogen-bonded, dispersion-dominated or other). A notable outlier is the neopentane dimer (index 36). By fitting to all dispersion-dominated complexes from the S66x8 database, we find
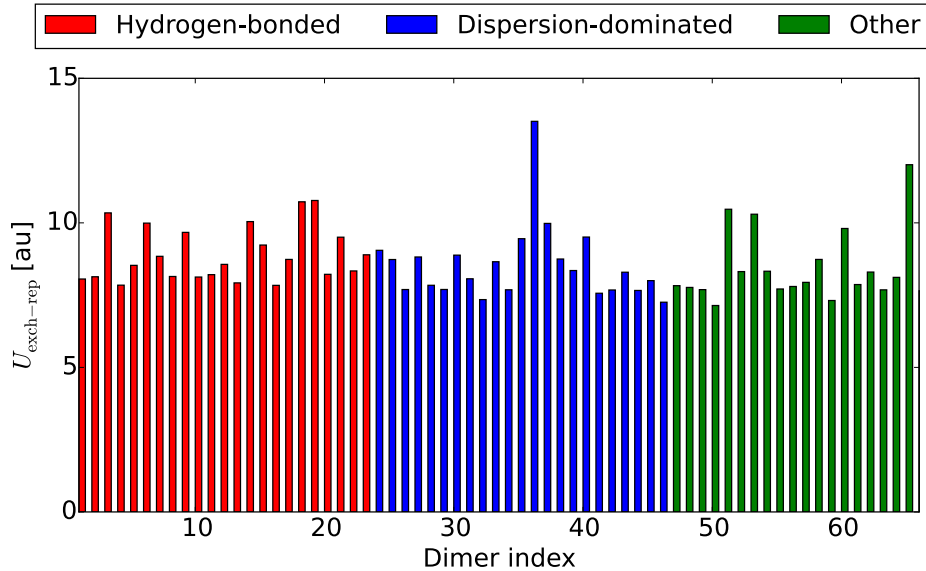

Figure 3: Values of $U_{\text{exch-rep}}$ specific for each dimer.

$U_{\text{exch}} = 8.13$ in atomic units. In Figure 4 we show that our model is capable of reproducing

SAPT exchange-repulsion energies, justifying the assumed functional expression that is proportional to the overlap of monomer electron densities. It turns out that for hydrogen-bonded complexes the model slightly underestimates the exchange-repulsion energy. Fortunately, this error is (at least partly) compensated by an underestimation of the electrostatic attraction energy for the same complexes (see figure 1). Our model seems to be better suited to reproduce the sum of exchange-repulsion and electrostatic energies, rather than both individual SAPT components. This can probably be related to the somewhat arbitrary classification of some terms, one of the limitations of the energy-decomposition methods mentioned earlier.
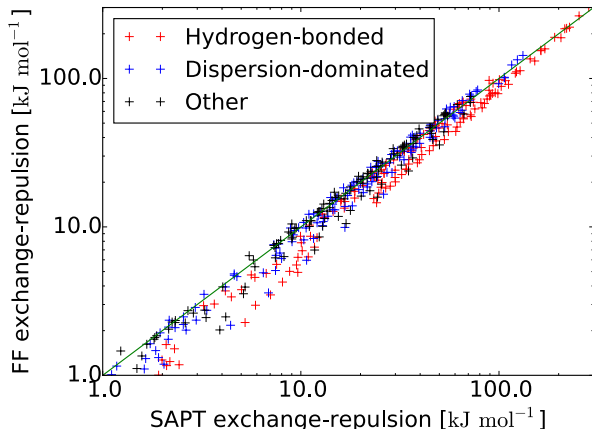
Figure 4: Correlation plot of SAPT and force-field exchange-repulsion energies for the S66x8 database. Note the use of a logarithmic scale.

Dispersion

Our expression for the dispersion energy contains one interaction parameter U_{s8} to scale the contribution of the term decaying as r^{-8}. To avoid interference from possible deficiencies of our Tang-Toennies damping function, we choose to fit U_{s8} to relatively large intermolecular separations (relative displacements 1.50, 1.60, 1.70, 1.80, 1.90, 2.00, 2.25 and 2.50, some of which are not included in the original S66x8 database). The fitting is first performed for each dimer separately and the resulting values of U_{s8} are reported in Figure 5. We note a rather large spread on the values of the interaction parameter. Only within the class of

23

dispersion-dominated complexes it is reasonable to attribute a universal value to U_{s8} and we find $U_{s8} = 0.57$. If we compare the SAPT dispersion energies with our model energies (for
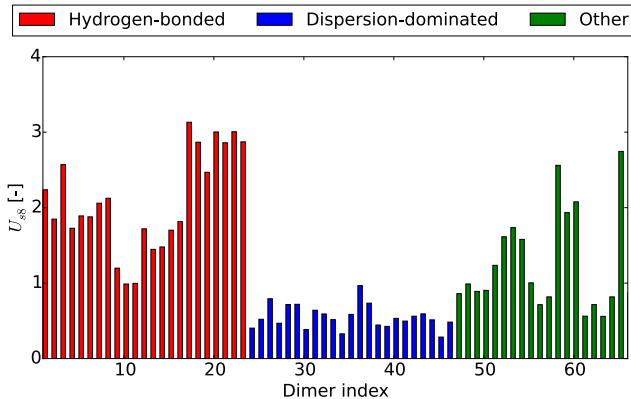

Figure 5: Values of s_8 specific for each dimer.

$U_{s8} = 0.57$) we find a good agreement for the dispersion-dominated complexes in the S66x8 dataset as shown in figure 6. Surprisingly the model also performs acceptably well for the hydrogen-bonded and other complexes. This is due to a fortuitous error cancellation between the underestimated value for U_{s8} and the overestimated damping for those complexes at short intermolecular separation. The value $U_{s8} = 0.57$ can thus also be considered to be universally applicable.

Induction

Our current induction model has a simple pairwise-additive functional form, again involving only one interaction parameter $U_{\mathrm{ind,ct}}$ that provides a linear relationship between the induction energy and monomer density properties. Figure 7 reveals that fitting $U_{\mathrm{ind,ct}}$ for every dimer separately results in a large spread of values. It is notable, but perhaps not surprising, that $U_{\mathrm{ind,ct}}$ is systematically higher for hydrogen-bonded than for dispersion-dominated complexes. It also becomes clear that only within the group of dispersion-dominated complexes it makes sense to define a universal value for $U_{\mathrm{ind,ct}}$. Nevertheless we tested the approach with a universal value for $U_{\mathrm{ind,ct}}$ for the entire dataset on the one hand and the different subsets on the other hand, and report the results in Figure 8. Figure 8 (c) confirms that by
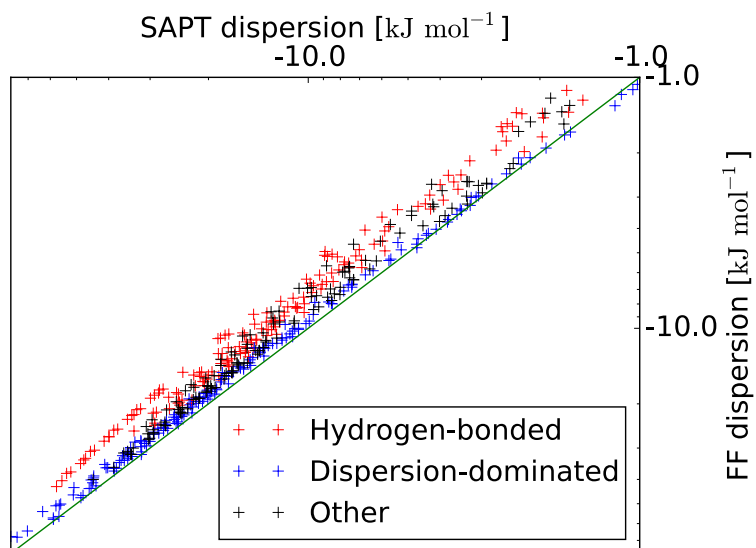
Figure 6: Correlation plot of SAPT and force-field dispersion energies for the S66x8 database. Note the use of a logarithmic scale.
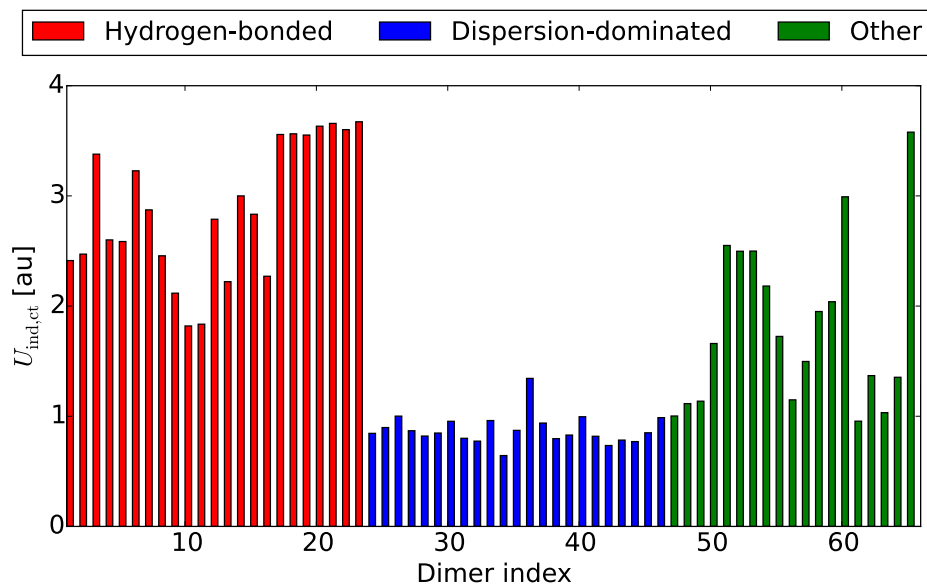

Figure 7: Values of $U_{\mathrm{ind,ct}}$ specific for each dimer.

setting $U_{\text{ind,ct}} = 0.87$ we obtain a quantitatively accurate induction model for the dispersion-dominated complexes. From Figures 8 (b) and (d) we learn that a similar approach is not viable for hydrogen-bonded and *other* complexes, although at the same time it seems that the correlation between SAPT induction and overlap is rather good for each separate complex (i. e. the curves connecting data points belonging to the same complex are relatively straight.)
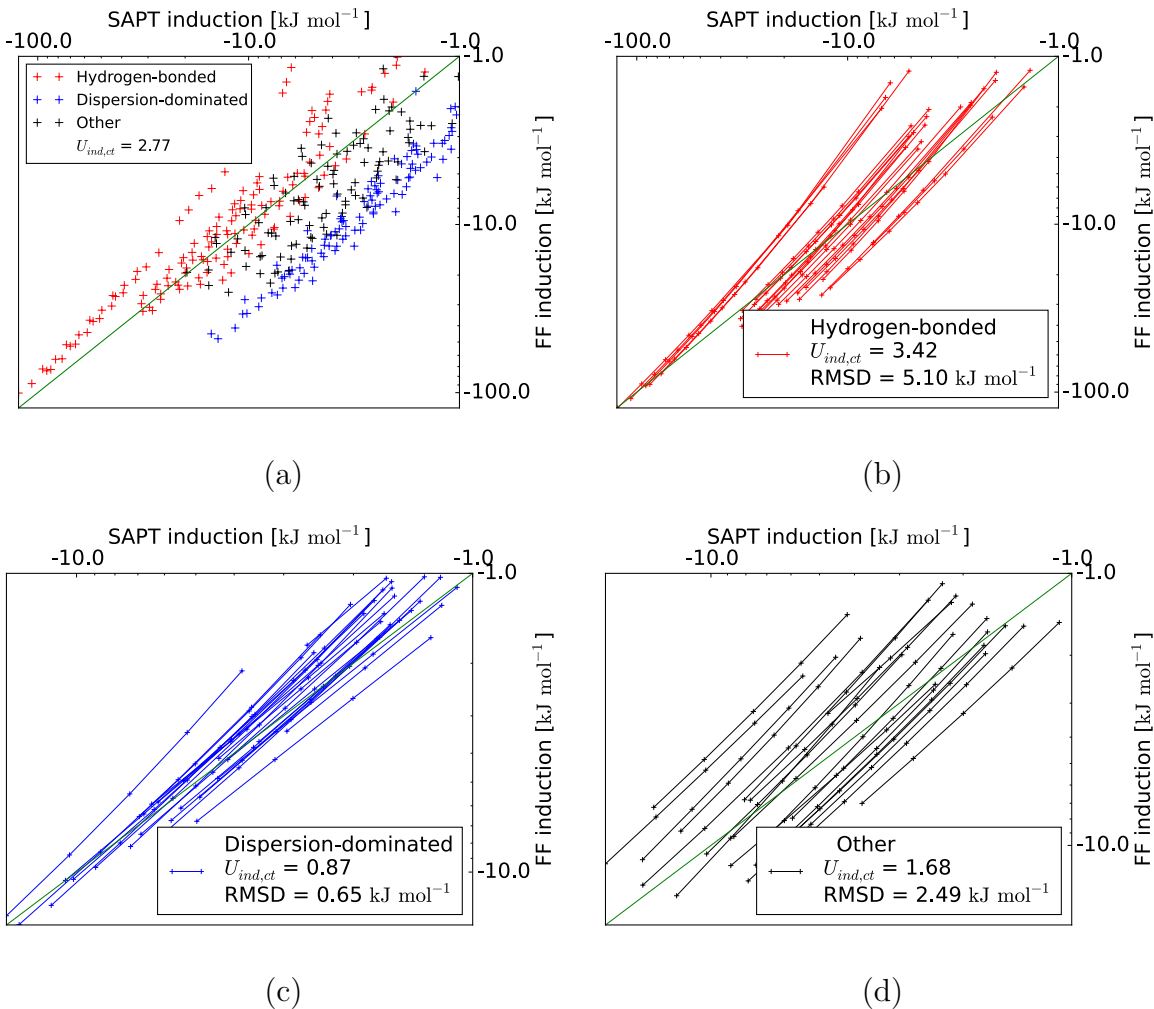

Figure 8: Correlation plots of SAPT and force-field induction energies. (a): complete S66x8 database with $U_{\text{ind,ct}} = 2.77$, (b): only hydrogen-bonded complexes in cost function ($U_{\text{ind,ct}} = 3.42$), (c): only dispersion-dominated complexes in cost function ($U_{\text{ind,ct}} = 0.87$), (d): only "other" complexes in cost function ($U_{\text{ind,ct}} = 1.68$). Note the use of a logarithmic scale. Data points on the dissociation curve of each dimer are connected with a line.

We investigate this further on a complex-specific basis. In Figure 9 (a) we compare our

model induction with the SAPT counterpart for the water dimer (index 01 of the S66x8 dataset, containing one hydrogen bond) where the parameter $U_{\mathrm{ind,ct}} = 2.42$ is obtained by fitting only to the 8 points on the dissociation curve of the water dimer. To make sure that the model still holds for other configurations on the water-dimer potential energy surface, we validate it for the geometries of the water dimer in the S66a8 dataset.[94] This dataset contains angular-displaced nonequilibrium geometries for the complexes present in the S66x8 dataset. Similar plots are shown in Figure 9 (b), (c) and (d) for the water-peptide dimer (index 04, containing one hydrogen bond), the acetamide dimer (index 21, containing two hydrogen bonds) and the benzene-methanol dimer (index 55, containing an OH-π bond) respectively. In general the absolute value of the SAPT induction energy is underestimated at long intermolecular distances (small absolute values of SAPT induction energies) while the absolute value is overestimated at short intermolecular distances (large absolute values of SAPT induction energies). This is because at long range the SAPT induction energy decays as a power of $\frac{1}{r}$, while our model shows a much faster exponential decay. This could be solved by the inclusion of for instance inducible dipoles which are not present in our pairwise-additive force field. In general the MEDFF induction energies for the complexes shown in Figure 9 are in reasonable agreement with SAPT induction energies, and especially the transferability to angular-displaced geometries (very challenging for directional hydrogen bonds) is encouraging. We thus conclude that our induction model cannot describe all dimers in the S66x8 dataset with just one universal value for $U_{\mathrm{ind,ct}}$; only for the subset of dispersion-dominated dimers this is a viable approach. Other dimers such as hydrogen-bonded complexes can still be described relatively well using our induction model, provided one uses a value for $U_{\mathrm{ind,ct}}$ specifically fitted to the dimer at hand. Plots similar to Figure 9 are included in the Supporting information for all complexes of the S66x8 dataset.
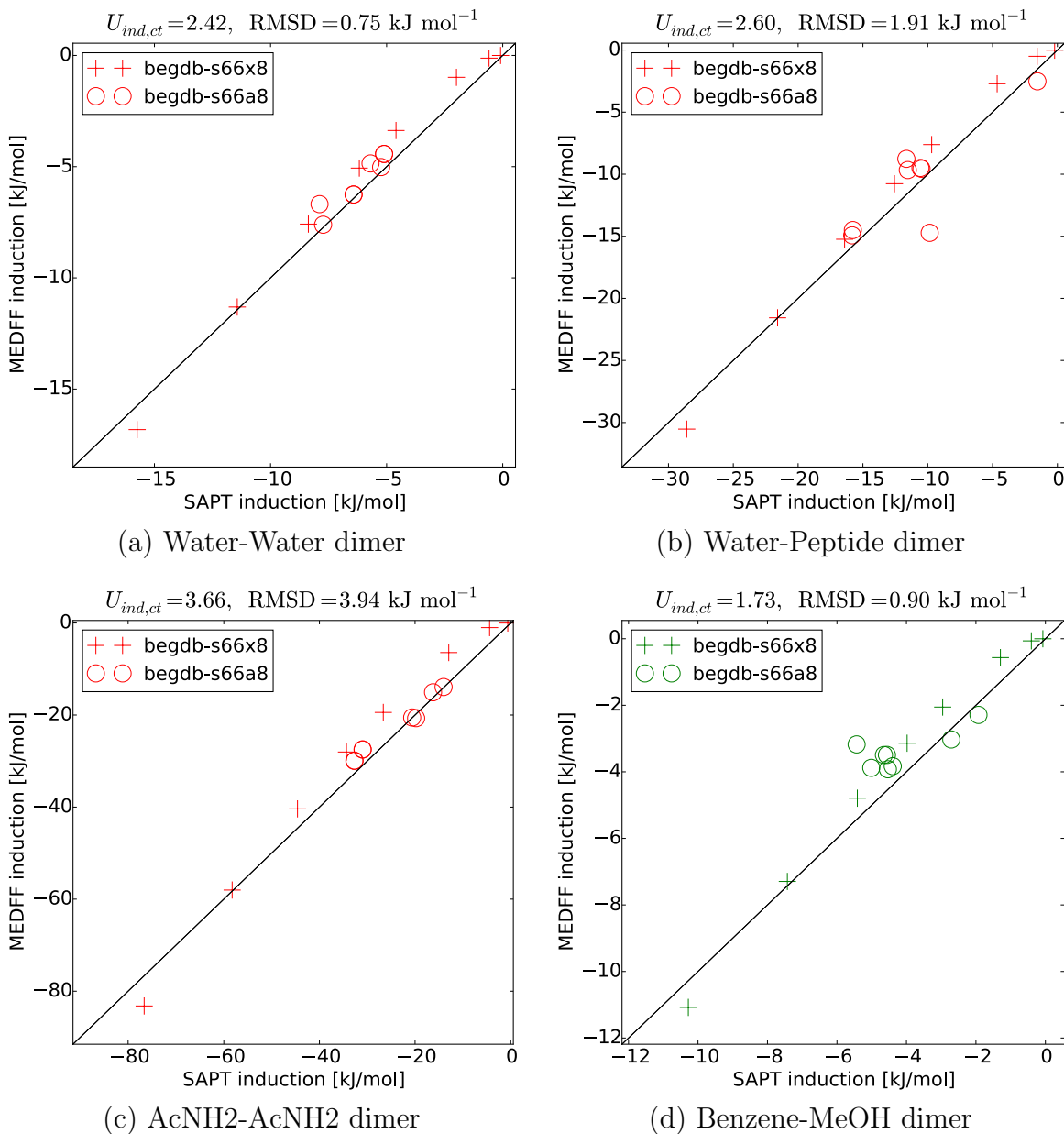
(a) Water-Water dimer

(b) Water-Peptide dimer

(c) AcNH2-AcNH2 dimer

(d) Benzene-MeOH dimer

Figure 9: Correlation plots of SAPT and force-field induction energies. The parameter $U_{\mathrm{ind,ct}}$ has a different value for each complex, obtained by fitting to the 8 geometries of that complex in the S66x8 dataset. The RMSD is computed for all geometries of that complex in S66x8 and S66a8.

3.3   Refining parameters by calibrating to CCSD(T)/CBS energies

By comparison of SAPT2+(3) interaction energies with CCSD(T)/CBS reference values, it became clear in Section 3.1 that the perturbation expansion is not fully converged in SAPT2+(3), especially at shorter intermolecular distances. It is also important to note that the total force-field energy is composed of several terms that cancel each other to a large extent. The exchange-repulsion term is always positive, while induction, dispersion and electrostatic are usually negative contributions. The typical behavior of all these terms is illustrated in Figure 10 for the dissociation of two dimers selected from the S66x8 database. An important consequence is that, in order to get reasonably accurate total energies, all the
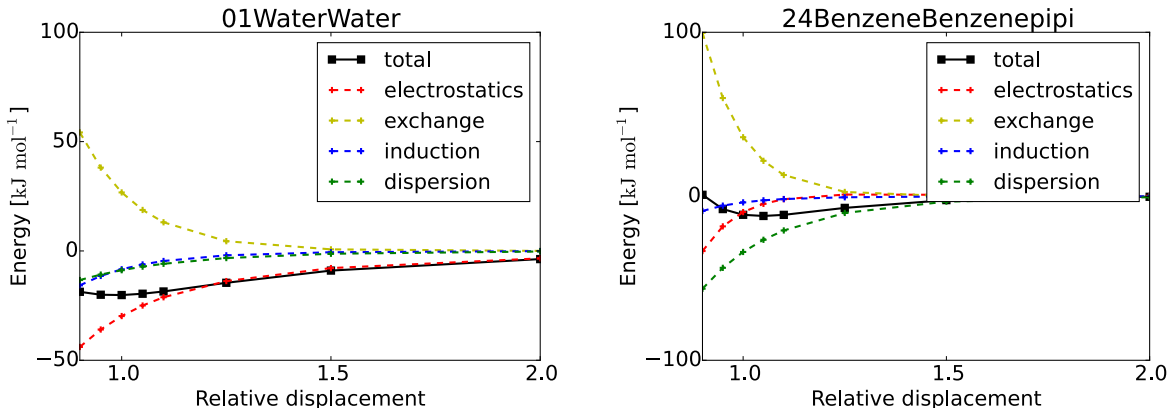

Figure 10: Dissociation curves showing components of the SAPT energy for the water dimer (left, hydrogen-bonded) and $\pi - \pi$ benzene dimer (right, dispersion-dominated).

components need to be very accurate. Otherwise small relative errors on one component will result in large relative errors on the total energy. This is why we prefer to further refine the interaction parameters together by fitting to total interaction energies from CCSD(T)/CBS calculations. A conventional minimization of a quadratic cost function often leads to an ill-conditioned fitting procedure in a high-dimensional parameter space, with an imminent danger of overfitting. Even when the number of parameters is rather modest, so-called sloppy modes can lead to parameters sets that are unphysical.[95,96] Furthermore all information deduced from the SAPT decomposition would be lost by using a conventional quadratic

29

cost function. Although the total SAPT energy may not be very accurate, the interaction parameters of our force field fitted to individual SAPT components certainly are an excellent first guess for further refinement of the parameters to reproduce CCSD(T)/CBS energies. This is why we will use ridge regression in the next subsection to refine the interaction parameters.

We note that only the difference of the interaction parameters $U_{\text{exch-rep}} - U_{\text{ind,ct}}$ is important, as both $U_{\text{exch-rep}}$ and $-U_{\text{ind,ct}}$ are prefactors of the same functional form (the overlap of electron densities). We however choose to treat these parameters separately in the remainder, as knowing the contribution of exchange-repulsion and induction can be useful for the physical interpretation of the interactions in the systems at hand.

In the previous sections, we have shown that our force field is particularly suited for the dispersion-dominated complexes. We will refer to this force field as MEDFF (Monomer Electron Density based Force Field) and will obtain universal values of the interaction parameters for MEDFF from fitting to the dispersion-dominated complexes from the S66x8 dataset. It is clear that especially the induction term needs improvement to also describe hydrogen-bonded and other complexes. To demonstrate that the functional form of our force field is appropriate for those complexes, we show results for a few selected cases where the interaction parameters are determined specifically for each case.

Ridge regression or Tikhonov regularization

A refinement of the interaction parameters can be performed by applying Tikhonov regularization, sometimes also termed ridge regression, to the conventional quadratic cost function, similar to Bayesian statistics.[97–99] In this work we will use the following cost function:

$$\chi^2 = \frac{1}{2} \sum_{n=1}^{N_d} \frac{\left(E_{\text{FF}}^n \left[\mathbf{U} \right] - E_{\text{ref}}^n \right)^2}{N_d} + \frac{1}{2\sigma_{\text{pr}}^2} \sum_{\alpha=1}^{N_u} \left(\frac{U_\alpha - U_\alpha^{\text{SAPT}}}{U_\alpha^{\text{SAPT}}} \right)^2 \qquad (26)$$

The cost function χ^2 has dimension [energy^2] (note that the parameter σ_{pr} has dimension [energy^{-1}] as all terms in the second summation are dimensionless). The sum over n runs over all N_d data points in the database, while the sum over α runs over all N_u interaction parameters, thus $N_u = 3$ in the case of MEDFF. The vector $\mathbf{U}$ contains all interaction parameters and the force-field energy $E^n_{\text{FF}}[\mathbf{U}]$ of dimer n depends on these interaction parameters. The second term, that regularizes the cost function, is called the prior and its influence is controlled by the parameter σ_{pr}. Our prior term in the cost function penalizes large relative deviations of the interaction parameters from their values computed by fitting to individual SAPT energy components, denoted by U^{SAPT}_α.

The parameter σ_{pr} cannot be computed a priori, instead its optimal value has to be determined as part of the fitting procedure. This can be achieved by first computing the estimated prediction error (EPE) as a function of σ_{pr}. The EPE can be estimated from the training dataset by cross validation (CV):

$$\text{EPE}^{\text{CV}} = \left(\sum_{n=1}^{N_d} \frac{(E^n_{\text{FF}}[\mathbf{U}(D_n)] - E^n_{\text{ref}})^2}{N_d} \right)^{1/2} \tag{27}$$

Here each of the N_d points in the database are in turn left out of the fitting procedure and the point that is left out, contributes to EPE^{CV}. We denote the database without point n as D_n. The notation $\mathbf{U}(D_n)$ is used to indicate that the parameters $\mathbf{U}$ are fitted to points in the database D_n, in this case the database that contains all data points except for point n. The optimal values of the parameters $\mathbf{U}$ depend on the value of σ_{pr} in the cost function, and in this way also EPE^{CV} depends on σ_{pr}. We computed EPE^{CV} for several values of σ_{pr}. In figure 11 we only consider dispersion-dominated complexes of the S66x8 database and use the force field for all energy contributions. We see that small values of σ_{pr} result in a high EPE because all interaction parameters are constrained to their SAPT values, which results in large differences with CCSD(T)/CBS energies. As soon as σ_{pr} is larger than $0.1\,\text{mol\,kJ}^{-1}$, the EPE does not decrease appreciably. Therefore we consider the parameters that are found

by minimizing the cost function with $\sigma_{\mathrm{pr}} = 0.1\,\mathrm{mol\,kJ^{-1}}$ to offer an optimal combination of being physical (close to parameters from fitting to separate SAPT components) and at the same time leading to good correspondence between CCSD(T)/CBS and force-field energies. The fact that the EPE does not increase again, indicates that we are not overfitting the parameters.

The values of the interaction parameters are given in Table 3 for different contributions of the prior to the cost function. If $\sigma_{\mathrm{pr}} = 0.0\,\mathrm{mol\,kJ^{-1}}$, the prior dominates the cost function and we obtain the values from fitting to the SAPT components separately. On the other hand, if $\sigma_{\mathrm{pr}} = \infty$, the prior does not contribute to the cost function and we obtain parameters from an unrestrained fit to CCSD(T)/CBS reference energies. If we choose $\sigma_{\mathrm{pr}} = 0.1\,\mathrm{mol\,kJ^{-1}}$ for MEDFF, we obtain parameters that show small relative deviations from parameters fitted to SAPT components and at the same time result in energies close to CCSD(T)/CBS reference values. In the remainder of this paper we will work with the parameters obtained by minimizing the cost function with $\sigma_{\mathrm{pr}} = 0.1\,\mathrm{mol\,kJ^{-1}}$.

We note that the refinement of the parameters mainly influences the interaction energies at rather small intermolecular separations. The resulting differences are in general smaller than the remaining errors with respect to the reference ab initio interaction energies.
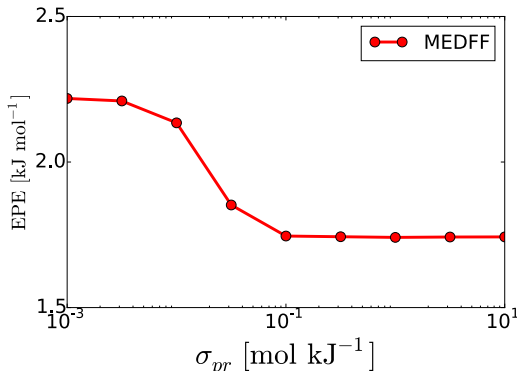

Figure 11: The EPE initially decreases as σ_{pr} increases (the prior contribution in the cost function diminishes). However, the EPE remains nearly constant as soon as $\sigma_{\mathrm{pr}} \geq 0.1\,\mathrm{mol\,kJ^{-1}}$.

Table 3: Optimal values of the MEDFF interaction parameters (all in atomic units) for different prior contributions (σ_{pr} in $\mathrm{mol\,kJ}^{-1}$) to the cost function.

$\sigma_{\mathrm{pr}}\,[\mathrm{mol\,kJ}^{-1}]$	0.0	0.1	∞
$U_{\mathrm{exch\text{-}rep}}$	8.13	8.43	8.27
$U_{\mathrm{ind,ct}}$	0.87	0.86	0.86
U_{s8}	0.57	0.57	0.48

Results

With the refinement of the interaction parameters according to the cost function (Equation 26), our force field MEDFF is completely constructed. The final values of the interaction parameters are tabulated in Table 3 in the column with $\sigma_{\mathrm{pr}} = 0.1\,\mathrm{mol\,kJ}^{-1}$. In a first internal validation we check whether MEDFF succeeds in reproducing the reference data that were used in the fitting procedure. We present the RMSD with respect to CCSD(T)/CBS interaction energies for our force field MEDFF in Table 4. We compare with the recently developed QMDFF,[18] where we note that QMDFF was fitted to B3LYP-D3/def2-QZVP reference data for the S22 dataset, a subset of the S66 dataset. MEDFF performs very well and can be considered chemically accurate (RMSD $< 1\,\mathrm{kcal\,mol}^{-1}$).We show the dissociation curves for a few selected dimers obtained with our force field MEDFF in Figure 12, and again compare with QMDFF. Dissociation curves for dispersion-dominated complexes are presented in the Supporting Information. These results reveal that the MEDFF force field performs remarkably well for all dispersion-dominated complexes. The neopentane dimer (index 36) is a notable exception, where also QMDFF fails to reproduce the dissociation curve correctly. A detailed analysis reveals that the main source of error is in the exchange-repulsion contribution, which is severely underestimated. This can also be seen in Figure 3, where it is clear that the value of the interaction parameter $U_{\mathrm{exch,rep}}$ fitted only to the neopentane dimer (index 36) is significantly above average.

We now turn our attention to some dimers that are not dominated by dispersion. Just as in Section 3.2 we will show the results for the water dimer (containing one hydrogen bond), the water-peptide dimer (containing one hydrogen bond), the acetamide dimer (containing

Table 4: RMSD of CCSD(T)/CBS and force-field interaction energies in $kJ\,mol^{-1}$ for the dispersion-dominated complexes in the S66x8 database. The definition of relative displacements is discussed elsewhere.[75] CCSD(T)/CBS and force-field energies are always compared for the same geometry.

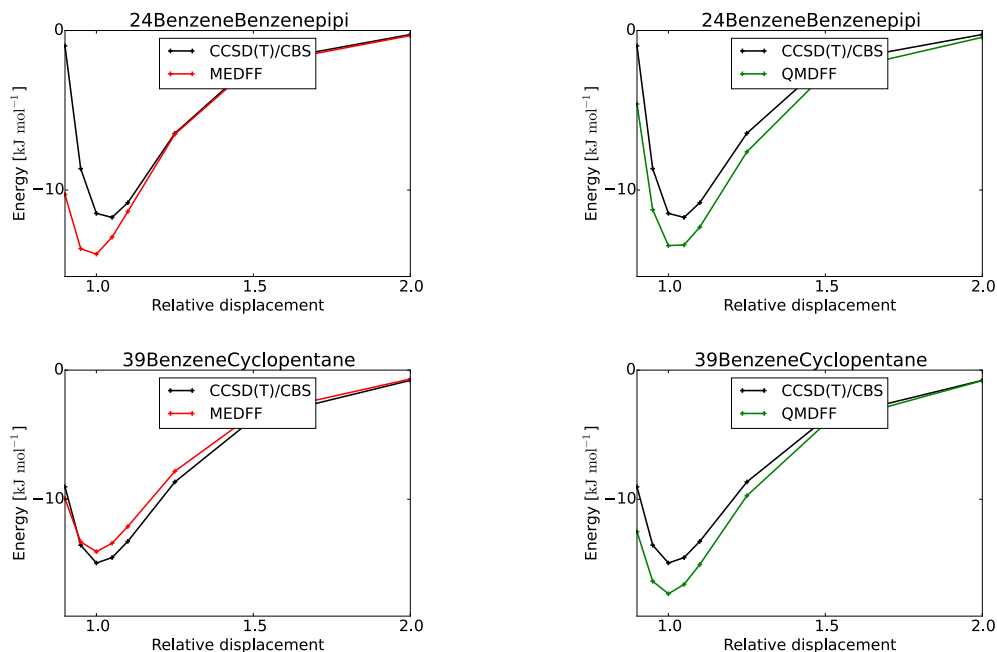| Relative displacement	MEDFF	QMDFF
0.90	3.7	6.2
0.95	2.2	5.3
1.00	1.5	4.7
1.05	1.1	4.2
1.10	0.9	3.8
1.25	0.6	2.7
1.50	0.4	1.5
2.00	0.2	0.5

Figure 12: Dissociation curves of the π-π pyridine dimer and the benzene-cyclopentane complex (both dispersion-dominated complexes).

two hydrogen bonds) and the benzene-methanol dimer (containing an OH-π bond) where we determine the parameters specifically for those complexes using the regularization procedure described at the beginning of this section. Again we use the 8 geometries of each complex in the S66x8 dataset as training data for the parameters for that specific complex. The geometries of the same complex in S66a8 is used afterwards as a validation. All results for the 4 complexes mentioned are summarized in Figure 13. For the water dimer the results are not entirely satisfying. A detailed analysis reveals that this is due to errors in the electrostatic term, most likely because the MBIS model gives isotropic atomic densities while atomic multipoles are indispensable for this specific case. For the other cases shown here the agreement is acceptable and the transferability of the model to the angular-displace geometries of the S66a8 dataset is remarkable. The results for all complexes are provided in the Supporting Information.

In Table 5 we summarize the results for the force field where parameters are obtained specifically for each complex, in contrast to our universal approach presented earlier. The errors with respect to CCSD(T)/CBS are well below 1 kcal mol^{-1} thereby demonstrating that our proposed energy expression can be used to reproduce high-level ab initio data. Of course one has to keep in mind that to use the force field in this way for complexes not present in the S66x8 dataset, a limited number of new ab initio calculations need to be performed.

4 External Validation

4.1 External validation on the hsg dataset

The next step in the validation of the new force field is an assessment in how far MEDFF succeeds in reproducing interaction energies of non-covalent complexes not included in the S66x8 dataset. For this external validation we select the hsg dataset.[100] The hsg dataset consists of 21 interacting fragment pairs, extracted from an HIV-II protease crystal struc-
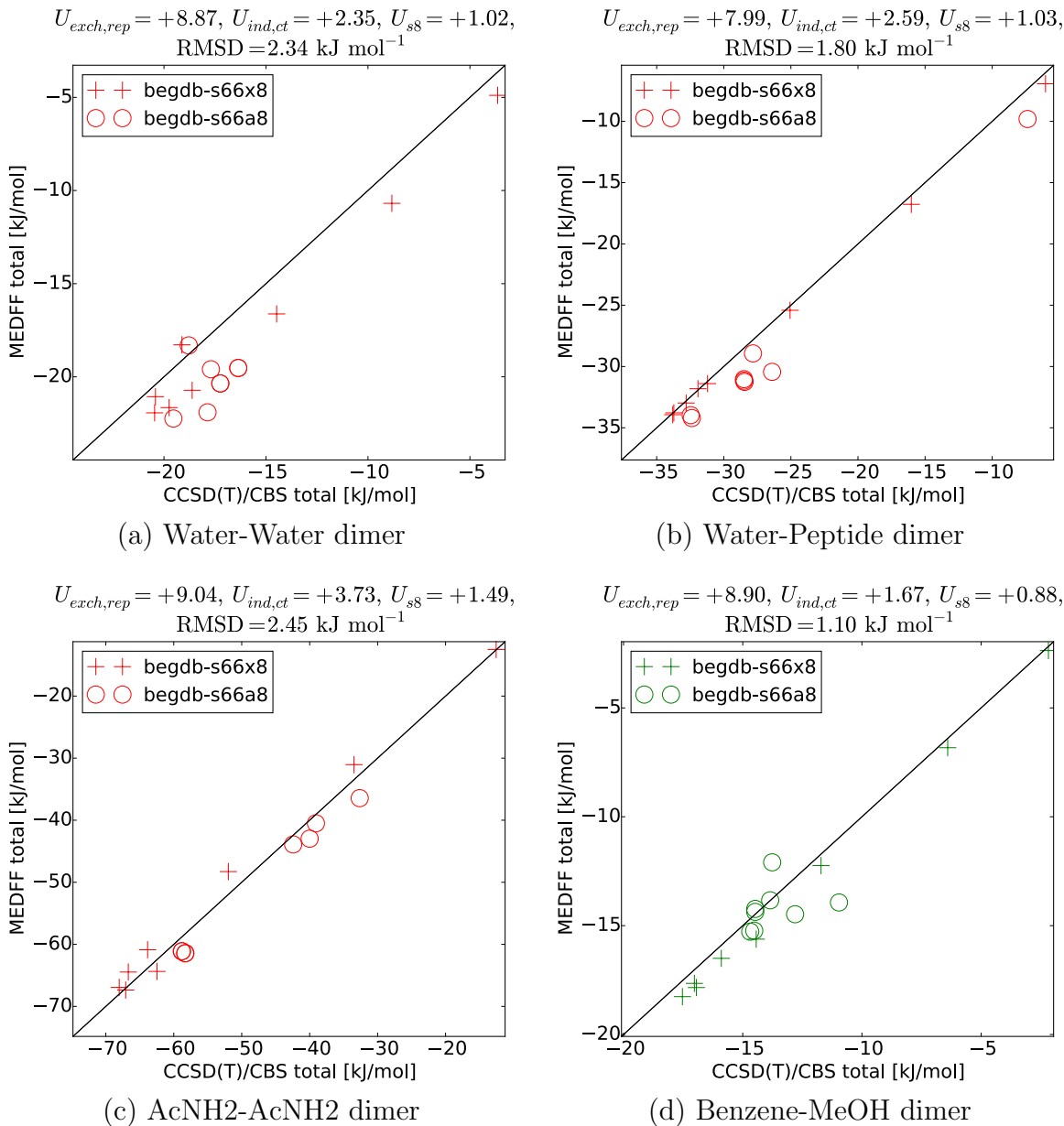
(a) Water-Water dimer

(b) Water-Peptide dimer

(c) AcNH2-AcNH2 dimer

(d) Benzene-MeOH dimer

Figure 13: Correlation plots of CCSD(T) and force-field interactions energies. The parameters $U_{\text{exch-rep}}$, U_{s_8} and $U_{\text{ind,ct}}$ have a different value for each complex, obtained by fitting to the 8 geometries of that complex in the S66x8 dataset. The RMSD is computed for all geometries of that complex in S66x8 and S66a8.

Table 5: RMSD of CCSD(T)/CBS and force-field interaction energies in $\mathrm{kJ\,mol^{-1}}$ for all complexes in the S66x8 database. The force-field interaction energies are obtained using parameters determined specifically for each separate complex.

Relative Displacement	Hydrogen-bonded	Dispersion-dominated	Others	All
0.90	1.89	0.72	1.34	1.40
0.95	0.46	0.39	0.88	0.60
1.00	1.50	0.58	1.13	1.13
1.05	2.39	0.69	1.32	1.64
1.10	2.94	0.73	1.38	1.94
1.25	3.31	0.68	1.20	2.10
1.50	2.26	0.51	0.72	1.42
2.00	0.70	0.24	0.26	0.46
All	2.15	0.59	1.09	1.45

ture with a bound ligand (indinavir). The molecules included in this dataset are, amongst others, 2-formamidoacetamide, N-methylacetamide, benzene, toluene, ethane, propane, butane, pyridine, methanol and acetic acid. We only retain the 13 dimers that are dominated by dispersion in this study. The values of the interaction parameters of our force field are not reoptimized, but instead taken from the fitting for the S66x8 dataset discussed earlier. This means that the computationally most expensive step to construct a force field for these complexes, is a DFT calculation for the monomers. We note that a similar computational effort is required for force fields where charges are obtained from an ab initio density (for example from ESP fitting or partitioning).

We compare the performance of our force field with some other force fields (Quantum Mechanically Derived Force Field[18] (QMDFF), the Generalized Amber Force Field[101] (GAFF) and the Merck Molecular Force Field[102] (MMFF)) in Figure 14. It is encouraging to see that our force field, MEDFF, performs well for dispersion-dominated complexes, even though the molecules present in this dataset were not used to fit the interaction parameters. Also the good performance of QMDFF for these complexes should be noted. Here we consider a complex to be dispersion-dominated if $\frac{E^{\mathrm{SAPT}}_{\mathrm{dispersion}}}{E^{\mathrm{SAPT}}_{\mathrm{electrostatics}}} \geq 1.7$, a definition corresponding to the classification used for the S66 dataset.[75]
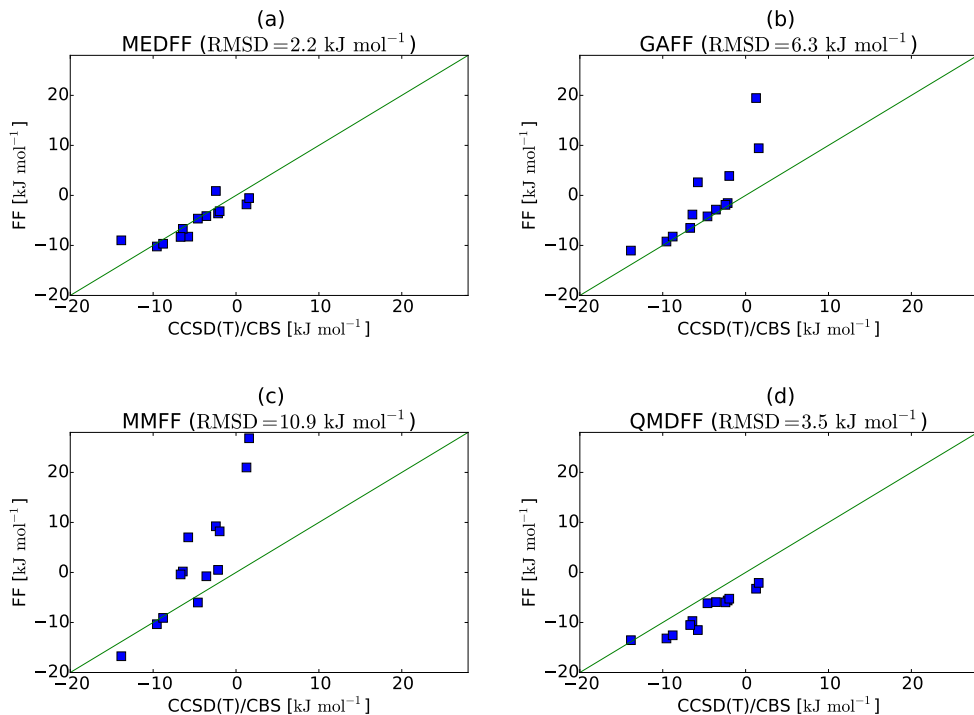
Figure 14: Performance of several force fields in reproducing CCSD(T)/CBS interaction energies for the dispersion-dominated complexes of the hsg dataset.

4.2 Second virial coefficients

To further validate our force field, we used it to compute second virial coefficients for some small alkanes and alkenes. Existing force fields can successfully predict this quantity, but such force fields are either based on experimental data,[103] on first principles calculations on alkane and alkene dimers[23] or on both.[104] Here, most force-field parameters for these systems will be derived solely from monomer computations on alkanes and alkenes. The three interaction parameters retain the values that were determined earlier without fitting to calculations for the dimers considered here, making this a proper validation of our method. The second virial coefficient describes the non-ideality of a gas in a first-order approximation. They form an interesting test case for our force field because experimental results are widely available and they only depend on the interaction between two molecules. The second virial

coefficient B_2 can be computed as follows[105]

$$B_2 = 2\pi \int \left(1 - \langle \exp\left[-\beta U_{\text{inter}}\left(r\right)\right]\rangle\right) r^2 dr \tag{28}$$

In this formula r is the distance between the centers of mass of the two molecules and $\beta = \frac{1}{k_B T}$. The brackets $\langle \ldots \rangle$ indicate an average over all possible configurations and orientations of the molecules. This average is weighted with a Boltzmann factor of the *intra*molecular energies of the configurations. To sample the molecular configurations with these weights, we performed MD simulations of an isolated molecule in the NVT ensemble for 5 ns with a time step of 0.5 fs. Where necessary, the dihedral force constants are rescaled in order to correctly reproduce the energy barriers obtained from a dihedral scan using the same ab initio level of theory. For instance in butane this is crucial to correctly sample gauche and trans states. The resulting intramolecular parameters are provided in the Supporting Information. For non-covalent interactions, MEDFF is used. Note that we neglect non-covalent interactions between $1 - 2$ and $1 - 3$ neighbours. To compute the second virial coefficient, two random configurations were selected from the MD run and one of the monomers was randomly rotated. This procedure enables the computation of B_2 by numerical integration over r with $1.5\text{Å} \leq r \leq 40\text{Å}$ in steps of 0.02Å. At each value of r, the orientational average was obtained by calculating the interaction energy for 100 different randomly generated orientations. The whole procedure was repeated 1 000 times with different randomly selected configurations, in order to properly compute the average over all possible configurations.

As mentioned, our force field is expected to work well for molecules where polarization plays a minor role. Thus, for a series of alkanes and alkenes we expect good performance. We limit the simulations to methane, ethane, propane, butane, ethene and propene, as for longer chains a proper intramolecular force field is strictly necessary to correctly sample molecular configurations. In general the agreement with experiment is rather good, as shown in Figure 15. The second virial coefficient is however slightly underestimated for the larger

molecules propene, propane and butane. We stress again that no dimer calculations for these molecules have to be performed to construct our force field. All necessary parameters come from monomer calculations with addition of a few interaction parameters obtained by fitting to reference data from the S66x8 database (accidentally ethene is present in the S66 database). This is in sharp contrast to most ab initio force fields where the parameters of a certain molecule are fitted to reproduce interaction energies for dimers containing that specific molecule.
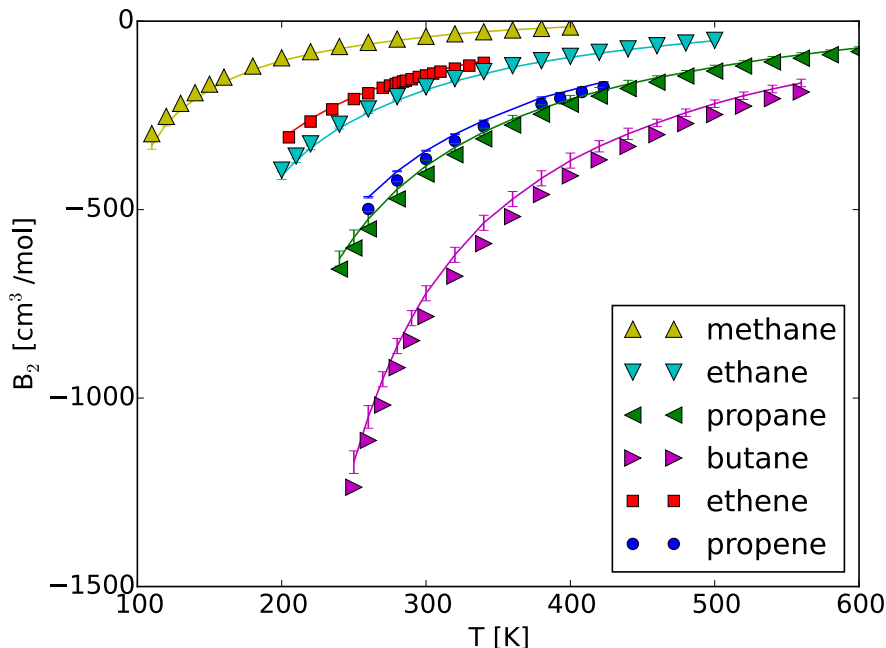

Figure 15: Comparison of second virial coefficients from our force field (symbols) with experiment (solid line). Experimental data for alkanes from Ref. 106, for propene from Ref. 107,108 and for ethene from Ref. 109.

4.3 Liquid properties

As a final validation, we applied the force field to compute properties of liquids. Because our force field is derived entirely based on ab initio computations in the gas phase, condensed-phase simulations are a stringent test to validate the assumptions made in the construction of the force field.

Computational details

We performed molecular dynamics simulations in the NPT-ensemble with our in-house code YAFF.[110] We used a Nosé-Hoover chain thermostat[111,112] with a time constant of 100 fs and chainlength 3, in combination with a Martyna-Tobias-Klein barostat[112,113] with time constant 1 000 fs. To treat the long-range electrostatic terms correctly, the Ewald summation[38,39] was used, while other non-bonded interactions were cut off at 12 Åusing a third-order polynomial to ensure smoothness of the potential and forces. The penetration contribution to the electrostatic interaction (the second term in 10) is also cut off at 12 Å. This is justified by the exponential decay of this term, which is clear from Equation 11. All simulations were carried out with 100 molecules in a cubic box and a time step of 0.5 fs. Every MD run lasted for 1 ns, of which the last 800 ps is used to collect data. The same force field as for the computation of virial coefficients, with an intramolecular force field obtained using QuickFF, was used.

Following quantities were computed based on the NPT simulation runs. The density ρ follows trivially from $\rho = \frac{M}{V}$, with M the mass of the system and V the volume. The enthalpy of vaporization ΔH_{vap} is approximated as:

$$\Delta H_{\text{vap}} \approx \langle E_{\text{pot}}^{\text{gas}} \rangle + k_B T - \langle E_{\text{pot}}^{\text{liq}} \rangle \tag{29}$$

where $E_{\text{pot}}^{\text{gas}}$ is obtained from a 100 ps NVT simulation of a single molecule.

Results

In line with calculations of the second virial coefficients, we performed MD simulations of methane, ethane, propane, butane, ethene and propene. We study liquid properties just below the boiling point of these molecules at a pressure of 1 atm. The experimental data for methane, ethane, propane and butane are taken from Ref. 114, for ethene from Ref. 115 and for propene from Ref. 116. Results for the density and enthalpy of vaporization are shown

in Table 6. In general the agreement with experiment is not very good, and we cannot claim that our force field predicts liquid properties that are quantitatively correct. The densities are systematically overestimated with errors ranging from +11% for ethane up to +20% for ethene. The enthalpies of vaporization are also overestimated for all molecules we studied, with errors ranging from +12% for ethane up to +23% for butane.

Table 6: Comparison of simulated and experimental liquid density ρ and enthalpy of vaporization ΔH_{vap}

Molecule	T [K]	ρ [g l^{-1}]		ΔH_{vap} [kJ mol^{-1}]	
		Experiment	Simulation	Experiment	Simulation
methane	111.6	422.8	502.5	8.2	9.2
ethane	184.5	544.6	618.2	14.7	16.4
propane	231.1	581.2	684.1	18.8	22.8
butane	272.6	601.1	713.8	22.4	28.3
ethene	169.2	568.0	673.1	13.5	16.1
propene	225.4	609.1	674.0	18.5	18.8

Discusssion

In the condensed phase, many-body effects will have a much larger impact than in the gas phase. With our pairwise-additive force field, it is impossible to properly describe many-body effects such as induction or Axilrod-Teller-Muto dispersion. Recently, it was shown how pairwise-additive force fields can be extended with explicit three-body terms for dispersion and exchange.[117] In the same paper, it was shown that neglect of these terms in ab initio derived force fields can have a large impact on predicting quantities such as density and heat of vaporization. Similar conclusions have been presented for liquid argon,[118] where the two-body force field overestimates densities with an error of 12%, while the inclusion of three-body corrections leads to good agreement with experimental results. Indeed, the necessity of many-body terms in general has been discussed before.[119,120]

In force fields based on experimental data, all these effects are included *implicitly* in the parameters. We illustrate this for the case of butane with the Optimized Potential for Liquid Simulations All Atom (OPLS-AA) force field.[121] The parameters of the intermolecular

potential are determined to correctly reproduce experimental liquid densities, heats of vaporization and radial distribution functions. For butane at the boiling point, this results in an error for the density of only -0.5% and only $+1.4\%$ for the heat of vaporization.[121] These errors are considerably smaller than errors obtained with MEDFF, $+19.0\%$ and $+24.8\%$ for liquid density and heat of vaporization respectively. In Figure 16 however, we show that the OPLS-AA force field fails to correctly reproduce the second virial coefficient of butane.[122] As its name suggests, OPLS-AA is not well suited to compute gas-phase properties. The fact that MEDFF performs better at reproducing the second virial coefficient than the empirical OPLS-AA force field, indicates indeed that the empirical OPLS-AA force field *implicitly* contains many-body effects (of critical importance in the condensed phase) in its parameters. Similar conclusions are obtained for the TraPPE force field for n-alkanes.[8]
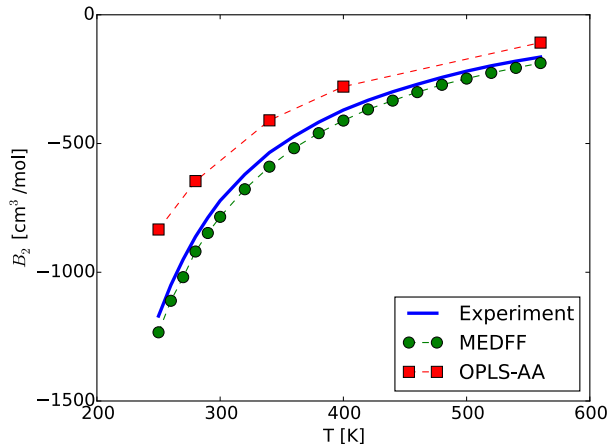

Figure 16: Second virial coefficient of butane. Experimental data from Ref. 106, OPLS-AA results from Ref. 122.

We elaborate on this point by comparing our force field MEDFF with a two-body force field derived from SAPT.[117] For methane under similar conditions, the predicted relative errors on density and heat of vaporization there are $+14\%$ and $+26\%$ respectively. By including explicit three-body terms, the errors are reduced to $+8\%$ and $+13\%$ respectively. Omission of repulsive three-body terms seems to result in an overestimation of liquid densities and of enthalpies of vaporization. This observation provides further evidence that the

errors in our current force field are to some extent due to its pairwise-additive character, and that these errors can be made significantly smaller by including explicit three-body terms in the force field. Another approach to include many-body dispersion effects is to map the valence-electron response onto a set of atom-centered quantum harmonic oscillators interacting through a dipole potential.[123–125] In this work, such effects are not taken into account yet, but this will be investigated in future extensions of the proposed method.

To obtain better results while maintaining the pairwise-additivity of our force field, we studied the following approach. Instead of deriving the MBIS parameters from the electron density in the gas phase, we derive them from the electron density from an ab initio calculation in the condensed phase. This offers an interesting way to *implicitly* include some many-body effects missing in the force field. We tested this approach for the case of methane. The resulting density and enthalpy of vaporization are indeed in closer agreement with experiment, compared to values obtained using MBIS parameters from gas-phase methane. The error is however still considerable, possibly because the interaction parameters have been derived using gas-phase interaction energies. In other words, not all many-body effects will be included using the suggested procedure. Another explanation is the limited accuracy of the two-body potential: after all MEDFF gives very good, but not perfect agreement with the second virial coefficient of methane. More details concerning this remark are given in Section 6 of the Supporting Information.

5 Summary and Conclusions

We present a procedure to derive pairwise-additive non-covalent force fields based on monomer electron densities. Our approach differs from existing methods because we avoid ill-conditioned fitting of parameters, we employ physically-motivated energy terms and we are able to use CCSD(T)/CBS data as a reference, because a limited number of data points is sufficient to fit the three interaction parameters. Similar approaches based on universal parameters for

non-covalent interactions have been presented before, but they usually involve more parameters. Furthermore our model is based on a distributed model of monomer electron densities, which allows to include the electrostatic penetration effect in an elegant and simple manner. By comparison with a SAPT decomposition of interaction energies for the S66x8 dataset, we show that all terms of our force field correctly reproduce the reference data. For the exchange-repulsion and dispersion contribution, the associated interaction parameter values are shown to be universal, i. e. largely independent of which specific dimer is considered. For the induction term, the same holds only for the dispersion-dominated dimers. For complexes not dominated by dispersion (for example hydrogen-bonded dimers) it is necessary to determine the induction parameter on a case-specific basis, so it takes on a different value for each complex studied. Next, the parameters for this pairwise-additive force field were refined by combining information from a SAPT energy decomposition with CCSD(T)/CBS interaction energies. The resulting force field is capable of accurately describing the dissociation curves that were used to fit the parameters.

As external validation, we first tested our force field on the hsg dataset, consisting of dimers extracted from an HIV-II protease crystal structure with a bound ligand (indinavir). Again, our force field performs well for the dispersion-dominated complexes. Furthermore experimental second virial coefficients for small alkanes and alkenes were successfully predicted. This is a remarkable success as the interaction parameters are not fitted specifically for these molecules and at no point experimental input is needed. Future extensions are needed to describe condensed-phase properties and hydrogen-bonded complexes. These extensions will cover many-body dispersion and exchange-repulsion terms, as well as a proper non-additive polarization model to describe induction interactions.

Acknowledgement

Supporting Information Available

In the Supporting Information, we derive analytical expressions for the integrals involving s-type Slater densities that appear in the force-field energy expression. Furthermore, we briefly review the method that is used to compute atomic pairwise C_{AB}^6 dispersion coefficients and show the dissociation curves for the dispersion-dominated dimers included in the S66 dataset computed using MEDFF, QMDFF and CCSD(T)/CBS. Scatter plots showing the performance of MEDFF with complex-specific parameters are given as well as the performance of some polarizable models for the induction energies. Finally a discussion on using condensed-phase electron densities to obtain the MBIS parameters is included. Additionally we also provide text files containing geometries, force-field parameters and interaction energies for the S66x8 and the hsg dataset and force-field parameters used to determine second virial coefficients and liquid densities for methane, ethane, propane, butane, ethene and propene. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Joos, L.; Swisher, J. A.; Smit, B. Molecular Simulation Study of the Competitive Adsorption of H2O and CO2 in Zeolite 13X. *Langmuir* **2013**, *29*, 15936–15942.

(2) Yot, P. G.; Boudene, Z.; Macia, J.; Granier, D.; Vanduyfhuys, L.; Verstraelen, T.; Van Speybroeck, V.; Devic, T.; Serre, C.; Férey, G.; Stock, N.; Maurin, G. Metal-organic frameworks as potential shock absorbers: the case of the highly flexible MIL-53(Al). *Chem. Commun.* **2014**, *50*, 9462–9464.

(3) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.

(4) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.* **1984**, *106*, 6638–6646.

(5) Allinger, N.; Yuh, Y.; Lii, J. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566.

(6) Rappé, A.; Casewit, C.; Colwell, K.; Goddard, W.; Skiff, W. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

(7) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(8) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569–2577.

(9) Ercolessi, F.; Adams, J. B. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Europhys. Lett.* **1994**, *26*, 583.

(10) Brommer, P.; Gähler, F. Potfit: effective potentials from *ab initio* data. *Model. Simul. Mater. Sci. Eng.* **2007**, *15*, 295.

(11) Li, A. H.-T.; Chao, S. D.; Chang, C.-C. Determination of a silane intermolecular force field potential model from an *ab initio* calculation. *Phys. Rev. A* **2010**, *82*, 062520.

(12) Transtrum, M. K.; Machta, B. B.; Brown, K. S.; Daniels, B. C.; Myers, C. R.; Sethna, J. P. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *J. Chem. Phys.* **2015**, *143*.

(13) Spackman, M. A. The use of the promolecular charge density to approximate the penetration contribution to intermolecular electrostatic energies. *Chem. Phys. Lett.* **2006**, *418*, 158 – 162.

(14) Zgarbová, M.; Otyepka, M.; Sponer, J.; Hobza, P.; Jurecka, P. Large-scale compensation of errors in pairwise-additive empirical force fields: comparison of AMBER in-

termolecular terms with rigorous DFT-SAPT calculations. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10476–10493.

(15) Donchev, A. G.; Ozrin, V. D.; Subbotin, M. V.; Tarasov, O. V.; Tarasov, V. I. A quantum mechanical polarizable force field for biomolecular interactions. *Proc. Natl. Acad. Sci.* **2005**, *102*, 7829–7834.

(16) Riley, K. E.; Platts, J. A.; Řezáč, J.; Hobza, P.; Hill, J. G. Assessment of the Performance of MP2 and MP2 Variants for the Treatment of Noncovalent Interactions. *J. Phys. Chem. A* **2012**, *116*, 4159–4169.

(17) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic Structure Theory*; John Wiley & Sons, LTD, Chichester, 2000.

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

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

(20) Yu, K.; McDaniel, J. G.; Schmidt, J. R. Physically Motivated, Robust, ab Initio Force Fields for CO2 and N2. *J. Phys. Chem. B* **2011**, *115*, 10054–10063.

(21) McDaniel, J. G.; Yu, K.; Schmidt, J. R. Ab Initio, Physically Motivated Force Fields for CO2 Adsorption in Zeolitic Imidazolate Frameworks. *J. Phys. Chem. C* **2012**, *116*, 1892–1903.

(22) McDaniel, J. G.; Schmidt, J. R. Robust, Transferable, and Physically Motivated Force Fields for Gas Adsorption in Functionalized Zeolitic Imidazolate Frameworks. *J. Phys. Chem. C* **2012**, *116*, 14031–14039.

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

(24) Van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R. Beyond BornMayer: Improved Models for Short-Range Repulsion in ab Initio Force Fields. *J. Chem. Theory Comput.* **2016**, *12*, 3851–3870, PMID: 27337546.

(25) Misquitta, A. J.; Stone, A. J. Ab Initio AtomAtom Potentials Using CamCASP: Theory and Application to Many-Body Models for the Pyridine Dimer. *J. Chem. Theory Comput.* **2016**, *12*, 4184–4208, PMID: 27467814.

(26) Tafipolsky, M.; Ansorg, K. Toward a Physically Motivated Force Field: Hydrogen Bond Directionality from a Symmetry-Adapted Perturbation Theory Perspective. *J. Chem. Theory Comput.* **2016**, *12*, 1267–1279.

(27) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 2609–2618.

(28) Qi, R.; Wang, Q.; Ren, P. General van der Waals potential for common organic molecules. *Bioorg. Med. Chem.* **2016**, *24*, 4911 – 4919.

(29) Wu, Q.; Ayers, P. W.; Zhang, Y. Density-based energy decomposition analysis for intermolecular interactions with variationally determined intermediate state energies. *J. Chem. Phys.* **2009**, *131*, 164112.

(30) Zhou, N.; Lu, Z.; Wu, Q.; Zhang, Y. Improved parameterization of interatomic potentials for rare gas dimers with density-based energy decomposition analysis. *J. Chem. Phys.* **2014**, *140*, 214117.

(31) Lu, Z.; Zhou, N.; Wu, Q.; Zhang, Y. Directional Dependence of Hydrogen Bonds: A Density-Based Energy Decomposition Analysis and Its Implications on Force Field Development. *J. Chem. Theory Comput.* **2011**, *7*, 4038–4049.

(32) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960–1986.

(33) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development. *J. Chem. Theory Comput.* **2016**, *12*, 3894–3912, PMID: 27385073.

(34) Wang, W.-P.; Parr, R. G. Statistical atomic models with piecewise exponentially decaying electron densities. *Phys. Rev. A* **1977**, *16*, 891–902.

(35) Wang, W.-P. Fixed-shell statistical atomic models with piecewise exponentially decaying electron densities. *Phys. Rev. A* **1982**, *25*, 2901–2912.

(36) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(37) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(38) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys. (Berlin)* **1921**, *369*, 253–287.

(39) de Leeuw, S. W.; Perram, J. W.; Smith, E. R. Simulation of Electrostatic Systems in Periodic Boundary Conditions. I. Lattice Sums and Dielectric Constants. *Proc. R. Soc. A* **1980**, *373*, 27–56.

(40) Hockney, R.; Eastwood, J. *Computer Simulation Using Particles*; CRC Press, 1988.

(41) Greengard, L.; Rokhlin, V. A Fast Algorithm for Particle Simulations. *J. Comput. Phys.* **1997**, *135*, 280–292.

(42) Tafipolsky, M.; Engels, B. Accurate Intermolecular Potentials with Physically Grounded Electrostatics. *J. Chem. Theory Comput.* **2011**, *7*, 1791–1803.

(43) Piquemal, J.-P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. Towards a force field based on density fitting. *J. Chem. Phys.* **2006**, *124*, 104101.

(44) Elking, D. M.; Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A.; Pedersen, L. G. Gaussian Multipole Model (GMM). *J. Chem. Theory Comput.* **2010**, *6*, 190–202.

(45) Stone, A. *The Theory of Intermolecular Forces*, 2nd ed.; Oxford University Press, 2013; Chapter 6.

(46) Kim, Y. S.; Kim, S. K.; Lee, W. D. Dependence of the closed-shell repulsive interaction on the overlap of the electron densities. *Chem. Phys. Lett.* **1981**, *80*, 574–575.

(47) Kita, S.; Noda, K.; Inouye, H. Repulsive potentials for ClR and BrR (R=He, Ne, and Ar) derived from beam experiments. *J. Chem. Phys.* **1976**, *64*, 3446–3449.

(48) Wheatley, R. J.; Price, S. L. An overlap model for estimating the anisotropy of repul-

sion. *Mol. Phys.* **1990**, *69*, 507–533.

(49) Mitchell, J. B. O.; Price, S. L. A Systematic Nonempirical Method of Deriving Model Intermolecular Potentials for Organic Molecules: Application To Amides. *J. Phys. Chem. A* **2000**, *104*, 10958–10971.

(50) 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. *J. Chem. Phys.* **1984**, *80*, 3726–3741.

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

(52) Hirshfeld, F. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **1977**, *44*, 129–138.

(53) Starkschall, G.; Gordon, R. G. Calculation of Coefficients in the Power Series Expansion of the Long-Range Dispersion Force between Atoms. *J. Chem. Phys.* **1972**, *56*, 2801–2806.

(54) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(55) Misquitta, A. J.; Stone, A. J. Accurate Induction Energies for Small Organic Molecules: 1. Theory. *J. Chem. Theory Comput.* **2008**, *4*, 7–18.

(56) Misquitta, A. J.; Stone, A. J.; Price, S. L. Accurate Induction Energies for Small Organic Molecules. 2. Development and Testing of Distributed Polarizability Models against SAPT(DFT) Energies. *J. Chem. Theory Comput.* **2008**, *4*, 19–32.

(57) Dick, B. G.; Overhauser, A. W. Theory of the Dielectric Constants of Alkali Halide Crystals. *Phys. Rev.* **1958**, *112*, 90–103.

(58) Schröder, K.-P.; ; Sauer, J. Potential Functions for Silica and Zeolite Catalysts Based on ab Initio Calculations. 3. A Shell Model Ion Pair Potential for Silica and Aluminosilicates. *J. Phys. Chem.* **1996**, *100*, 11043–11049.

(59) Lewis, G. V.; Catlow, C. R. A. Potential models for ionic oxides. *J. Phys. C: Solid*

State Phys. **1985**, *18*, 1149.

(60) Schmollngruber, M.; Lesch, V.; Schroder, C.; Heuer, A.; Steinhauser, O. Comparing induced point-dipoles and Drude oscillators. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14297–14306.

(61) Huang, L.; Roux, B. Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data. *J. Chem. Theory Comput.* **2013**, *9*, 3543–3556.

(62) Savelyev, A.; MacKerell, A. D. All-atom polarizable force field for DNA based on the classical drude oscillator model. *J. Comput. Chem.* **2014**, *35*, 1219–1239.

(63) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341–350.

(64) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046–4063.

(65) Elking, D.; Darden, T.; Woods, R. J. Gaussian induced dipole polarization model. *J. Comput. Chem.* **2007**, *28*, 1261–1274.

(66) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143–3161.

(67) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. ACKS2: Atom-condensed Kohn-Sham DFT approximated to second order. *J. Chem. Phys.* **2013**, *138*, 074108.

(68) Verstraelen, T.; Vandenbrande, S.; Ayers, P. W. Direct computation of parameters for accurate polarizable force fields. *J. Chem. Phys.* **2014**, *141*, 194114.

(69) Murrell, J. N.; Randic, M.; Williams, D. R. The Theory of Intermolecular Forces in the Region of Small Orbital Overlap. *Proc. R. Soc. A* **1965**, *284*, 566–581.

(70) Gresh, N.; Claverie, P.; Pullman, A. Computations of intermolecular interactions: Expansion of a charge-transfer energy contribution in the framework of an additive procedure. Applications to hydrogen-bonded systems. *Int. J. Quantum Chem.* **1982**, *22*, 199–215.

(71) Gresh, N.; Claverie, P.; Pullman, A. Intermolecular interactions: Elaboration on an additive procedure including an explicit charge-transfer contribution. *Int. J. Quantum Chem.* **1986**, *29*, 101–118.

(72) Hagberg, D.; Karlström, G.; Roos, B. O.; Gagliardi, L. The Coordination of Uranyl in Water: A Combined Quantum Chemical and Molecular Simulation Study. *J. Am. Chem. Soc.* **2005**, *127*, 14250–14256.

(73) Kumar, R.; Wang, F.-F.; Jenness, G. R.; Jordan, K. D. A second generation distributed point polarizable water model. *J. Chem. Phys.* **2010**, *132*, 014309.

(74) Stone, A. J.; Misquitta, A. J. Charge-transfer in Symmetry-Adapted Perturbation Theory. *Chem. Phys. Lett.* **2009**, *473*, 201–205.

(75) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.

(76) Morokuma, K.; Kitaura, K. *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Springer US, New York, 1981; pp 215–242.

(77) Bagus, P. S.; Hermann, K.; Bauschlicher, C. W. A new analysis of charge transfer and polarization for ligandmetal bonding: Model studies of Al4CO and Al4NH3. *J. Chem. Phys.* **1984**, *80*, 4378–4386.

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

(79) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. Psi4: an open-source ab initio electronic structure program. *WIREs Comput. Mol. Sci.* **2012**, *2*, 556–565.

(80) Hohenstein, E. G.; Sherrill, C. D. Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory. *J. Chem. Phys.* **2010**, *133*, 014101.

(81) Hohenstein, E. G.; Sherrill, C. D. Wavefunction methods for noncovalent interactions.

WIREs Comput. Mol. Sci. **2012**, *2*, 304–326.

(82) Hohenstein, E. G.; Sherrill, C. D. Efficient evaluation of triple excitations in symmetry-adapted perturbation theory via second-order Møller-Plesset perturbation theory natural orbitals. *J. Chem. Phys.* **2010**, *133*, 104107.

(83) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J. Chem. Phys.* **2014**, *140*, 094106.

(84) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N2, and H2O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

(85) Patkowski, K.; Korona, T.; Jeziorski, B. Convergence behavior of the symmetry-adapted perturbation theory for states submerged in Pauli forbidden continuum. *J. Chem. Phys.* **2001**, *115*, 1137–1152.

(86) Szalewicz, K.; Patkowski, K.; Jeziorski, B. In *Intermolecular Forces and Clusters II*; Wales, D. J., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; Chapter Intermolecular Interactions via Perturbation Theory: From Diatoms to Biomolecules, pp 43–117.

(87) 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.; Dannen-

berg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09 Revision D.01*; Gaussian Inc.: Wallingford CT, 2013.

(88) Toon Verstraelen, Katharina Boguslawski, Pawel Tecmer, Farnaz Heidar-Zadeh, Matthew Chan, Taewon D. Kim, Yilin Zhao, Steven Vandenbrande, Derrick Yang, Cristina E. González-Espinoza, Peter A. Limacher, Diego Berrocal, Ali Malek, Paul W. Ayers HORTON 2.0.0, http://theochem.github.com/horton/, 2015.

(89) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(90) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Modell.* **2006**, *25*, 247–260.

(91) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **2007**, *126*, 144111.

(92) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.

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

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

(95) Waterfall, J. J.; Casey, F. P.; Gutenkunst, R. N.; Brown, K. S.; Myers, C. R.; Brouwer, P. W.; Elser, V.; Sethna, J. P. Sloppy-Model Universality Class and the Vandermonde Matrix. *Phys. Rev. Lett.* **2006**, *97*, 150601.

(96) Gutenkunst, R. N.; Waterfall, J. J.; Casey, F. P.; Brown, K. S.; Myers, C. R.; Sethna, J. P. Universally Sloppy Parameter Sensitivities in Systems Biology Models.

PLoS Comput. Biol. **2007**, *3*, 1–8.

(97) Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*; Springer-Verlag New York, Inc.: Secaucus, NJ, USA, 2006.

(98) Petzold, V.; Bligaard, T.; Jacobsen, K. Construction of New Electronic Density Functionals with Error Estimation Through Fitting. *Top. Catal.* **2012**, *55*, 402–417.

(99) Cailliez, F.; Pernot, P. Statistical approaches to forcefield calibration and prediction uncertainty in molecular simulation. *J. Chem. Phys.* **2011**, *134*, 054124.

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

(101) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(102) Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720–729.

(103) Rodríguez, A. L.; Vega, C.; Freire, J. J.; Lago, S. Potential parameters of methyl and methylene obtained from second virial coefficients of n-alkanes. *Mol. Phys.* **1991**, *73*, 691–701.

(104) Nagy, J.; Jr., V. H. S.; Weaver, D. F. Calculation of second virial coefficients of alkanes with the MM2 and MM3 force fields. *Mol. Phys.* **1994**, *81*, 1039–1047.

(105) McQuarrie, D. *Statistical Mechanics*; University Science Books, Sausalito, 2000.

(106) Dymond, J.; Cholinski, J.; Szafranski, A.; Wyrzykowska-Stankiewicz, D. Second virial coefficients for N-alkanes; recommendations and predictions. *FFE* **1986**, *27*, 1–13.

(107) Glos, S.; Kleinrahm, R.; Wagner, W. Measurement of the (p, ρ, T) relation of propane, propylene, n-butane, and isobutane in the temperature range from (95 to 340) K at pressures up to 12MPa using an accurate two-sinker densimeter. *J. Chem. Thermodyn.* **2004**, *36*, 1037 − 1059.

(108) Warowny, W.; Wielopolski, P.; Stecki, J. Compressibility factors and virial coefficients for propane, propene and their mixtures by the burnett method. *Phys. A* **1978**, *91*,

73–87.

(109) Nowak, P.; Kleinrahm, R.; Wagner, W. Measurement and correlation of the (p, ρ, T) relation of ethylene I. The homogeneous gaseous and liquid regions in the temperature range from 105K to 340K at pressures up to 12 MPa. *J. Chem. Thermodyn.* **1996**, *28*, 1423–1439.

(110) Vanduyfhuys, L.; Vandenbrande, S.; Verstraelen, T. YAFF, Yet Another Force Field. `http://molmod.ugent.be/software/`, (Accessed 09/30/2016).

(111) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(112) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **1996**, *87*, 1117–1157.

(113) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

(114) Younglove, B. A.; Ely, J. F. Thermophysical Properties of Fluids. II. Methane, Ethane, Propane, Isobutane, and Normal Butane. *J. Phys. Chem. Ref. Data* **1987**, *16*, 577–798.

(115) Smukala, J.; Span, R.; Wagner, W. New Equation of State for Ethylene Covering the Fluid Region for Temperatures From the Melting Line to 450K at Pressures up to 300MPa. *J. Phys. Chem. Ref. Data* **2000**, *29*, 1053–1121.

(116) Lemmon, E.; McLinden, M.; Friend, D. *"Thermophysical Properties of Fluid Systems" in NIST Chemistry WebBook*, 1st (revised) ed.; National Institute of Standards and Technology, Gaithersburg MD, 20899, 1996.

(117) McDaniel, J. G.; Schmidt, J. R. First-Principles Many-Body Force Fields from the Gas Phase to Liquid: A Universal Approach. *J. Phys. Chem. B* **2014**, *118*, 8042–8053.

(118) Desgranges, C.; Delhommelle, J. Many-Body Effects on the Thermodynamics of Fluids, Mixtures, and Nanoconfined Fluids. *J. Chem. Theory Comput.* **2015**, *11*, 5401–5414.

(119) DiStasio Jr., R. A.; Gobre, V. V.; Tkatchenko, A. Many-body van der Waals interactions in molecules and condensed matter. *J. Phys. Condens. Matter* **2014**, *26*, 213202.

(120) Schmidt, J. R.; Yu, K.; McDaniel, J. G. Transferable Next-Generation Force Fields from Simple Liquids to Complex Materials. *Acc. Chem. Res.* **2015**, *48*, 548–556.

(121) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. Free Energies of Hydration and Pure Liquid Properties of Hydrocarbons from the OPLS All-Atom Model. *J. Phys. Chem.* **1994**, *98*, 13077–13082.

(122) Nagy, J.; Weaver, D. F.; Smith Jr., V. H. A Comprehensive Study of Alkane Non-bonded Empirical Force Fields. Suggestions for Improved Parameter Sets. *J. Phys. Chem.* **1995**, *99*, 8058–8065.

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

(124) Tkatchenko, A.; Ambrosetti, A.; DiStasio Jr., R. A. Interatomic methods for the dispersion energy derived from the adiabatic connection fluctuation-dissipation theorem. *J. Chem. Phys.* **2013**, *138*, 074106.

(125) Ambrosetti, A.; Reilly, A. M.; DiStasio Jr., R. A.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *J. Chem. Phys.* **2014**, *140*, 18A508.

Graphical TOC Entry