

# Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development

Toon Verstraelen,<sup>\*,†</sup> Steven Vandenbrande,<sup>†</sup> Farnaz Heidar-Zadeh,<sup>‡</sup> Louis Vanduyfhuys,<sup>†</sup> Veronique Van Speybroeck,<sup>†</sup> Michel Waroquier,<sup>†</sup> and Paul W. Ayers<sup>‡</sup>

<sup>†</sup>*Center for Molecular Modeling (CMM), Member of the QCMM Ghent–Brussels Alliance, Ghent University, Technologiepark 903, B9000 Ghent, Belgium*

<sup>‡</sup>*Department of Chemistry and Chemical Biology, McMaster University, 1280 West Main Street, Hamilton, Ontario L8S 4M1, Canada*

E-mail: toon.verstraelen@ugent.be

## Abstract

Atomic partial charges appear in the Coulomb term of many force-field models and can be derived from electronic structure calculations with a myriad of atoms-in-molecules (AIM) methods. More advanced models have also been proposed, using the distributed nature of the electron cloud and atomic multipoles. In this work, an electrostatic force field is defined through a concise approximation of the electron density, for which the Coulomb interaction is trivially evaluated. This approximate “pro-density” is expanded in a minimal basis of atom-centered s-type Slater density functions, whose parameters are optimized by minimizing the Kullback-Leibler divergence of the pro-density from a reference electron density, e.g. obtained from an electronic structure calculation. The proposed method, Minimal Basis Iterative Stockholder (MBIS), is a variant of the Hirshfeld AIM method but it can also be used as a density-fitting

technique. An iterative algorithm to refine the pro-density is easily implemented with a linear-scaling computational cost, enabling applications to supramolecular systems. The benefits of the MBIS method are demonstrated with systematic applications to molecular databases and extended models of condensed phases. A comparison to 14 other AIM methods shows its effectiveness when modeling electrostatic interactions. MBIS is also suitable for rescaling atomic polarizabilities in the Tkatchenko-Sheffler scheme for dispersion interactions.

## 1 Introduction

The importance of force-field models is evident from recent hallmarks of atomistic force-field simulations in biology, such as the full characterization of  $\beta 2$  adrenergic receptor with a Markov state model<sup>1</sup> and the Anton 2 computer that can perform 10  $\mu$ s molecular dynamics simulations per day on a system containing one million atoms.<sup>2</sup> In many other domains, impressive scientific breakthroughs were also realized with atomistic force-field simulations, such as the virtual screening of 87 000 zeolites for the selective adsorption of CO<sub>2</sub>.<sup>3</sup>

An efficient and reliable model for electrostatic interactions is a fundamental component of a force-field model. For example, molecular recognition in proteins can be driven by electrostatic interactions.<sup>4</sup> Several authors studied partial charges derived from electronic wavefunctions to model electrostatic interactions in metal-organic frameworks.<sup>5–8</sup> Depending on the framework type, the values of the partial charges can strongly affect the predicted adsorption isotherms and self-diffusion coefficients.<sup>9</sup> Energy decomposition methods have also shown that electrostatic interactions are one of the main driving forces in the formation of hydrogen bonds.<sup>10</sup>

In this work, we propose a new and transparent method to derive, from an electronic wavefunction, a robust, compact and reliable model for electrostatic interactions that is easily included in a force-field model. The goal is thus an efficient computation of electrostatic interactions between the molecules in the frozen density approximation,<sup>10–12</sup> i.e. without

accounting for induction or polarization effects. Although it is important and challenging to account for polarization in force fields,<sup>13–15</sup> the development of polarizable force fields goes beyond the scope of this paper. We also do not consider so-called polarized force fields, where polarization is described effectively by computing the charges from an electronic structure calculation with a polarizable continuum model.<sup>16</sup>

The most basic and widespread electrostatic force-field model consists of interacting atomic point charges placed at the positions of the nuclei. In older works, e.g. the TraPPE force field for CO<sub>2</sub>,<sup>17</sup> the partial charges are fitted to experimental thermodynamic reference data. More recently, e.g. as in the TraPPE-EH models,<sup>18</sup> charges are often derived from electronic wavefunctions. Plenty of methods exist to compute such partial charges but usually, for force-field purposes, they are fitted to the electrostatic potential around model compounds of interest,<sup>19</sup> e.g. extensions of the AMBER force field often use the RESP method for partial charges.<sup>20</sup> The point-charge model is only a very crude representation of the molecular charge distribution; it does not account for finer details such as atomic multipoles<sup>21</sup> and the spatial distribution of the electron cloud. The spatial distribution becomes important when electron densities of two atoms or molecules begin to overlap: in that regime point-multipole models neglect a relatively large attractive electrostatic force, which is known as the penetration effect.<sup>12,22–26</sup>

In principle, atomic multipoles are easily computed with an atoms-in-molecules (AIM) method. In some works, the acronym AIM is used exclusively for Quantum Theory of Atoms in Molecules<sup>27</sup> (QTAIM). Here it is used more generally, to refer to any method that partitions the molecular electron density,  $\rho(\mathbf{r})$ , into atomic contributions,  $\rho_A(\mathbf{r})$ , from which e.g. atomic multipole moments can be derived. The spatial distribution of the electron density is sometimes also modeled with density-fitting techniques, e.g. as in the Gaussian Electrostatic Model<sup>28,29</sup> or related methods.<sup>30–32</sup> This leads to very accurate models of the electronic density but the use of such advanced charge distributions in force-field simulations poses some difficulties: the conformational dependence of atomic multipoles can be very

complex and it is far from trivial to include torques acting on higher moments in a force-field model. Several authors have proposed methods to overcome these challenges, e.g. with rigid molecules<sup>33</sup> or with machine learning methods.<sup>34</sup> Such advanced techniques are not always feasible for large-scale simulations. In this work, we propose a mathematically elegant and compact approximation of the electron distribution that results in relatively accurate electrostatic interactions in force-field models, without compromising computational efficiency. Only spherically symmetric models for atoms are considered and generalizations toward non-spherical atoms will be studied in future work.

Our new method minimizes the Kullback-Leibler (KL) Divergence of a pro-density, a minimal expansion in atom-centered s-type Slater functions, from a given molecular electron density.<sup>35,36</sup> This approach is closely related to the Iterative Stockholder (IS) method,<sup>37</sup> where the pro-molecule density is a sum of spherical non-negative pro-atom densities, without any restrictions on their radial dependence. As will be discussed in section 2, the algorithm to optimize our pro-density parameters is also very similar to IS. Hence, we refer to our new method as Minimal Basis Iterative Stockholder (MBIS).

MBIS can be perceived in two different ways. In the first place, it is a variant of the Hirshfeld method:<sup>38</sup> a partitioning of the molecular electron density inspired by information theory.<sup>39</sup> Second, it can also be seen as a density fitting technique that uses the KL-divergence,<sup>36</sup> instead of the more common least-squares approach with a Coulomb metric,<sup>40</sup> to optimize the model density. This duality permits many applications, also beyond the scope of modeling electrostatic interactions. For example, the Hirshfeld method is extensively used in different dispersion corrections for Density Functional Theory computations.<sup>41–43</sup> Furthermore, AIM populations are widely used in conceptual density functional theory to compute condensed reactivity indicators.<sup>44</sup>

Several related AIM methods were proposed in the literature, each trying to improve certain properties of their predecessors. The original Hirshfeld method<sup>38</sup> has some well-known weaknesses, such as the relatively low partial charges<sup>45</sup> and some deficiencies in its moti-

vation from information theory.<sup>46</sup> These issues were mostly fixed in the Iterative Hirshfeld (HI) method:<sup>46</sup> charges computed with this method reproduce well the electrostatic potential around a molecule.<sup>47–49</sup> Compared to ESP-fitted charges, HI charges are also relatively robust with respect to conformational changes, choice of basis set, etc.<sup>49,50</sup> Unfortunately, also Iterative Hirshfeld has its deficiencies. For example, when the method is applied to highly polar oxides, it requires the spherically averaged density of the non-existing oxygen dianion as input.<sup>51</sup> When this dianion density is computed with a localized basis set, iterative Hirshfeld charges severely overestimate electrostatic potentials of metal oxides.<sup>51,52</sup> The Iterative Stockholder (IS) analysis was developed independently from the Iterative Hirshfeld method and it addresses most of the issues mentioned so far.<sup>51,53</sup> However, IS charges are not very robust with respect to conformational changes, similar to ESP-fitted charges.<sup>54</sup> A recent analysis revealed that the lack of robustness is strongly related to the ill-defined density tails of the IS pro-atoms, while the core region of the IS pro-atom is usually well defined.<sup>30</sup> Several authors have presented solutions to overcome the weaknesses of the Iterative Hirshfeld and Iterative Stockholder methods.<sup>30,51,54–57</sup> A general difficulty with these recent efforts is that they all significantly increase the algorithmic complexity and/or introduce many tuned parameters that are needed as extra input for the partitioning. In this work, we will reverse this trend and propose a method that is mathematically elegant, straightforward to implement for large systems and free from empirical input (like atomic radii) or pre-computed pro-atoms.

In the development of the MBIS method, we paid special attention to its applicability to condensed phases and extended systems. One of the applications of interest is the automatic derivation of environment specific force-field parameters for supramolecular systems<sup>58</sup> and porous materials.<sup>6,59</sup> In such applications, density partitioning is applied to DFT calculations of large atomistic models, from which force-field parameters are derived. Besides the obvious requirement that an accurate model for electrostatics must be obtained, it is also essential that the atoms-in-molecules method is computationally feasible for large systems.

In practice, this means that the computational cost must scale linearly with the system size. This is achieved in MBIS by using only well-behaved integrals over atomic regions whose cost is independent of the system size.

The paper is organized as follows. In section 2, the Minimal Basis Iterative Stockholder (MBIS) method is derived using arguments from information theory, followed by more practical aspects such as numerical algorithms and software implementations. Section 3 showcases typical MBIS results with two brief applications. Section 4 compares MBIS to 14 other AIM methods, assessing the robustness of charges and the accuracy of electrostatic potentials and electrostatic interactions. Some specific advantages of MBIS over (Iterative) Hirshfeld, are presented in section 5, by testing different variants of the Tkatchenko-Scheffler dispersion model.<sup>42</sup> Finally, our conclusions and an outlook on future work are given in section 6.

## 2 Minimal Basis Iterative Stockholder Method

### 2.1 Information theory approach to Hirshfeld partitioning

It is instructive to review the information theory arguments<sup>39</sup> that support the Hirshfeld method.<sup>38</sup> The amount of information lost when atoms-in-molecules (AIM) densities are approximated by pro-atoms, can be expressed as the sum of the KL-divergence for every atom:

$$\Delta S[\{\rho_A\}; \{\rho_A^0\}] = \sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) \ln \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} d\mathbf{r}, \quad (1)$$

Traditionally, the pro-atoms,  $\rho_A^0$ , are fixed and the AIM densities,  $\rho_A$ , are the unknowns to be determined. In the original Hirshfeld method, spherically averaged isolated neutral atoms are used as pro-atoms. To obtain AIM densities that are maximally similar to the pro-atoms, one minimizes the information loss with the constraint that the AIM densities have to add up to the total density:  $\sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) = \rho(\mathbf{r})$ . Hence, the optimal AIM densities

are a stationary point of the following Lagrangian:

$$\begin{aligned}
L_0[\{\rho_A\}, \lambda(\mathbf{r}); \{\rho_A^0\}] = & \\
& \sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) \ln \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} d\mathbf{r} \\
& + \int \lambda(\mathbf{r}) \left( \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) - \rho(\mathbf{r}) \right) d\mathbf{r}.
\end{aligned} \tag{2}$$

with the Lagrange multiplier  $\lambda(\mathbf{r})$  and with fixed pro-atom densities  $\rho_A^0(\mathbf{r})$ . The Lagrange equations take the following form

$$0 = \frac{\delta L_0}{\delta \rho_A(\mathbf{r})} = \ln \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} + 1 + \lambda(\mathbf{r}) \quad \forall A. \tag{3}$$

The solution is:

$$\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} = \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \quad \forall A \neq B. \tag{4}$$

After multiplication by  $\rho_A^0(\mathbf{r})\rho_B^0(\mathbf{r})$  and summing over all atoms  $B$ , one obtains the well-known stockholder partitioning:<sup>39</sup>

$$\rho_A(\mathbf{r}) = \rho(\mathbf{r}) \frac{\rho_A^0(\mathbf{r})}{\rho^0(\mathbf{r})} \quad \text{with} \quad \rho^0(\mathbf{r}) = \sum_B \rho_B^0(\mathbf{r}), \tag{5}$$

which corresponds to the definition originally given by Hirshfeld.<sup>38</sup> The name *stockholder* comes from the ratio  $\rho_A^0(\mathbf{r})/\rho^0(\mathbf{r})$ : at every point in space it represents the *share* of pro-atom  $A$  in the total pro-density. It can be interpreted as an atomic weight function that assigns part of the total electron density to atom  $A$ . In most Hirshfeld variants,<sup>38,46,51,53,54</sup> the weight function varies smoothly over the range  $[0, 1]$ . In QTAIM,<sup>27</sup> a similar atomic weight function, derived from the topology of  $\rho(\mathbf{r})$ , is either 1 inside the atomic basin or 0 elsewhere. Due to the minimization of the KL-divergence, the Hirshfeld AIM densities are maximally similar to the pro-atoms, ensuring some degree of transferability between AIM densities in different molecules.<sup>60</sup>

The use of fixed pro-atoms has some important disadvantages. Results obtained with the Hirshfeld partitioning method depend largely on the choice of the fixed pro-atoms, which is essentially arbitrary.<sup>46</sup> Furthermore,  $\rho_A^0(\mathbf{r})$  and  $\rho_A(\mathbf{r})$  do not necessarily have the same norm ( $N_A^0 \neq N_A$ ), such that the KL-divergence cannot be used as a proper measure for information loss.<sup>61</sup> This shortcoming was one of the motivations to develop the iterative Hirshfeld (HI) method.<sup>46</sup> In HI, the pro-atoms are not fully fixed a priori but rather updated iteratively to achieve consistency between the charge of the pro-atom and the AIM density.

## 2.2 Definition of the MBIS partitioning

In this paper, we will make use of the information theory concepts reviewed in the previous subsection, yet with a different model for the pro-atomic density:

$$\rho_A^0(\mathbf{r}) = \sum_{i=1}^{m_A} \rho_{Ai}^0(\mathbf{r}), \quad (6)$$

with

$$\rho_{Ai}^0(\mathbf{r}) = N_{Ai} f_{Ai}(\mathbf{r}) = \frac{N_{Ai}}{\sigma_{Ai}^3 8\pi} \exp\left(-\frac{|\mathbf{r} - \mathbf{R}_A|}{\sigma_{Ai}}\right), \quad (7)$$

where the number of Slater functions,  $m_A$ , is the number of shells of atom  $A$ , i.e. its row in the periodic table. Both the population,  $N_{Ai}$ , and the width,  $\sigma_{Ai}$ , of each atomic shell are free variables. The shape functions,  $f_{Ai}(\mathbf{r})$ , are normalized 1s Slater-type density functions ( $\int f_{Ai}(\mathbf{r}) d\mathbf{r} = 1$ ) and hence the population of a pro-atom is simply  $N_A^0 = \sum_{i=1}^{m_A} N_{Ai}$ . Figure 1a illustrates the expansion of the density in Slater functions, for the case of a carbon dioxide molecule.

It is clear that the pro-atom parameterization with s-type Slater functions is only applicable to (reconstructed) all-electron densities. Regardless of this requirement, the MBIS method has many advantages over existing methods, as will be extensively shown in the remainder of the paper. Future work will focus on more advanced pro-atom models, e.g. to make them also suitable for pseudo densities, while still maintaining a numerically robust



algorithm. In this work, only the most minimal, yet very effective, parameterization of the pro-atoms is considered.

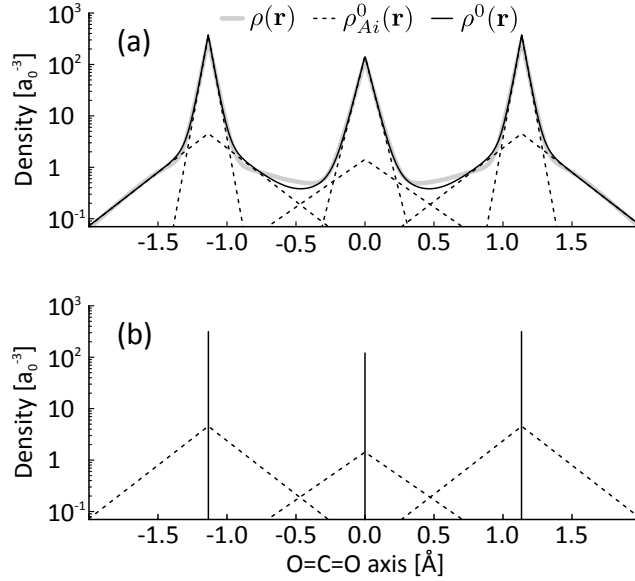


Figure 1: (a) An expansion of the molecular electron density of carbon dioxide (solid gray curve,  $\rho$ ) in a minimal number of 1s Slater-type density functions (black dashed curves,  $\rho_{Ai}^0$ ). The sum of all Slater functions is the pro-molecular density (solid black curve,  $\rho^0$ ). (b) A reduction suitable for force-field models: for every atom  $A$ , the nuclear charge and the core Slater functions are condensed into an effective core charge (solid vertical line,  $q_{A,c}$ ), while the valence Slater function (dashed black curve, parameters  $N_{A,v}$  and  $\sigma_{A,v}$ ) is retained.

All the pro-atom parameters,  $\{N_{Ai}\}$  and  $\{\sigma_{Ai}\}$ , and the AIM densities,  $\rho_A(\mathbf{r})$  will be optimized by minimizing the information loss. The main difference with the conventional Hirshfeld method is that also a set of pro-atom parameters is varied, such that the pro-atom densities become a good approximation of the AIM densities. These additional degrees of freedom also allow us to constrain the population of each pro-atom and corresponding AIM to be equal, avoiding any ambiguity in the statistical interpretation of Eq. (1).<sup>61</sup> The Lagrangian

for this problem is an extension of Eq. (2) with additional variables and constraints:

$$\begin{aligned}
L_1[\{\rho_A\}, \lambda(\mathbf{r}), \{N_{Ai}\}, \{\sigma_{Ai}\}, \{\mu_A\}] = & \\
& \sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) \ln \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} d\mathbf{r} \\
& + \int \lambda(\mathbf{r}) \left( \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) - \rho(\mathbf{r}) \right) d\mathbf{r} \\
& + \sum_{A=1}^{N_{\text{atoms}}} \mu_A \int (\rho_A^0(\mathbf{r}) - \rho_A(\mathbf{r})) d\mathbf{r},
\end{aligned} \tag{8}$$

where  $\mu_A$  are new Lagrange multipliers associated with the consistency of the pro-atom and AIM populations.

Independent variation of the Lagrangian  $L_1$  with respect to each variable ( $\rho_A(\mathbf{r})$ ,  $N_{Ai}$  or  $\sigma_{Ai}$ ) leads to a set of Lagrange equations, which, together with the constraints, determine the MBIS AIM and pro-atom densities and the Lagrange multipliers  $\lambda(\mathbf{r})$  and  $\{\mu_A\}$ .

We first consider the derivative of  $L_1$  toward  $N_{Ai}$ :

$$0 = \frac{\partial L_1}{\partial N_{Ai}} = \int \frac{\delta L_1}{\delta \rho_A^0(\mathbf{r})} \frac{\partial \rho_A^0(\mathbf{r})}{\partial N_{Ai}} d\mathbf{r} \tag{9}$$

$$= \int \left( -\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} + \mu_A \right) f_{Ai}(\mathbf{r}) d\mathbf{r} \tag{10}$$

$$= \mu_A - \int \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} f_{Ai}(\mathbf{r}) d\mathbf{r}. \tag{11}$$

When we multiply by  $N_{Ai}$  and sum over the shells  $i$  of atom  $A$ , we get:

$$0 = \sum_i^{m_A} N_{Ai} \frac{\partial L_1}{\partial N_{Ai}} = \mu_A N_A^0 - \int \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \sum_i^{m_A} N_{Ai} f_{Ai}(\mathbf{r}) d\mathbf{r} \tag{12}$$

$$= \mu_A N_A^0 - N_A. \tag{13}$$

Due to the constraint  $N_A = N_A^0$ , we have  $\mu_A = 1$  for each atom. Next, we take the functional

derivative of  $L_1$  toward  $\rho_A(\mathbf{r})$  and make use of  $\mu_A = 1$ :

$$0 = \frac{\delta L_1}{\delta \rho_A(\mathbf{r})} = \ln \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} + \lambda(\mathbf{r}), \quad (14)$$

whose solution is the stockholder partitioning formula in Eq. (5). Finally, we consider the derivative of  $L_1$  toward  $\sigma_{Ai}$ :

$$0 = \frac{\partial L_1}{\partial \sigma_{Ai}} = \int \frac{\delta L_1}{\delta \rho_A^0(\mathbf{r})} \frac{\partial \rho_A^0(\mathbf{r})}{\partial \sigma_{Ai}} d\mathbf{r} \quad (15)$$

$$= - \int \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \frac{\partial \rho_A^0(\mathbf{r})}{\partial \sigma_{Ai}} + \mu_A N_{Ai} \frac{\partial}{\partial \sigma_{Ai}} \int f_{Ai}(\mathbf{r}) d\mathbf{r} \quad (16)$$

$$= \int \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \left( \frac{3}{\sigma_{Ai}} - \frac{|\mathbf{r} - \mathbf{R}_A|}{\sigma_{Ai}^2} \right) \rho_A^0(\mathbf{r}) d\mathbf{r}, \quad (17)$$

where we made use of  $\int f_{Ai}(\mathbf{r}) d\mathbf{r} = 1$ .

The AIM densities can be eliminated from the Lagrange equations (11) and (17) by making use of  $\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} = \frac{\rho(\mathbf{r})}{\rho_0(\mathbf{r})}$ . They can be rewritten in the following form:

$$N_{Ai} = \int \rho(\mathbf{r}) \frac{\rho_{Ai}^0(\mathbf{r})}{\rho_0(\mathbf{r})} d\mathbf{r}, \quad (18)$$

$$\sigma_{Ai} = \frac{1}{3N_{Ai}} \int \rho(\mathbf{r}) \frac{\rho_{Ai}^0(\mathbf{r})}{\rho_0(\mathbf{r})} |\mathbf{r} - \mathbf{R}_A| d\mathbf{r}. \quad (19)$$

These identities form the basis for the self-consistent algorithm that will be explained in the next subsection.

## 2.3 Self-consistent algorithm

Fig. 2 depicts a flow chart of the self-consistent algorithm discussed in this subsection. The individual steps are described in more detail below.

In order to find all the pro-atom parameters,  $\{N_{Ai}\}$  and  $\{\sigma_{Ai}\}$ , an initial guess is generated first, which will be refined later. Because the parameters  $\{\sigma_{Ai}\}$  are non-linear, it is not guaranteed that  $L_1$  is convex or has a unique minimum. Hence, multiple stationary points

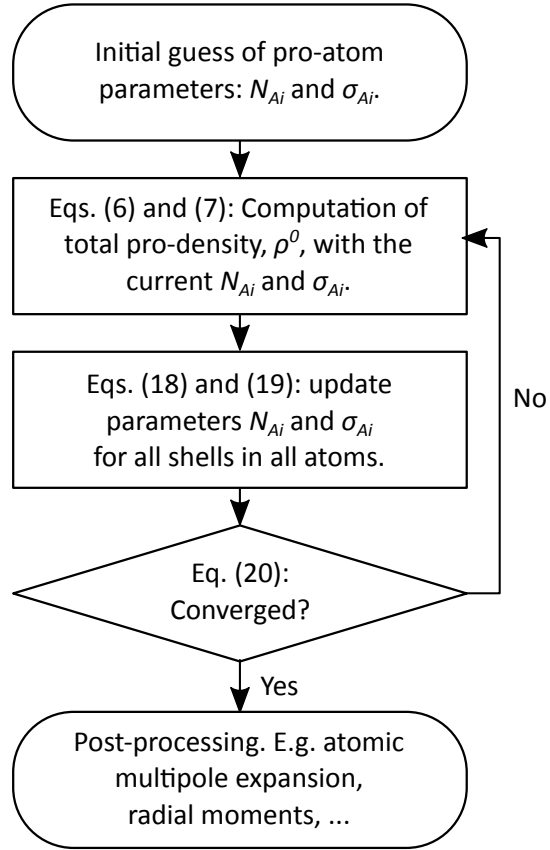


Figure 2: Flow chart of the self-consistent algorithm for the refinement of the MBIS parameters.

may exist and a reasonable initial guess is needed to find the solution of interest. The initial values of the parameters  $\{N_{Ai}\}$  of atom  $A$  are set to the number of electrons in each shell of the corresponding neutral isolated atom. The initial guess of  $\{\sigma_{Ai}\}$  is inspired by hydrogenic s-type orbitals. For the innermost and outermost shell of atom  $A$ , we take  $a_0/2Z_A$  and  $a_0/2$ , respectively, where  $Z_A$  is the atomic number. Initial values of  $\sigma_{Ai}$  for the intermediate shells are fixed by geometric interpolation:  $\sigma_{Ai} = a_0/2Z_A^{\left(1-\frac{i-1}{m_A-1}\right)}$ .

Given the initial guess, the parameters are refined iteratively with a self-consistent update. In a single iteration, Eqs. (18) and (19) are evaluated, using the “old” parameters in the right-hand side, yielding the “new” parameters in the left-hand side. These iterations are repeated until the pro-atom parameters no longer change significantly. In this work, the iterative algorithm (of MBIS and other Iterative Hirshfeld flavors) is stopped after the root-mean-square deviation between the pro-atom densities of the last and the previous iteration drops below a threshold of  $10^{-8}$  a.u:

$$\max_A \sqrt{\int d\mathbf{r} [\rho_{A,\text{last}}^0(\mathbf{r}) - \rho_{A,\text{previous}}^0(\mathbf{r})]^2} < 10^{-8}. \quad (20)$$

Other convergence criteria could be used as well, e.g. based on the gradient of the Lagrangian  $L_1$ .

It should also be possible to optimize the pro-atom parameters with a quasi-Newton optimizer. However, robust quasi-Newton optimizers that can handle various types of equality and inequality constraints (to fix the total population and to keep all parameters positive) are non-trivial. The algorithm sketched above satisfies all constraints at every iteration, is much easier to implement and converges smoothly, even with very tight convergence settings. The Levenberg-Marquardt algorithm is not applicable because it is specifically designed for least-squares objective functions, while MBIS uses the KL-divergence as the objective function.

The integrals in Eqs. (18) and (19) must be evaluated numerically. Because the integrands

are well localized on one atom, it is possible to implement the self-consistent update with a cost that scales linearly with the number of atoms.

## 2.4 Relevant pro-atom parameters for modeling electrostatic interactions with force fields

After the optimization of pro-atom parameters, one may reduce the pro-density to a simpler picture, which is suitable for force-field models. The nuclear charge and the Slater functions associated with core electrons can be condensed into a single effective *core charge*,  $q_{A,c}$ . This is illustrated in Fig. 1b. The remaining *valence* Slater function is characterized by two parameters, its *valence population*,  $N_{A,v}$ , and its *valence width*,  $\sigma_{A,v}$ . The net atomic charge,  $q_A = q_{A,c} - N_{A,v}$  can be used to approximate long-range electrostatic interactions. The two remaining degrees of freedom can be used to model the penetration effect,<sup>12,22–26</sup> i.e. the deviation of the electrostatic interactions from the simple point-charge model when the electronic densities begin to overlap. It can be computed efficiently with analytic expressions for the Coulomb interaction between Slater densities.<sup>25,32</sup>

## 2.5 MBIS Implementation

In the remainder of this work, MBIS will be tested extensively with applications to theoretical electron densities of molecules and condensed phases. In these applications, the all-electron density is first computed on an integration grid suitable for the numerical evaluation of Eqs. (18) and (19). The implementation of these numerical integrals differs significantly between isolated molecules and periodic systems.

For isolated molecules, all-electron densities are computed with Gaussian09<sup>62</sup>, using Density Functional Theory (DFT). Different functionals and Gaussian basis sets were used, as will be explained in the following sections. The MBIS partitioning of isolated molecule densities is carried out with HORTON 2.0.0,<sup>63</sup> which uses a standard atom-centered Becke-

Lebedev integration grid.<sup>64</sup> This implementation can be combined with any level of theory in Gaussian09 that produces an all-electron 1-particle reduced density matrix (1RDM) with the “density=current” option. Because Gaussian09 does not write out the 1RDM when relativistic corrections are used, our tests on isolated molecules are limited to molecules containing no elements heavier than krypton.

Electron densities of periodic crystals are computed with the Projector Augmented Wave (PAW) method<sup>65</sup> as implemented in GPAW-0.11.0.<sup>66–68</sup> Integrals involving the all-electron density of periodic systems are carried out as follows. In the PAW formalism, the total electron density is separated in a smoothly varying part, denoted as  $\tilde{\rho}(\mathbf{r})$ , and a correction for every atom  $A$  in the so-called augmentation sphere.<sup>65</sup>

In GPAW, the smooth density is represented on an equidistant real-space grid and the corrections are evaluated on atom-centered grids in spherical coordinates. This combination of integration grids makes it possible to perform very accurate integrations involving (reconstructed) all-electron densities of periodic systems. Our second MBIS implementation can handle any type of integration grid and we therefore used the same grid structure as in GPAW for periodic calculations. The advanced numerical techniques in this implementation, such as linear-scaling computational cost<sup>69</sup> and convergence acceleration, will be discussed in future work.

## 2.6 Relation to other partitioning methods

The MBIS pro-atom model has been used previously, however not yet in the context of Hirshfeld partitioning. For example, a similar pro-atom model (with fixed parameters) was also used in an ESP fitting scheme.<sup>70</sup> A similar density model is also used in the Stewart-Slater method.<sup>71</sup> Although our pro-atom model is obviously inspired by Slater’s work on atomic shielding constants,<sup>72</sup> the typical polynomial prefactors are omitted. This omission is inspired by the piece-wise exponential ansatz from statistical models for atomic densities.<sup>73–75</sup> The reduced model for force-field applications in subsection 2.4 has also been used before in

the development of force-field models.<sup>31,32,76</sup>

Our approach is comparable to the Iterative Stockholder (IS) method.<sup>53</sup> In IS, spherical pro-atoms are defined by generic radial functions without further restrictions in terms of density basis functions; in practice they are represented by function values on a radial grid. A self-consistent update, in the same spirit as Eqs. (18) and (19), guarantees that the optimal IS atoms minimize the KL-divergence over all possible spherically symmetric pro-atoms.<sup>53,77</sup> Even though this is a convex problem, a well-documented weakness of IS is that the density tails of the pro-atoms are ill-defined, which leads to numerical instabilities and poorly defined atomic charges.<sup>30,54,77</sup> In MBIS, this is resolved by modeling the density tail of each atom with only a single Slater function, which is comparable to the BS-ISA+DF method.<sup>30</sup> The MBIS self-consistent update algorithm is also very similar to the iterative Hirshfeld (HI) algorithm.<sup>46</sup> The main difference with HI is that MBIS makes use of an analytic ansatz for each pro-atom with several parameters per atom, i.e. the populations and widths of all shells in each atom, while HI varies just one population parameter per atom and makes use of pre-computed isolated atom densities. Furthermore, HI cannot be defined as a constrained optimization problem.<sup>35</sup>

It is also important to realize that density fitting<sup>78–80</sup> is closely related to MBIS. This connection becomes clear by considering the following Lagrangian:

$$L_2[\{N_{Ai}\}, \{\sigma_{Ai}\}, \mu] = \int \rho(\mathbf{r}) \ln \frac{\rho(\mathbf{r})}{\rho^0(\mathbf{r})} d\mathbf{r} + \mu \int \rho^0(\mathbf{r}) - \rho(\mathbf{r}) d\mathbf{r}. \quad (21)$$

The self-consistent update equations (18) and (19) can also be derived from  $L_2$ . This shows that the optimal MBIS pro-atom parameters can also be found by minimizing the KL-divergence of the pro-molecule density,  $\rho^0(\mathbf{r})$ , from a given molecular density,  $\rho(\mathbf{r})$ . This interpretation is similar to density fitting in force-field development,<sup>28,29</sup> except for the following two points. First, the MBIS pro-density is expanded in Slater functions while density-fitting techniques usually rely on contracted Gaussian functions, also with higher



multipoles.<sup>40</sup> Second, MBIS uses the KL-divergence as a cost function to fit the pro-atom parameters, while conventional density-fitting makes use of a least-squares cost, often with a Coulomb metric. The least-squares cost function was also used in other related works, e.g. the least-squares analog of IS is known as Stewart atoms<sup>81</sup> and the least-squares analog of MBIS is very similar to Stewart-Slater atoms.<sup>71</sup> Hybrid approaches, combining least-squares and KL-divergence cost functions, were also proposed, such as Hirshfeld-E,<sup>51</sup> Gaussian ISA<sup>54</sup> and BS-ISA+DF.<sup>30</sup> In the development of the MBIS method, a least-squares cost function was avoided because it was recently found to lead to non-local AIM densities.<sup>36</sup> Finally, note that the pro-atom parameters are sufficient to construct monopolar electrostatic force fields. We therefore expect that the direct optimization of the pro-atoms with a Lagrangian similar to Eq. (21), i.e. without constructing AIM densities, can be an attractive alternative to conventional AIM methods.

### 3 Example MBIS applications

This section provides two illustrative applications of the MBIS method. Their main purpose is to show the applicability of MBIS in very different scenarios and to provide the reader with some typical results. The first example discusses the robustness of MBIS and its compatibility with chemical intuition, when applied to rather extreme variations of the oxidation states of oxygen. The second example shows that MBIS is also sufficiently robust when studying subtle variations of the electron density of water between the gas, liquid and solid phase.

The examples below only illustrate the usefulness of the MBIS method. A more systematic assessment can be found in section 4.

#### 3.1 Oxygen in different oxidation states

In previous studies, Hirshfeld-I (HI) partitioning was criticized for its poor applicability to oxides.<sup>51,52,55</sup> During the iterative convergence of the charges, HI requires reference densities

for the oxygen dianion (or sometimes even trianion) in vacuum, which does not exist.<sup>55</sup> When the oxygen dianion is computed with a finite basis, one obtains a very diffuse density that is not representative for the oxygen atom in a molecule or crystal. This mismatch results in very large absolute values for the atomic charges in oxides, overestimating the polarity of oxide clusters or the electrostatic potential in solid oxides.<sup>52,55</sup> Many modifications of HI were proposed to surmount this limitation,<sup>51,55–57,82–84</sup> often using different (somewhat arbitrary) techniques for the computation of unstable anions. The MBIS method does not need (unstable) ion densities as input and one would therefore expect that it does not suffer from the same overpolarization issues as HI.

Table 1 compares MBIS and HI results for the oxygen element, in a series of systems where the oxidation state of oxygen varies from -2 to +3: MgO, chabazite, MIL-53(Al), quartz, H<sub>2</sub>O, CO<sub>2</sub>, H<sub>2</sub>O<sub>2</sub>, LiO<sub>2</sub>, O<sub>3</sub>, CO, O<sub>2</sub> and OF<sub>2</sub>. Results for other elements in these systems are given as well for the sake of completeness. Also the isolated oxygen cation, atom and anion are included because these have different valence electron densities that result in different parameters for the outer shell in the MBIS pro-density. Atom types, which are used to differentiate all non-equivalent atoms, are defined in section S1 in the supporting information. All electron densities are computed at the PBE level of theory. For isolated molecules, Gaussian09<sup>62</sup> was used with the 6-311+G(2df,p) basis and charges were computed with HORTON.<sup>63</sup> Electron densities of crystal unit cells were computed with GPAW,<sup>66–68</sup> using a grid spacing of 0.1 Å and charges were derived from the periodic densities with a second implementation of MBIS. (See subsection 2.5.)

The main trend in Table 1 is the strong correlation between HI and MBIS charges. For systems where HI was found to be useful for force-field development, MBIS gives very comparable results. However, when oxygen has an oxidation state of -2 and has (semi-ionic) bonds to cations with a high oxidation number, MBIS charges for oxygen are less negative, making them more suitable for force-field development.

A reasonable correlation between atomic charges and oxidation numbers is found. Such

Table 1: Comparison of HI and MBIS charges for a selection of molecules and solids. (See text.) Results are grouped per element and sorted by the MBIS charge within each group. ON stands for oxidation number. For the MBIS method, also the quantities from subsection 2.4 are reported: core charge ( $q_{A,c}$ ), valence charge ( $q_{A,v}$ ) and valence width ( $\sigma_{A,v}$ ).

Molecule or solid	Atom <sub>(type)</sub>	ON	HI	MBIS	$q_{A,c}$ [e]	$q_{A,v}$ [e]	$\sigma_{A,v}$ [Å]
			$q$ [e]	$q$ [e]			
MIL-53(Al)	H <sub>ph</sub>	0	0.132	0.155	1.0	−0.845	0.198
H <sub>2</sub> O <sub>2</sub>	H	1	0.387	0.414	1.0	−0.586	0.186
H <sub>2</sub> O	H	1	0.436	0.443	1.0	−0.557	0.187
MIL-53(Al)	H <sub>hy</sub>	1	0.565	0.519	1.0	−0.481	0.175
LiO <sub>2</sub> <sup>·</sup>	Li	1	0.912	0.825	1.076	−0.251	0.387
MIL-53(Al)	C <sub>pc</sub>	0	−0.132	−0.148	4.359	−4.507	0.266
MIL-53(Al)	C <sub>ph</sub>	0	−0.087	−0.110	4.354	−4.463	0.266
CO	C	3	0.144	0.108	4.327	−4.218	0.270
MIL-53(Al)	C <sub>ca</sub>	4	0.910	0.849	4.340	−3.491	0.246
CO <sub>2</sub>	C	4	0.847	0.863	4.340	−3.477	0.243
MgO	O	−2	−2.220	−1.934	6.243	−8.177	0.247
Chabazite	O <sub>b</sub>	−2	−1.497	−1.261	6.335	−7.596	0.222
Chabazite	O <sub>r</sub>	−2	−1.480	−1.250	6.337	−7.588	0.222
MIL-53(Al)	O <sub>hy</sub>	−2	−1.952	−1.220	6.329	−7.548	0.223
Chabazite	O <sub>x</sub>	−2	−1.431	−1.219	6.344	−7.563	0.220
Quartz	O	−2	−1.473	−1.213	6.343	−7.555	0.221
O <sup>−</sup>	O	−1	−1.0	−1.0	6.194	−7.194	0.246
H <sub>2</sub> O	O	−2	−0.872	−0.885	6.333	−7.219	0.220
MIL-53(Al)	O <sub>ca</sub>	−2	−0.781	−0.748	6.354	−7.103	0.214
CO <sub>2</sub>	O	−2	−0.424	−0.431	6.380	−6.811	0.208
H <sub>2</sub> O <sub>2</sub>	O	−1	−0.387	−0.414	6.351	−6.765	0.212
LiO <sub>2</sub> <sup>·</sup>	O	−1/2	−0.456	−0.412	6.330	−6.743	0.215
O <sub>3</sub>	O <sub>t</sub>	0	−0.194	−0.177	6.364	−6.541	0.207
CO	O	−3	−0.144	−0.108	6.398	−6.506	0.201
O	O	0	0.0	0.0	6.348	−6.348	0.207
O <sub>2</sub>	O	0	0.0	0.0	6.371	−6.371	0.203
OF <sub>2</sub>	O	2	0.156	0.121	6.367	−6.247	0.203
O <sub>3</sub>	O <sub>c</sub>	0	0.389	0.354	6.386	−6.032	0.197
O <sup>+</sup>	O	1	1.0	1.0	6.431	−5.431	0.182
OF <sub>2</sub>	F	−1	−0.078	−0.060	7.381	−7.441	0.182
MgO	Mg	2	2.220	1.934	10.551	−8.617	0.117
MIL-53(Al)	Al	3	2.780	2.111	3.215	−1.104	0.352
Quartz	Si	4	2.946	2.425	4.425	−2.000	0.314
Chabazite	Si	4	2.944	2.490	4.395	−1.905	0.316

correlations are not expected to be perfect because the oxidation number is based on simple counting rules that do not account for the (partial) covalent character of chemical bonds. It may be surprising that the core charge,  $q_{A,c}$ , is systematically larger than the integer value one would get by combining the nuclear charge and an integer number of core electrons. Because the valence Slater function does not decay toward the nucleus, it also contributes to the core region, which is compensated by a slightly more positive core charge. All the variations in the net charge are reflected in the valence charge,  $q_{A,v}$ . The valence width,  $\sigma_{A,v}$ , linearly correlates with the net charge: within each group of a given element, more negative atoms tend to have a slightly larger valence width.

### 3.2 Application of MBIS to the three phases of water

The MBIS method will first be illustrated with an application to an isolated water molecule, 38 clusters of water molecules, a model for the hexagonal phase of ice and 10 snapshots of a liquid water MD simulation. The electron densities of the isolated systems were computed with Gaussian09<sup>62</sup> at the BLYP/6-311+G(2df,p) level of theory.<sup>85,86</sup> The 38 clusters, ranging from 2 to 10 water molecules in size, were taken from the work of Temelso *et al.*<sup>87</sup> We used the 3x3x2 model for the ice-1h phase of water from the work of Hayward and Reimers.<sup>88</sup> This model contains 96 water molecules and is tuned for computational applications: the water molecules have realistic randomized orientations, yet the net dipole moment of the unit cell is constrained to zero. The geometry of the ice-1h model is refined with CP2K-2.6.0<sup>89-92</sup> at the BLYP-D3 level of theory<sup>64,85,93,94</sup> using the MOLOPT-DZVP-SR-GTH<sup>95</sup> basis set and GTH pseudopotentials.<sup>96,97</sup> CP2K was also used to generate periodic structures of liquid water (32 molecules per unit cell). These geometries were sampled every 10 ps from a 100 ps NVT<sup>98</sup> molecular dynamics run at 300 K and at the experimental density, using the same level of theory. The electron densities of all periodic structures (ice and liquid water snapshots) were computed with GPAW<sup>66-68</sup> using the BLYP functional and a grid spacing of 0.1 Å, as explained above. (CP2K was not used for this purpose because it cannot print

out a reconstructed all-electron density on suitable integration grids.)

Figure 3 displays the key MBIS results for the water systems in this section. The partitioning of the density into atomic contributions,  $\rho_A(\mathbf{r})$ , is first used to construct electron densities of separate water molecules from which multipoles can be derived (relative to the molecular center of mass). The most obvious result is the increase of the molecular dipole moment as water forms hydrogen bonds with surrounding molecules (Figure 3a). Similar trends are usually found in simulations of water with polarizable force fields.<sup>99–102</sup> This increase is seen throughout all water clusters and the solid ice 1h phase. Liquid water exhibits relatively large random fluctuations in the molecular dipole moment, due to variations in the water geometry and its local environment. Another clear trend is that water molecules tend to exchange a small fraction of an electron with their surrounding, leading to non-zero molecular charges (Figure 3b).

Atomic charges and dipole moments are directly derived from the AIM densities,  $\rho_A(\mathbf{r})$ . The increased polarization of water in larger clusters (Figure 3c) is due to the decrease of the (negative) oxygen charge, while the norm of the oxygen dipole moment follows the opposite trend (Figure 3d), slightly reducing the overall polarization. Hydrogen atoms have a small and constant dipole moment, showing that they are only weakly polarizable.

Because the MBIS pro-molecular density is a sum of spherical atoms, it was to be expected that AIM dipole fluctuations play a minor role compared to atomic and molecular charge fluctuations. In general, the partitioning of the total polarization into contributions from atomic charges and/or dipoles is inherently ambiguous and can depend strongly on the AIM method. However, Mei *et al.* observed, for a large set of molecules and for all AIM methods tested in their work, that the overall polarization always involves a significant amount of charge fluctuations.<sup>103</sup>

MBIS pro-atom parameters reveal additional trends that are not easily observed with other methods. The *valence width*,  $\sigma_{A,v}$  as introduced in subsection 2.4, is also sensitive to the molecular environment, which is most notable for the hydrogen atoms while oxygen has

a more constant valence width (Figure 3e). Polarizable force fields usually consider fluctuating atomic charges and/or dipoles, but fluctuations in the width of the atomic electron distribution are rarely included. Our results indicate that these may also be relevant to model electronic polarization.

Finally, Figure 3f shows the *atomic core charge*,  $q_{A,c}$  defined in subsection 2.4. This quantity varies relatively little, in line with the expectation that the properties of core electrons should be transferable. There is a small but notable difference between the core charge for Gaussian09 (1-10 H<sub>2</sub>O) and GPAW (Ice 1h, Liquid) calculations. GPAW calculations on the isolated clusters confirm that this is due to the different treatment of the core electrons in both programs (results not shown).

In general, the MBIS results for different phases of water show that the method is robust enough to uncover several subtle trends in the electronic polarization, which is very helpful for the interpretation of these trends and the construction of polarizable force fields.

## 4 Systematic comparison to other AIM methods

In this section, the MBIS method is compared to a series of other AIM methods, using several molecular datasets. This assessment focuses on properties that are relevant for modeling electrostatic interactions in force fields: the quality of the electrostatic potential (ESP), the accuracy of electrostatic interactions and the robustness of the charges.

### 4.1 Molecular datasets

Several datasets of molecular dimers and isolated molecules are considered in this section: five were taken from the literature and two new datasets are introduced below. All molecular electron densities were computed at the B3LYP/6-311+G(2df,p) level of theory with Gaussian09.<sup>62</sup>

Three sets of molecular dimers were taken from the work of Hobza et al., namely S66

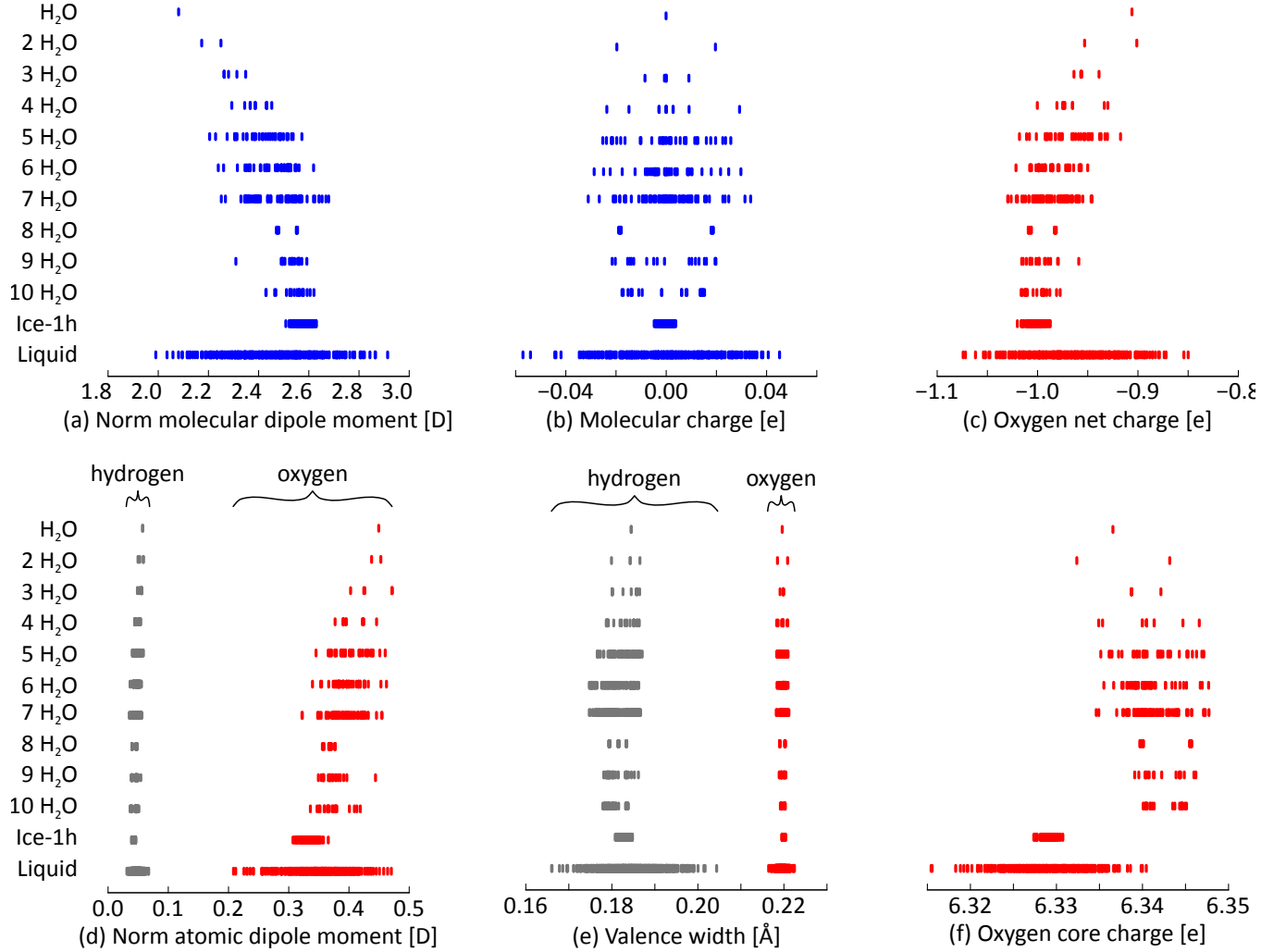


Figure 3: MBIS results for different phases of water (isolated water molecule, water clusters, ice 1h and liquid water) derived from all-electron densities computed at the BLYP level of theory: (a) the norm of the molecular dipole moment of each water molecule (b) the net charge of each water molecule (c) the oxygen charge of each water molecule, (d) the norm of each atomic dipole moment, (e) the valence width of each atom ( $\sigma_{A,v}$ , see subsection 2.4), (f) the atomic core charge of each oxygen atom ( $q_{A,c}$ , see subsection 2.4).

(diverse non-covalent interactions between neutral organic molecules),<sup>104</sup> IHB15 (ionic hydrogen bonds)<sup>105</sup> and X40 (halogen bonds).<sup>106</sup> From the X40 set, dimers containing iodine were omitted because proper all-electron densities for such heavy elements can only be computed with relativistic corrections. (See subsection 2.5.) Also, a new set of molecular dimers is introduced, i.e. ZG237, a set of 237 dimers between silica clusters and typical guest/template molecules for porous media. Neutral and anionic silica clusters are present in ZG237 and the guest molecules include noble gases, neutral and cationic organic molecules. (More details are provided in section S2 of the supporting information.) The goal of this assessment with molecular dimers is to test how well atomic charges obtained with different AIM methods can reproduce the electrostatic interaction.

Three datasets of larger isolated molecules are also used in the tests below, of which two were taken from earlier work: PENTA103 (103 random penta-alanine conformers)<sup>49</sup> and SILICA245 (topologically different hydrogen-terminated silica clusters containing up to 8 Si atoms).<sup>51</sup> One new set, MIL53(M)10 was created based on our experience with the development of a flexible force field for the metal-organic framework MIL-53(AL).<sup>6</sup> This set contains ten organometallic clusters with the same structure, see Fig. 4, but with different metals in oxidation state III: Al, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Ga. The spin multiplicity of each cluster was fixed by coupling the spins of the transition metals to obtain a maximal  $\langle \hat{S}_z \rangle$  value. The  $\mu$ -OH group is located at the center and the cluster is carefully terminated by four malondialdehyde anions and one formic acid anion. This neutral configuration is stable for many first-row transition metals (excluding Cu and Zn) and it resembles well the metal-oxide structure found in the MIL-53 framework.<sup>107</sup>

## 4.2 Selection of AIM methods

Three categories of AIM methods are used for comparison: ESP-fitted charges, density partitioning methods (Hirshfeld variants and QTAIM) and Hilbert-space partitioning methods.

Atomic charges fitted to the electrostatic potential (ESP) are among the most ubiquitous



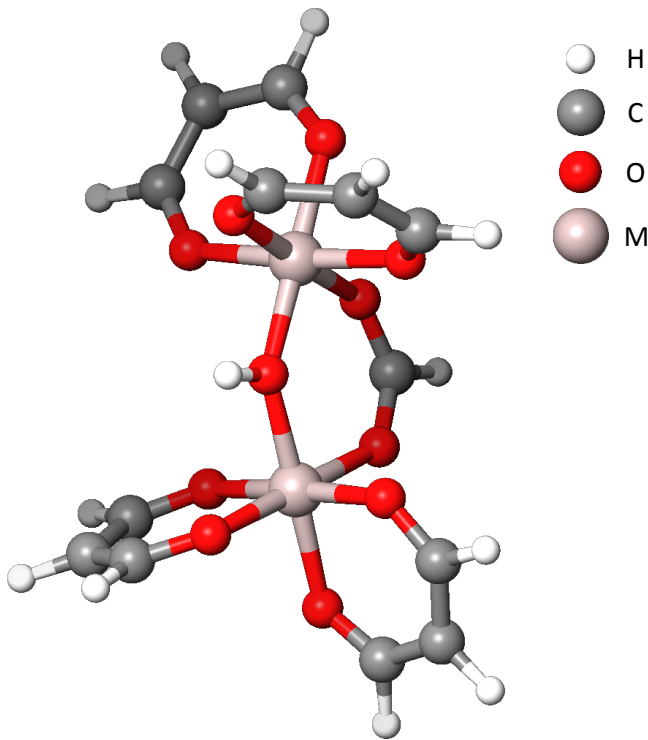


Figure 4: Metal-oxide cluster where M is a metal atom in oxidation state III (Al, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Ga).

for the development of force field models. We selected four such variants: Merz-Singh-Kollman (MSK),<sup>108</sup> CHelpG,<sup>109</sup> Restrained ESP (RESP)<sup>20</sup> and Hu-Lu-Yang (HLY).<sup>70</sup> MSK, CHelpG and HLY differ in the way the volume around the molecule is sampled in the fitting procedure but they all make use of a standard least-squares procedure. The RESP method extends the MSK cost function with hyperbolic restraints to penalize large absolute atomic charges.

A large selection of Hirshfeld variants is used in our comparison, starting with the original Hirshfeld method (H).<sup>38</sup> CM5 is a popular empirical correction to the Hirshfeld method to better reproduce experimental dipole moments of small molecules.<sup>110</sup> Another popular related method is Hirshfeld-I (HI).<sup>46</sup> While the Hirshfeld method simply uses spherically averaged neutral atoms as pro-atoms, HI iteratively updates the pro-atoms to enforce consistency between the AIM charges and the charges of the pro-atoms. Charged pro-atoms are constructed by a linear interpolation between spherically averaged densities of isolated neu-

tral atoms and ions. A particular improvement of HI over Hirshfeld is that HI charges make a good estimate of the electrostatic potential of organic molecules.<sup>47</sup> This is no longer the case for metal oxides, e.g. the ESP in the pores of zeolites, which inspired several groups to further improve the method, leading to variants such as Hirshfeld-E<sup>51</sup> (HE) and DDEC4.<sup>56</sup> The Iterative Stockholder (IS) analysis is another variant of the Hirshfeld method, proposed independently of HI but with many similarities.<sup>37</sup> IS also iteratively updates its pro-atoms but just uses the spherical averages of the AIM densities from the previous iteration as the new pro-atoms. Besides all the Hirshfeld variants mentioned so far (H, CM5, HI, HE, DDEC4 and IS), one more density-based method, QTAIM,<sup>27</sup> is also included in the comparison.

The third group of methods are Hilbert-space methods, which partition the density matrix instead of the density. Mulliken (M) is the oldest AIM method<sup>111</sup> and two popular improvements of this scheme are also widely used: Löwdin (L)<sup>112</sup> and Natural (N) charges.<sup>113</sup> These three Hilbert-space methods assume that each orbital basis function is centered on one of the atoms, which is always the case when using standard Gaussian basis sets. However, in several popular periodic DFT codes, e.g. VASP, CPMD and GPAW, such information is not available, because they use delocalized basis sets.

Results for methods MSK, CHELPG, M, L and N were obtained with Gaussian09.<sup>62</sup> The MSK or CHELPG methods make use of van der Waals radii but do not define them for all elements, in which case UFF radii were used instead. Gaussian formatted checkpoint files were used to post-process the densities with HORTON-2.0.0<sup>63</sup> to compute the HLY, H, HI, HE, IS and MBIS charges. The RESP program from the Antechamber program<sup>114</sup> was used to compute the RESP charges. DDEC4 charges were computed with Chargemol-09.15.2014,<sup>115</sup> QTAIM charges with AIMAll-11.06.19.<sup>116</sup> Practically all MBIS results below are obtained from the pro-atom parameters discussed in section 2.4. MBIS AIM densities as such are not used unless noted otherwise.

### 4.3 Quality of the electrostatic potential (ESP)

In the context of force-field development, it is assumed that whenever atomic charges accurately reproduce the electrostatic potential (ESP) around a molecule, they also make good predictions of the electrostatic interactions.<sup>19</sup> Hence, one of the desirable properties of atomic charges is their ability to reproduce the ESP as well as possible, which is the topic of this subsection. A direct assessment of the quality of electrostatic interactions is discussed in subsection 4.4.

We have tested the quality of the ESP for all sets of isolated molecules discussed in section 4.1 and also for the monomers present in all dimer datasets. The HLY ESP cost function is used to measure the quality of the ESP and the results would not change much if we had used an MSK or CHELPG cost function instead. The HLY cost function is an integration over a volume surrounding the molecule,<sup>70</sup> which we converted to an RMSE value as follows:

$$\text{RMSE}_{\text{ESP}} = \sqrt{\frac{\int w_{\text{HLY}}(\mathbf{r}) \left( V_{\text{DFT}}(\mathbf{r}) - \sum_A \frac{q_A}{4\pi\epsilon_0|\mathbf{r}-\mathbf{R}_A|} \right)^2 d\mathbf{r}}{\int w_{\text{HLY}}(\mathbf{r}) d\mathbf{r}}} \quad (22)$$

where  $w_{\text{HLY}}(\mathbf{r})$  is the weight function designed by Hu, Lu and Yang: it becomes one in the region surrounding a molecule and goes smoothly to zero inside the molecule and at larger distances.<sup>70</sup>  $V_{\text{DFT}}(\mathbf{r})$  is the reference ESP from the DFT calculation and  $q_A$  are the atomic charges. The smoothness of the weight function guarantees that the ESP cost is not sensitive to the exact position of the grid points, which is a clear advantage over other ESP fitting methods.  $\text{RMSE}_{\text{ESP}}$ , which we computed for every molecule and every AIM method, is a measure for the error on the frozen-density interaction energy<sup>10-12</sup> of a unit charge with the molecule when it is placed near its van der Waals surface.

Fig. 5 compares for every AIM method the average of  $\text{RMSE}_{\text{ESP}}$  within five groups of isolated molecules: PENTA103 and MIL53(M)10 are those discussed in section 4.1. SILICA contains all those of the SILICA245 set plus all silica clusters from the ZG237 set. X40HMONO contains all halogenide molecules present in the X40 set of dimers. Fi-

nally, ORGANIC contains all other monomers from the dimer sets S66, IHB15, X40 and ZG237. (The noble gas atoms from ZG237 are not included.)












































































AIM Method		PENTA103	MIL53(M)10	SILICA	X40HMONO	ORGANIC
ESP fitting	MSK	4.3 	3.7 	6.2 	9.5 	5.4 
	RESP	33 	11 	13 	11 	8.7 
	CHELPG	5.2 	6.2 	5.1 	10 	6.5 
	HLY	4.2 	3.7 	4.5 	9.4 	5.3 
Density-based Hirshfeld variants	H	16 	17 	8.6 	17 	18 
	CM5	15 	13 	24 	12 	10 
	HI	7.3 	8.8 	23 	14 	7.8 
	DDEC4	7.3 	5.9 	12 	12 	7.2 
	HE	9.9 	14 	13 	13 	10 
	IS	7.9 	5.3 	9.0 	13 	6.8 
	MBIS	8.3 	7.3 	13 	12 	7.3 
	QTAIM	49 	101 	39 	37 	27 
Hilbert space partitioning	M	28 	36 	10 	18 	13 
	L	42 	36 	17 	38 	22 
	N	22 	44 	34 	22 	13 

Figure 5:  $\text{RMSE}_{\text{ESP}}$  in  $\text{kJ mol}^{-1}$  computed for all AIM methods in this work, averaged over groups of isolated molecules. (See text for definition of groups).

Obviously, the ESP-fitted charges (MSK, RESP, CHELPG and HLY) perform well in this test as they are optimized to reproduce the ESP surrounding each molecule. RESP charges are not as optimal as the other three because of the hyperbolic restraints, which becomes very pronounced for large molecules as in the PENTA set. Lowering the strength of the restraints could relieve this issue but it would also result in less robust charges. The halogenides in X40HMONO have an ESP that is relatively difficult to reproduce with point charges: the sigma-hole of the halogen atom corresponds to a large and local dipole moment whose effect on the ESP cannot be explained in terms of atomic monopoles.<sup>58,117</sup>

The original Hirshfeld method (H) usually predicts poor ESPs because the absolute values of the atomic charges are too low.<sup>45</sup> All variants of the Hirshfeld method (HI, CM5, DDEC4, HE, IS and MBIS) produce more accurate ESPs, except for CM5 and HI when tested with the SILICA set. The good performance of the IS method is not surprising: it partitions the electron density in AIMS that are as spherical as possible, sometimes by introducing

an unreasonable radial dependence,<sup>48</sup> thus having small atomic dipole and higher multipole moments. In fact, any method that performs better than IS likely biases the charges to mimic effects of atomic multipoles. Such overfitting clearly occurs in the ESP-fitting methods. In case of MBIS, we only considered MBIS point charges and not the more advanced model with valence Slater functions, see Fig. 1b, simply because Eq. (22) only tests the ESP outside the molecule where the density is very low. In this region, the ESP generated by the Slater functions is very well approximated by that of point charges. Of all Hirshfeld variants, DDEC4, IS and MBIS are comparably good.

In line with previous observations, QTAIM charges are inadequate for the purpose of modeling ESPs.<sup>51</sup> This can only be fixed by including higher atomic QTAIM multipoles, as is often done in QTAIM-based force fields.<sup>34</sup> Mulliken (M), Löwdin (L) and Natural (N) charges rarely produce useful ESPs in our tests.

Fig. 6 shows ESP maps plotted on the  $\rho = 0.002$  a.u. isosurface of two representative molecules: methylacetamide and methylsilanetriol, which have an electrostatic potential that is respectively easy and difficult to reproduce with point charges. The isosurface approximates the molecular van der Waals surface,<sup>118</sup> which is convenient for visualizing non-covalent interactions. In addition to the ESP of the DFT calculation, the ESPs obtained with a subset of atomic charge methods, and their deviation from the DFT result, are shown. The isosurfaces sample the ESP at a higher density than the HLY cost function, which has two important consequences. First, the scale of the ESP deviations is large compared to the reported  $\text{RMSE}_{\text{ESP}}$  values. Second, the ESP maps are different for MBIS point charges and MBIS core charges with delocalized valence shells, the latter accounting for the penetration effect.

The main observation is that all model ESPs qualitatively agree with the DFT result. (See left column in Fig. 6a and Fig. 6b.) Some quantitative differences are present but they only appear clearly in the isosurfaces on which the deviations from the DFT ESP are shown. (See right column in Fig. 6a and Fig. 6b.) Even though RESP charges have a relatively low

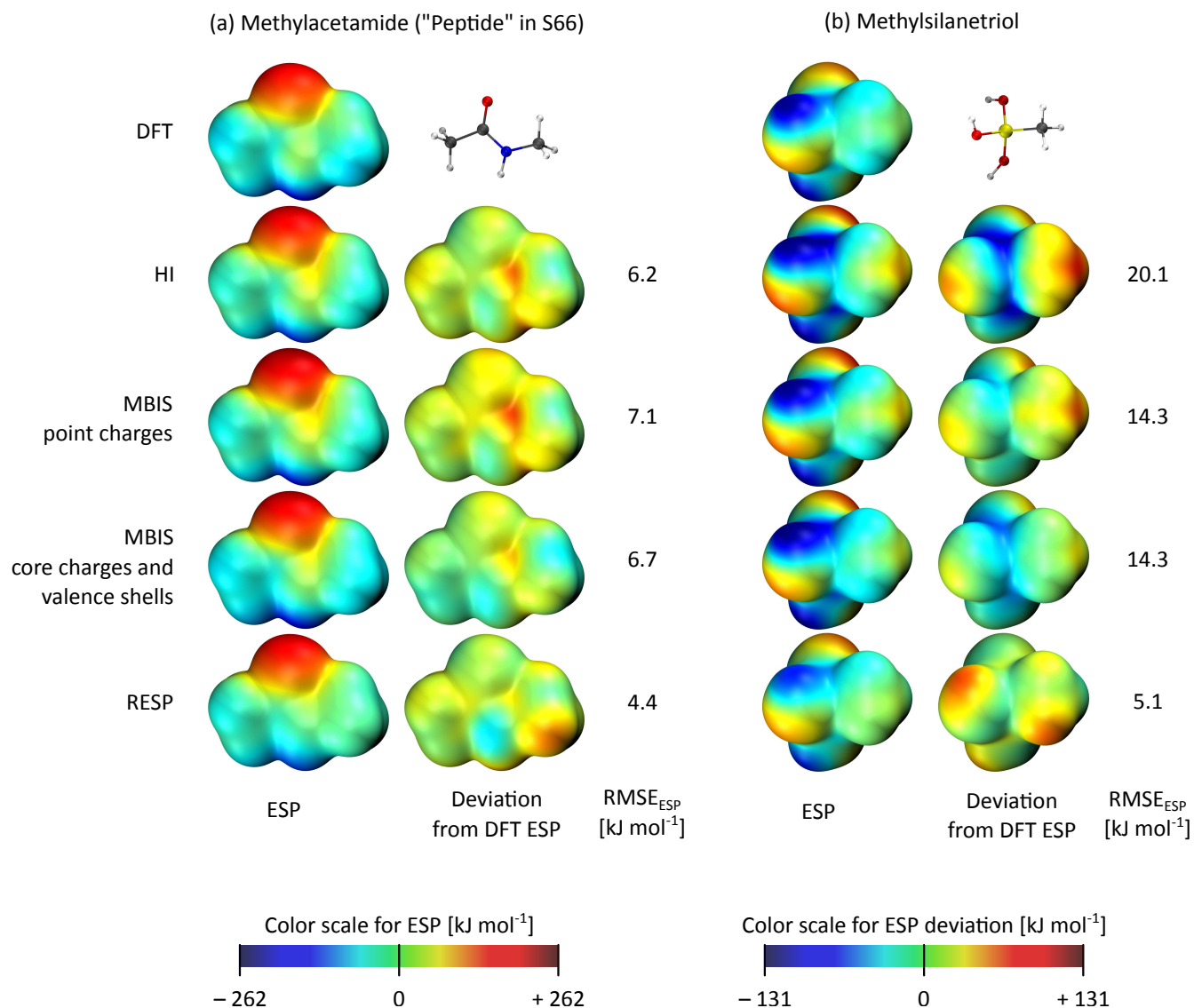


Figure 6: ESP maps plotted on the  $\rho = 0.002$  a.u. isosurface of (a) methylacetamide and (b) methylsilanetriol. In the left column for each molecule, the DFT ESP and the ESP due to selected point charge methods (HI, MBIS and RESP) and the core charge + valence shell (MBIS) are shown. For each approximate ESP, the deviation from the DFT reference is shown in the right column, for each molecule.

$\text{RMSE}_{\text{ESP}}$ , the deviations from the DFT ESP are not significantly smaller than for the other methods. When MBIS core charges and valence shells are used to estimate the ESP, a better visual agreement is found, because the penetration effect is already significant at the selected isodensity surface. Also note that the ESP obtained with HI charges for methylsilanetriol deviates the most from the DFT reference, in line with the limitations of HI for oxides, which were also discussed in subsection 3.1 and which is also seen in Fig. 5. Finally, note that this visualization of two representative molecules merely serves as an illustration. Solid conclusions can only be drawn from a thorough statistical analysis involving many molecules, such as the one presented in Fig. 5.

#### 4.4 Accuracy of electrostatic interactions

A common assumption in force-field development is that ESP-fitted charges also reproduce electrostatic interaction energies in general. Here, we assess the validity of this assumption for molecular dimers: the electrostatic interaction in the frozen-density approximation,<sup>10–12</sup>  $E_{\text{FD}}$ , will be used as a reference to test approximate electrostatic interactions obtained with atomic point charges from different AIM methods. The frozen-density approximation does not include any effects from polarization or charge-transfer. Such effects should be modeled with a polarizable (or polarized) force field, which is beyond the scope of this test.

The four sets of molecular dimers described in section 4.1 (S66, IHB15, X40 and ZG237) cover a large variety of electrostatic interactions, from as little as  $-0.06 \text{ kJ mol}^{-1}$  to rather extreme values of  $-664 \text{ kJ mol}^{-1}$ . Especially in the ZG237 set, it is often hard to classify dimers into specific interaction types, like hydrogen bonding, salt bridge, etc. To facilitate the interpretation of the results, we have classified the dimers more conveniently, just using thresholds on the strength of the electrostatic interaction in the frozen-density approximation: “Weak” ( $E_{\text{FD}} > -10 \text{ kJ mol}^{-1}$ ), “Medium” ( $-10 \text{ kJ mol}^{-1} \geq E_{\text{FD}} > -50 \text{ kJ mol}^{-1}$ ) and “Strong” ( $E_{\text{FD}} \leq -50 \text{ kJ mol}^{-1}$ ). Fig. 7a shows the numbers of dimers from each dataset in each class. The IHB15 set contributes exclusively to the “Strong” class while all other sets

have dimers in each class of interaction strength. Fig. 7b shows the root-mean-square error (RMSE) on the electrostatic interaction energy for each of the three classes and for each AIM method. The label “MBIS-S” refers to the interaction energy computed using effective core charges and valence Slater density functions, as shown schematically in Fig. 1b.

(a) Number of dimers from each dataset per category

Dataset	Weak	Medium	Strong
<b>S66</b>	16	39	11
<b>IHB15</b>	0	0	15
<b>X40</b>	13	13	5
<b>ZG237</b>	63	83	91

(b) RMSE on the electrostatic interaction



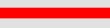





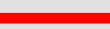


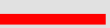




































AIM Method	Weak	Medium	Strong
<b>MSK</b>	5.9 	15 	41 
<b>RESP</b>	5.9 	17 	51 
<b>CHELPG</b>	5.9 	17 	43 
<b>HLY</b>	5.8 	14 	40 
<b>H</b>	6.1 	25 	87 
<b>CM5</b>	6.0 	20 	54 
<b>HI</b>	5.5 	15 	35 
<b>DDEC4</b>	5.8 	15 	36 
<b>HE</b>	5.9 	20 	47 
<b>IS</b>	5.9 	14 	31 
<b>MBIS</b>	5.7 	13 	30 
<b>MBIS-S</b>	2.5 	4.9 	30 
<b>QTAIM</b>	7.2 	22 	39 
<b>M</b>	5.8 	23 	61 
<b>L</b>	6.2 	26 	89 
<b>N</b>	5.9 	11 	29 

Figure 7: The accuracy of the electrostatic interactions for three classes of electrostatic interactions: Weak ( $E_{\text{FD}} > -10 \text{ kJ mol}^{-1}$ ), Medium electrostatics ( $-10 \text{ kJ mol}^{-1} \geq E_{\text{FD}} > -50 \text{ kJ mol}^{-1}$ ) and Strong ( $E_{\text{FD}} \leq -50 \text{ kJ mol}^{-1}$ ). Part (a) shows the number of dimers each dataset contributes to each class. Part (b) shows the RMSE of the electrostatic interaction in  $\text{kJ mol}^{-1}$ , obtained with atomic charges from different AIM methods, for each group. MBIS-S goes beyond the simple point-charge model and uses spherical valence Slater functions to model the penetration effect.

A surprising result in Fig. 7 is that electrostatic interactions computed with ESP-fitted charges are not the most accurate. MBIS point charges perform better in all three classes (Weak, Medium and Strong) than the best ESP-fitting method (HLY). This can be under-



stood as follows: as explained in subsection 4.3, ESP-fitted charges are biased to reproduce effects of atomic multipoles on the ESP. Although this may improve the accuracy with which the ESP is reproduced, it is a form of overfitting that may deteriorate other results obtained with ESP-fitted charges, as we observe here.

The second important result is that the MBIS method is a very effective model for the penetration effect. The MBIS-S results in Fig. 7 are obtained by describing every atom with an effective core charge and a valence Slater function. Even though this is a very simple (and thus computationally efficient) approach, it already reduces the RMSE by more than 50% in the classes “Weak” and “Medium”. Only for “Strong” electrostatic interactions, there are no apparent benefits from using such Slater density functions. A more detailed analysis, in which we computed electrostatic interactions with multipole expansions of MBIS AIM densities, showed that the largest error on the “Strong” electrostatic interactions is due to the neglect of atomic dipole moments.

## 4.5 Robustness of the atomic charges

For the development of an electrostatic force-field model or for the chemical interpretation of atomic charges, it is desirable that the charges are robust, i.e. not too sensitive to small details in the electronic structure calculations from which they are derived. Robustness is a prerequisite for transferability, i.e. the assumption that parameters derived from a molecule remain valid when that molecule is embedded (non-)covalently in a molecular environment. Even for environment-specific force-field parameters,<sup>58,119</sup> a robust partitioning is of interest to assure that such parameters remain valid as far as possible from the reference point for which they were computed. In this subsection, three kinds of sensitivity (the inverse of robustness) of atomic charges are investigated: sensitivity to conformational changes, to chemical changes in the environment and to changes in the basis set.

Fig. 8a shows the sensitivity of the atomic charges to conformational changes of the penta-alanine chain, for all AIM methods. For a given AIM method, the standard deviation

of the atomic charges in the PENTA103 set are computed with respect to the average charge of each atom over all 103 conformations. These 103 conformers are randomly generated meta-stable structures.<sup>49</sup> Although some fluctuation of the charges may be expected due to internal polarization, some methodological artifacts will cause larger fluctuations without physical origin.

Fig. 8b compares an averaged standard deviation of the Si charges in the SILICA245 set. The Si atoms are divided into three groups, based on the number of terminating hydrogen atoms they are bonded to (ranging from 1 to 3). Within each group, the standard deviation on the Si charge is computed and the average over the three groups is shown in Fig. 8b. This standard deviation should be small because the Si atoms within one group have a very comparable chemical environment.

Finally, Fig. 8c shows the sensitivity of the metal atom charge in the MIL53(M)10 clusters to the basis set. Their electron densities were computed with 6-311+G(2dp,f), 6-311+G\*, 6-31+G\*, 6-31+G or 6-31G\*. The standard deviation is computed relative to the average charge of each transition metal over all basis sets. Note that the sensitivity values for Mulliken and Löwdin fall literally off the chart and the corresponding bars in the bar plot were truncated for the sake of clarity.

The ESP-fitted and Mulliken charges have a very high sensitivity in all three cases, in line with earlier work.<sup>49</sup> This is problematic because it is almost impossible to provide definitive charges with such methods. The basis set sensitivity seems to be the most difficult to control: the standard deviation is larger than 0.1 e for the methods MSK, RESP, CHELPG, HLY, HI, HE, and M, L and N.

The basis set robustness of H, HI and HE can be improved as follows. Currently we used consistent levels of theory for pro-atom and molecular electron densities. If the pro-atoms were computed with a single level of theory and basis set, independent of the settings of the molecular calculation, the robustness would significantly improve. This is noticeable in the low sensitivity of the DDEC4 and especially the CM5 charges. Both CM5 and DDEC are

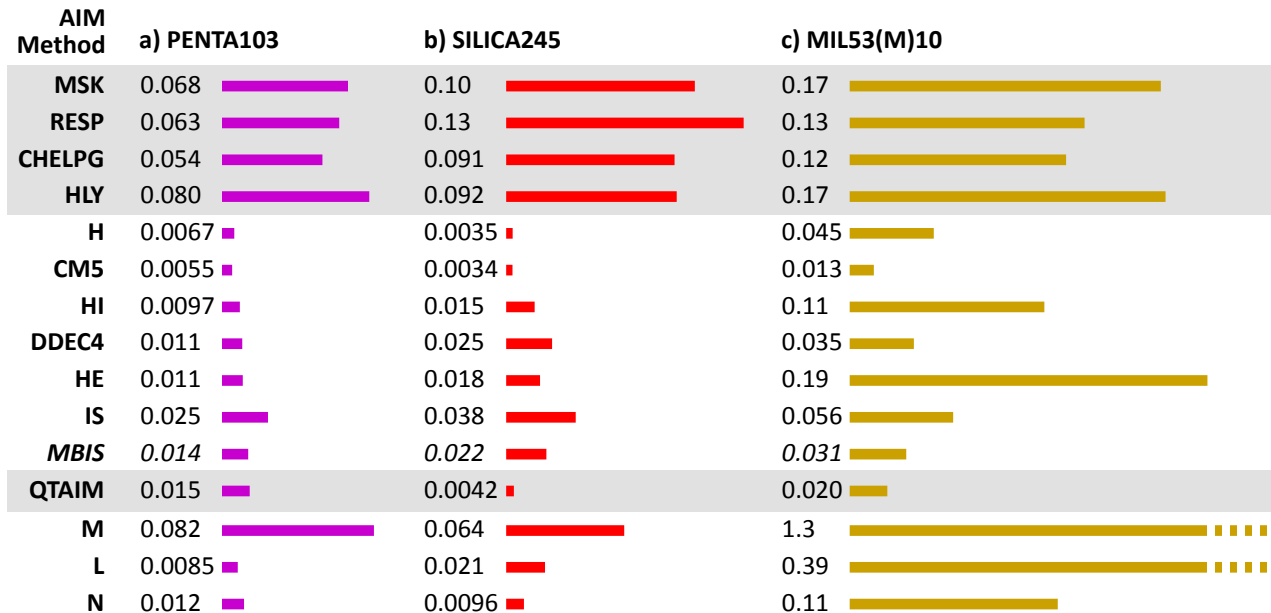


Figure 8: Standard deviations on atomic charges in e, within different sets of molecules. (a) The fluctuation on the atomic charges of 103 penta-alanine conformers. (b) The fluctuation on three different types of Si charges in a set of 245 silica clusters. (See text for definition of groups of Si atoms.) (c) The fluctuation on the metal charge in the MIL53(M)10 clusters, due to changes in basis set.

implemented with a unique set of pro-atoms.

The MBIS charges are more robust than the IS charges. This is simply because MBIS pro-atoms have fewer degrees of freedom than IS pro-atoms, in line with previous observations.<sup>54</sup>

The restraints in the RESP method only have a marginal impact on the robustness compared to MSK (the same method without restraints), showing that the restraints do not meet their purpose while they may cause a very poor fit to the ESP (see Fig. 5). The HLV method, a rather recent ESP-fitting method with a more carefully constructed cost function, does not guarantee robust results either.

QTAIM charges are among the most robust in our test, which is in line with previous studies assessing the transferability of QTAIM results.<sup>34,120</sup> Still, several Hirshfeld variants, such as CM5, MBIS and DDEC4 are comparably robust.

## 4.6 Pareto analysis

In our comparative analysis, we have considered three main criteria that atomic charges should meet for the development of force fields: accuracy of the ESP, accuracy of electrostatic interactions and robustness. Ideally, an AIM method should combine all these features, especially the last two. The Pareto plots in Fig. 9 visualize the trade-offs between different criteria discussed in the previous subsections. It is unavoidable that some subjective choices slightly affect the Pareto analysis, such as the selected molecules in the datasets, their classification into groups, the ESP cost function, etc. Nevertheless, some clear trends can be observed.

Fig. 9a compares the average  $\text{RMSE}_{\text{ESP}}$  over all groups in Fig. 5 (Y-axis) to the average RMSE of the weak and medium electrostatic interactions from Fig. 7 (X-axis). The strong electrostatic interactions are not included because we found that these can never be reproduced reliably with any model using just atomic monopoles. The datapoint for QTAIM was omitted due to its excessively large average  $\text{RMSE}_{\text{ESP}}$ . The Pareto front only considers genuine point-charge models. Obviously, the MBIS-S method performs far better for electrostatic interactions as it goes beyond the simple point-charge model. This figure mainly shows that an accurate electrostatic potential does not guarantee accurate electrostatic interactions, and vice versa. If both qualities are of interest, the methods N, MBIS, IS and HLY are Pareto optimal. Obviously, HLY is Pareto optimal as its cost function was used to compute  $\langle \text{RMSE}_{\text{ESP}} \rangle$ , which is the reason for its advantage over MSK and CHELPG.

Fig. 9b uses the same X-axis as figure 9a but has the average of the three sensitivity values from Fig. 8 on the Y-axis. The datapoint for the Mulliken method was omitted due to its excessively large average sensitivity value. The Pareto-optimal point-charge models are N, MBIS and CM5. Again, when going beyond point charges, MBIS-S has a very attractive performance. The poor performance of ESP-fitted charges in Fig. 9b is striking. The RESP method has been the method of choice in the development of many force field models, most notably in the AMBER community. Our results indicate that RESP and other ESP fitting

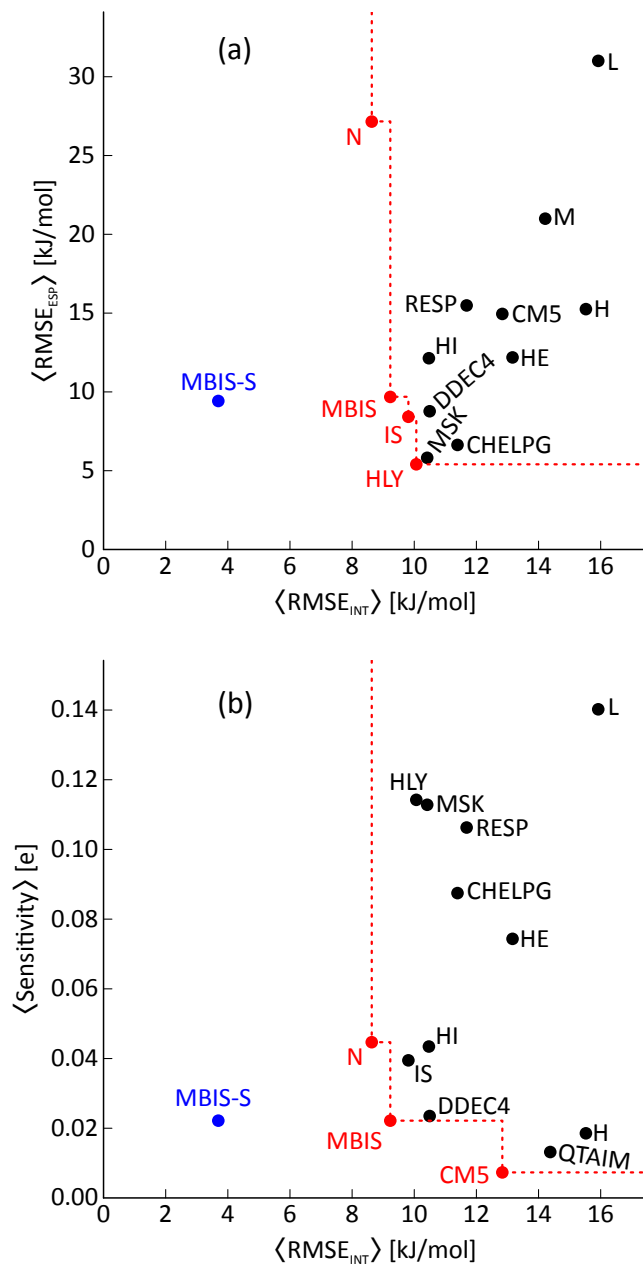


Figure 9: Pareto plots showing the trade-offs between different desirable properties of AIM methods: (a) The quality of the ESP versus the accuracy of electrostatic interactions and (b) the sensitivity of the atomic charges versus the accuracy of the electrostatic interactions. Red datapoints are on the Pareto front for all models using just point charges. The MBIS-S method goes beyond point charges by introducing a model for the distributed valence electron density.

methods are relatively poor methods for modeling electrostatic interactions in force fields.

## 5 Application to density-dependent dispersion models

Several dispersion models, typically used to correct DFT calculations, make use of the Hirshfeld partitioning method to estimate AIM polarizabilities. The polarizability of atom  $A$  in a molecule,  $\alpha_A$ , is obtained by rescaling the experimental value of the free neutral atom using the third radial moment of AIM density,  $\rho_A(\mathbf{r})$ :

$$\alpha_A = \alpha_{A,\text{free}} \frac{\langle r^3 \rangle_A}{\langle r^3 \rangle_{A,\text{free}}}, \quad (23)$$

where

$$\langle r^3 \rangle_A = \int d\mathbf{r} |\mathbf{r} - \mathbf{R}_A|^3 \rho_A(\mathbf{r}) \quad (24)$$

and similarly for  $\langle r^3 \rangle_{A,\text{free}}$ . These rescaled polarizabilities are used in various methods<sup>41–43</sup> to obtain environment-specific atomic  $C_6$  coefficients. The original Hirshfeld method is most often used in this context. The Iterative Hirshfeld method is sometimes used instead<sup>121</sup> and found to improve dispersion-corrected DFT calculations for ionic systems.<sup>82,83</sup> Density-based dispersion models are not only used for correcting DFT calculations but were recently also employed in force-field development.<sup>58</sup> In this section, we will directly compare the accuracy of molecular  $C_6$  coefficients when the Tkatchenko-Scheffler method is used in combination with different Hirshfeld variants.<sup>42</sup> We expect that similar results can be obtained with the Exchange-hole dipole model (XDM)<sup>41</sup> and related approaches such as dDsC.<sup>43</sup>

The expectation value  $\langle r^3 \rangle_{A,\text{free}}$  can be computed in two different ways. One may use the symmetry-broken ground state density of the free atom or one may constrain the atom to be spherically symmetric and closed-shell. (In both cases, the same level of theory is used as for the molecule.) The second choice is the most common in the context of dispersion models. For some elements however, this results in higher-energy states with fractionally

occupied orbitals. For the sake of consistency, compatible choices are made when computing the reference atoms for the (Iterative) Hirshfeld method. In this section, we consider in total six variants of the TS dispersion model, using three different partitioning methods: Hirshfeld (H), Iterative Hirshfeld (HI) and MBIS. For each partitioning method, ground state reference atoms (GS) or spherical closed-shell atoms (SCS) are used. The molecular  $C_6$  coefficients are also computed with Grimme’s D3 model.<sup>93</sup>

The dispersion models are tested with the database of  $C_6$  coefficients for dimers of neutral molecules by Tkatchenko and Scheffler,<sup>42</sup> which are derived from experimental dipole oscillator strengths. B3LYP/6-311+G(2df,p) densities<sup>86,122</sup> are computed for all molecules in this database and the atoms they contain. Gaussian09<sup>62</sup> is used for the B3LYP calculations, except for the SCS atoms, for which a new SCF program was written. Dimers with the xenon atom are omitted because their all-electron density can only be computed properly with relativistic corrections. As explained in subsection 2.5, the density cannot be written out by Gaussian09 when relativistic corrections are used.

Fig. 10 shows the scatter plots of the model  $C_6$  values versus the experimental reference data, including the mean percentage errors (MPE) and the root-mean-square percentage errors (RMSPE). The original TS model corresponds to TS-H-SCS, which is one of the better variants. When using the Iterative Hirshfeld method instead (TS-HI-SCS), the RMSPE increases, indicating that the model becomes less accurate. In both H and HI variants, the use of ground-state atoms, i.e. TS-H-GS or TS-HI-GS, leads to a systematic overestimation of the reference  $C_6$  coefficients. The situation is reversed when using MBIS partitioning, i.e. the dispersion model is most accurate when using ground state reference atoms (TS-MBIS-GS), while the use of spherical reference atoms (TS-MBIS-SCS) is clearly inferior. Finally, it is worth noting that Grimme’s D3 model performs slightly better than any TS variant, which is impressive given that it only makes use of the nuclear coordinates and not the electron density.

We will now analyze why the TS model only works well for certain methodological combi-

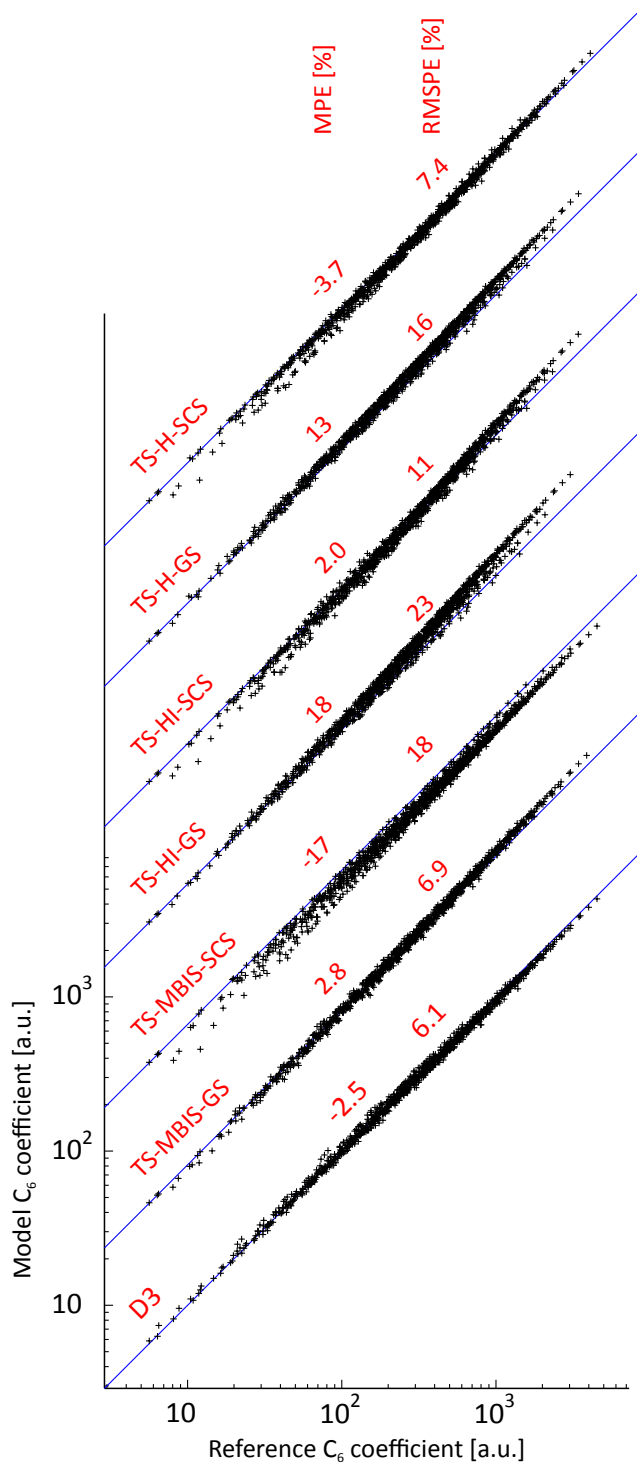


Figure 10: Scatter plots of model molecular  $C_6$  coefficients versus experimental reference values. Except for the D3 results, all data points are shifted up to avoid overlap between results from different models. The mean percentage error (MPE) and the root-mean-square percentage error (RMSPE) between model and reference molecular  $C_6$  coefficients are printed just above the corresponding data points.



nations. First, the spherical closed-shell atoms have slightly larger (or equal)  $\langle r^3 \rangle_{A,\text{free}}$  values compared to ground state atoms, as shown in Table 2. It turns out that, for completely different reasons, the Hirshfeld method exhibits some artifacts in the partitioning that also result in increased values of  $\langle r^3 \rangle_A$ , as explained below. For most molecules, these two effects balance out, except for the  $\text{H}_2$  molecule, which causes some outliers in Fig. 10 for TS-H-SCS and TS-HI-SCS at low  $C_6$  values.

Table 2: The value of  $\langle r^3 \rangle_{A,\text{free}}$  in  $\text{a}_0^3$  for the elements present in the molecules in the Tkatchenko-Sheffler set,<sup>42</sup> computed in two ways: using ground-state electron densities (GS) and using densities of atoms that are constrained to be spherical and closed-shell (SCS).

Element	$\langle r^3 \rangle_{A,\text{free,GS}} [\text{a}_0^3]$	$\langle r^3 \rangle_{A,\text{free,SCS}} [\text{a}_0^3]$	$\frac{\langle r^3 \rangle_{A,\text{free,SCS}}}{\langle r^3 \rangle_{A,\text{free,GS}}}$
H	7.9	9.7	1.23
Li	89.0	105.6	1.19
C	35.7	41.4	1.16
N	27.0	30.9	1.14
O	22.7	23.8	1.05
F	18.6	18.9	1.01
Ne	15.4	15.4	1.00
Si	103.0	114.3	1.11
S	77.0	79.7	1.04
Cl	66.7	67.4	1.01
Ar	57.2	57.2	1.00
Br	97.0	97.8	1.01
Kr	89.1	89.1	1.00

The artifact of the Hirshfeld method is very clear in Fig. 11: it shows the AIM density,  $\rho_A$ , of the hydrogen atom in hydrogen fluoride, computed with the Hirshfeld and MBIS methods. The Hirshfeld AIM density (solid blue line) is asymmetric, with more electron density toward the fluoride. This can be understood as follows. Any variant of the Hirshfeld method exhibits the same similarity principle,<sup>60</sup> due to Eq. (5): the Hirshfeld AIM density will be as close as possible to that of the pro-atom. Because the density tail of isolated hydrogen (dashed blue line) is so different from that of hydrogen in HF, the Hirshfeld AIM density is aspherical with too much density in the bonding region. This accumulation of density, relatively far away from the hydrogen nucleus, leads to larger values of high radial

moments, such as  $\langle r^3 \rangle_A$ . This is a general feature of the Hirshfeld method, also seen in other molecules. In the MBIS method, no such asymmetries are found because the parameters  $\sigma_{A,i}$  in Eq. (7), i.e. the widths of the Slater functions, are also optimized to match the molecular electron density. By consequence, the MBIS AIM density (solid green curve) almost coincides with the corresponding pro-atom (dashed green curve) and is therefore close to symmetric.

Fig. 11 also shows a side effect of the MBIS method: where the MBIS AIM density (solid green curve) passes through the nucleus of fluoride, some ripples can be seen, because some details in the fluoride core electron density cannot be reproduced by the Slater functions. These ripples are very local and therefore have a negligible effect on  $\langle r^3 \rangle_A$ .

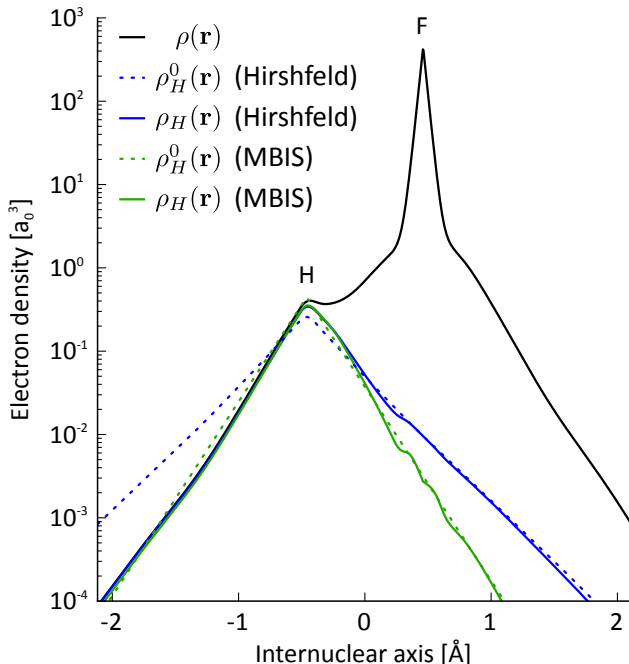


Figure 11: Total electron density (black) of hydrogen fluoride along the internuclear axis, including the AIM density (solid) and the pro-atomic density (dashed) of hydrogen for the Hirshfeld (blue) and MBIS method (green).

Fig. 12 further illustrates the mismatch between the density tails of the (Iterative) Hirshfeld pro-atoms and the molecular densities in the TS set. It shows the MBIS *valence width* of the carbon atoms of all molecules in the TS set<sup>42</sup> versus their MBIS charge. This figure also includes the data points for free carbon atoms and ions, which are used as pro-atoms in the (iterative) Hirshfeld method. The trends in this figure are general: neutral atoms

and anions, have density tails that decay slower (i.e. higher valence width) than molecular electron densities. This leads to the asymmetric (Iterative) Hirshfeld AIM densities as the one observed in Fig. 11. Especially when computing higher radial or multipole moments with the (Iterative) Hirshfeld method, this may result in undesirable artifacts. In previous work, the erroneous density tails of unstable anions were corrected, e.g. by computing these atoms in a Watson sphere<sup>123</sup>. For example, such corrections were used by Bučko *et al.*<sup>82,83</sup> in their tests of the TS-HI variant. The results in Fig. 12 suggest that it is also advantageous to reduce the density tails of stable anions and neutral atoms.

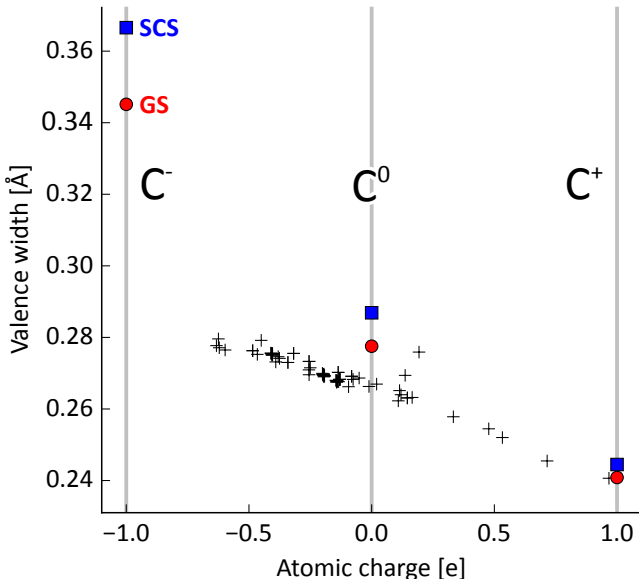


Figure 12: The MBIS *valence width* ( $\sigma_{A,v}$ , see subsection 2.4) of carbon atoms in the Tkatchenko-Sheffler set of molecules<sup>42</sup> versus their MBIS charge (black plus). The valence widths of isolated carbon atoms and ions are also included: ground states atoms (red circle) and spherical closed-shell atoms (blue square).

Finally, note that a few other Hirshfeld variants were proposed, in which the density tails of the pro-atoms are optimized to match the molecular electron density, most notably the Iterative Stockholder (IS) method<sup>37,53</sup> and some of its variants including the Gaussian ISA<sup>54</sup> and BS-ISA+DF.<sup>30</sup>

## 6 Conclusions and outlook

The MBIS method is a new density-based AIM method that is particularly suitable for the development of efficient and relatively accurate electrostatic force-field models. MBIS belongs to the family of Hirshfeld methods. Its pro-density is expanded in a minimal set of atom-centered s-type Slater density functions, whose parameters are fitted to a given molecular electron density by minimizing the Kullback-Leibler divergence. In that sense, it can also be interpreted as an information theory density-fitting method, where the Slater functions as such are used in applications, rather than the atoms-in-molecules densities.

The MBIS method is extensively tested for the development of electrostatic force-field models. When it is just used for the purpose of deriving atomic charges, it is one of the best methods available to date, in terms of robustness and accuracy of the electrostatic interactions. When the MBIS Slater functions are used to describe the valence electron density in a force field, the error on the electrostatic interactions can be reduced by 50%, if the electrostatic interaction is not too strong. This is a computationally efficient approach to describe the so-called penetration effect, i.e. the deviation of interatomic electrostatic interactions from that of point charges, when the atomic densities begin to overlap. MBIS is also useful beyond the scope of frozen-density electrostatics, e.g. when modeling dispersion interactions, or to analyze density fluctuations that a polarizable force field should reproduce.

In future work, we will focus on improving our method, its implementation and more applications in different areas. The obvious methodological improvement is a better model for the pro-molecular density, e.g. by including atomic multipoles<sup>30,32</sup> or by making it compatible with pseudo-densities. An improved pro-molecule model should not merely result in a better fit to a given electron density; one should also avoid too many degrees of freedom for the sake of robustness. Moreover, the use of an improved pro-density model in force fields should remain computationally efficient.

## Acknowledgement

The authors would like to thank F. De Proft and P. Bultinck for inspiring discussions and useful suggestions. T.V., S.V. and L.V. acknowledge the Foundation of Scientific Research - Flanders (FWO). T.V., L.V., M.W. and V.V.S. thank the Research Board of Ghent University (BOF), and BELSPO in the frame of IAP/7/05 for their financial support. V.V.S. acknowledges the European Research Council for funding the European Union’s Horizon 2020 research and innovation programme [consolidator ERC grant agreement no. 647755 – DYNPOR (2015–2020)]. The computational resources and services used were provided by Ghent University (Stevin Supercomputer Infrastructure). F.H.-Z. was supported by a Vanier-CGS fellowship and a Michael Smith Foreign Study Supplement from the National Sciences and Engineering Research Council of Canada (NSERC); P.W.A. was supported by a Discovery Grant and a E.W.R. Steacie Fellowship from NSERC.

## Supporting Information Available

Atom types in MIL-53(Al), chabazite and ozone. Description of the ZG237 set of molecular dimers and Cartesian coordinates of the MP2/cc-pVTZ+CP optimized dimer geometries in this set. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

## References

- (1) Kohlhoff, K. J.; Shukla, D.; Lawrenz, M.; Bowman, G. R.; Konerding, D. E.; Belov, D.; Altman, R. B.; Pande, V. S. Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nature Chem.* **2014**, *6*, 15–21.
- (2) Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Denneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.;

- Greskamp, B.; Ho, C. R.; Ierardi, D. J.; Iserovich, L.; Kuskin, J. S.; Larson, R. H.; Layman, T.; Lee, L.-S.; Lerer, A. K.; Li, C.; Killebrew, D.; Mackenzie, K. M.; Mok, S. Y.-H.; Moraes, M. A.; Mueller, R.; Nociolo, L. J.; Peticolas, J. L.; Quan, T.; Ramot, D.; Salmon, J. K.; Scarpazza, D. P.; Schafer, U. B.; Siddique, N.; Snyder, C. W.; Spengler, J.; Tang, P. T. P.; Theobald, M.; Toma, H.; Towles, B.; Vitale, B.; Wang, S. C.; Young, C. Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. *SC14 Proceedings* **2014**, 41–53.
- (3) Kim, J.; Abouelnasr, M.; Lin, L.-C.; Smit, B. Large-Scale Screening of Zeolite Structures for CO<sub>2</sub> Membrane Separations. *J. Am. Chem. Soc.* **2013**, *135*, 7545–7552.
- (4) Macchiarulo, A.; Costantino, G.; Sbaglia, R.; Aiello, S.; Meniconi, M.; Pellicciari, R. The role of electrostatic interaction in the molecular recognition of selective agonists to metabotropic glutamate receptors. *Proteins* **2003**, *50*, 609–619.
- (5) Ramsahye, N. A.; Maurin, G.; Burrelly, S.; Llewellyn, P.; Loiseau, T.; Ferey, G. Charge distribution in metal organic framework materials: transferability to a preliminary molecular simulation study of the CO(2) adsorption in the MIL-53 (Al) system. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1059–1063.
- (6) Vanduyfhuys, L.; Verstraelen, T.; Vandichel, M.; Waroquier, M.; Van Speybroeck, V. Ab initio parametrized force field for the flexible metal-organic framework MIL-53(Al). *J. Chem. Theory Comput.* **2012**, *8*, 3217–3231.
- (7) Kadantsev, E. S.; Boyd, P. G.; Daff, T. D.; Woo, T. K. Fast and Accurate Electrostatics in Metal Organic Frameworks with a Robust Charge Equilibration Parameterization for High-Throughput Virtual Screening of Gas Adsorption. *J. Phys. Chem. Lett.* **2013**, *4*, 3056–3061.
- (8) Gabrieli, A.; Sant, M.; Demontis, P.; Suffritti, G. B. Partial Charges in Periodic Sys-

- tems: Improving Electrostatic Potential (ESP) Fitting via Total Dipole Fluctuations and Multiframe Approaches. *J. Chem. Theory Comput.* **2015**, *11*, 3829–3843.
- (9) Hamad, S.; Balestra, S. R. G.; Bueno-Perez, R.; Calero, S.; Ruiz-Salvador, A. R. Atomic charges for modeling metal–organic frameworks: Why and how. *J. Solid State Chem.* **2015**, *223*, 144–151.
- (10) 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.
- (11) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050–8053.
- (12) Tafipolsky, M.; Engels, B. Accurate Intermolecular Potentials with Physically Grounded Electrostatics. *J. Chem. Theory Comput.* **2011**, *7*, 1791–1803.
- (13) Warshel, A.; Kato, M.; Pisiakov, A. V. Polarizable Force Fields: History, Test Cases, and Prospects. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.
- (14) 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.
- (15) Verstraelen, T.; Vandenbrande, S.; Ayers, P. Direct computation of parameters for accurate polarizable force fields. *J. Chem. Phys.* **2014**, *141*, 194114.
- (16) Ji, C.; Mei, Y. Some Practical Approaches to Treating Electrostatic Polarization of Proteins. *Acc. Chem. Res.* **2014**, *47*, 2795–2803.
- (17) Potoff, J. J.; Siepmann, J. I. Vapor-liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE Journal* **2001**, *47*, 1676–1682.

- (18) Rai, N.; Siepmann, J. I. Transferable potentials for phase equilibria. 10. Explicit-hydrogen description of substituted benzenes and polycyclic aromatic compounds. *J. Phys. Chem. B* **2013**, *117*, 273–288.
- (19) Fox, T.; Kollman, P. A. Application of the RESP Methodology in the Parametrization of Organic Solvents. *J. Phys. Chem. B* **1998**, *102*, 8070–8079.
- (20) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; 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.
- (21) Stone, A. J.; Alderton, M. Distributed multipole analysis. *Molec. Phys.* **1985**, *56*, 1047–1064.
- (22) Kairys, V.; Jensen, J. H. Evaluation of the charge penetration energy between non-orthogonal molecular orbitals using the Spherical Gaussian Overlap approximation. *Chem. Phys. Lett.* **1999**, *315*, 140–144.
- (23) Krapp, A.; Bickelhaupt, F. M.; Frenking, G. Orbital Overlap and Chemical Bonding. *Chem. Eur. J.* **2006**, *12*, 9196–9216.
- (24) Spackman, M. The use of the promolecular charge density to approximate the penetration contribution to intermolecular electrostatic energies. *Chem. Phys. Lett.* **2006**, *418*, 158–162.
- (25) 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.
- (26) 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 Electro-



- static Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 2609–2618.
- (27) Bader, R. F. W. A quantum theory of molecular structure and its applications. *Chem. Rev.* **1991**, *91*, 893–928.
- (28) 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.
- (29) Cisneros, G. A.; Piquemal, J.-P.; Darden, T. A. Generalization of the Gaussian electrostatic model: Extension to arbitrary angular momentum, distributed multipoles, and speedup with reciprocal space methods. *J. Chem. Phys.* **2006**, *125*, 184101.
- (30) Misquitta, A. J.; Stone, A. J.; Fazeli, F. Distributed Multipoles from a Robust Basis-Space Implementation of the Iterated Stockholder Atoms Procedure. *J. Chem. Theory Comput.* **2014**, *10*, 5405–5418.
- (31) Wang, B.; Truhlar, D. G. Screened Electrostatic Interactions in Molecular Mechanics. *J. Chem. Theory Comput.* **2014**, *10*, 4480–4487.
- (32) Öhrn, A.; Hermida-Ramon, J. M.; Karlström, G. Method for Slater-Type Density Fitting for Intermolecular Electrostatic Interactions with Charge Overlap. I. The Model. *J. Chem. Theory Comput.* **2016**, 10.1021/acs.jctc.5b01155.
- (33) Leslie, M. DL\_MULTI—A molecular dynamics program to use distributed multipole electrostatic models to simulate the dynamics of organic crystals. *Mol. Phys.* **2008**, *106*, 1567–1578.
- (34) Popelier, P. L. A. QCTFF: On the construction of a novel protein force field. *Int. J. Quant. Chem.* **2015**, *115*, 1005–1011.
- (35) Ghillemin, D.; Bultinck, P.; Van Neck, D.; Ayers, P. W. A self-consistent Hirshfeld

- method for the atom in the molecule based on minimization of information loss. *J. Comput. Chem.* **2011**, *32*, 1561–1567.
- (36) Heidar-Zadeh, F.; Ayers, P. W. How pervasive is the Hirshfeld partitioning? *J. Chem. Phys.* **2015**, *142*, 044107.
- (37) Lillestolen, T. C.; Wheatley, R. J. Redefining the atom: atomic charge densities produced by an iterative stockholder approach. *Chem. Commun.* **2008**, 5909–5911.
- (38) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chem. Acc.* **1977**, *44*, 129–138.
- (39) Nalewajski, R. F.; Parr, R. G. Information theory, atoms in molecules, and molecular similarity. *P. Natl. Acad. Sci.* **2000**, *97*, 8879–8882.
- (40) 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.
- (41) Becke, A. D.; Johnson, E. R. Exchange-hole dipole moment and the dispersion interaction revisited. *J. Chem. Phys.* **2007**, *127*, 154108.
- (42) 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*, 73005.
- (43) Steinmann, S. N.; Corminboeuf, C. Comprehensive Benchmarking of a Density-Dependent Dispersion Correction. *J. Chem. Theory Comput.* **2011**, *7*, 3567–3577.
- (44) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2003**, *103*, 1793–1874.
- (45) Davidson, E. R.; Chakravorty, S. A test of the Hirshfeld definition of atomic charges and moments. *Theor. Chem. Acc.* **1992**, *83*, 319–330.

- (46) 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.
- (47) Van Damme, S.; Bultinck, P.; Fias, S. Electrostatic Potentials from Self-Consistent Hirshfeld Atomic Charges. *J. Chem. Theory Comput.* **2009**, *5*, 334–340.
- (48) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. The electronegativity equalization method and the split charge equilibration applied to organic systems: parametrization, validation, and comparison. *J. Chem. Phys.* **2009**, *131*, 044127.
- (49) Verstraelen, T.; Pauwels, E.; De Proft, F.; Van Speybroeck, V.; Geerlings, P.; Waroquier, M. Assessment of Atomic Charge Models for Gas-Phase Computations on Polypeptides. *J. Chem. Theory Comput.* **2011**, *8*, 661–676.
- (50) Bultinck, P.; Ayers, P.; Fias, S.; Tiels, K.; Van Alsenoy, C. Uniqueness and basis set dependence of iterative Hirshfeld charges. *Chem. Phys. Lett.* **2007**, *444*, 205–208.
- (51) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. Hirshfeld-E partitioning: AIM charges with an improved trade-off between robustness and accurate electrostatics. *J. Chem. Theory Comput.* **2013**, *9*, 2221–2225.
- (52) Verstraelen, T.; Sukhomlinov, S. V.; Van Speybroeck, V.; Waroquier, M.; Smirnov, K. S. Computation of charge distribution and electrostatic potential in silicates with the use of chemical potential equalization models. *J. Phys. Chem. C* **2012**, *116*, 490–504.
- (53) Lillestolen, T. C.; Wheatley, R. J. Atomic charge densities generated using an iterative stockholder procedure. *J. Chem. Phys.* **2009**, *131*, 144101.
- (54) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. The conformational sensitivity of iterative stockholder partitioning schemes. *Chem. Phys. Lett.* **2012**, *545*, 138–143.

- (55) Manz, T. A.; Sholl, D. S. Chemically Meaningful Atomic Charges That Reproduce the Electrostatic Potential in Periodic and Nonperiodic Materials. *J. Chem. Theory Comput.* **2010**, *6*, 2455–2468.
- (56) Manz, T. A.; Sholl, D. S. Improved Atoms-in-Molecule Charge Partitioning Functional for Simultaneously Reproducing the Electrostatic Potential and Chemical States in Periodic and Nonperiodic Materials. *J. Chem. Theory Comput.* **2012**, *8*, 2844–2867.
- (57) Vanpoucke, D. E. P.; Bultinck, P.; Van Driessche, I. Extending Hirshfeld-I to bulk and periodic materials. *J. Comput. Chem.* **2013**, *34*, 405–417.
- (58) Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning. *J. Chem. Theory Comput.* **2016**, 10.1021/acs.jctc.6b00027.
- (59) Haldoupis, E.; Nair, S.; Sholl, D. S. Finding MOFs for Highly Selective CO<sub>2</sub>/N<sub>2</sub> Adsorption Using Materials Screening Based on Efficient Assignment of Atomic Point Charges. *J. Am. Chem. Soc.* **2012**, *134*, 4313–4323.
- (60) Ayers, P. Atoms in molecules, an axiomatic approach. I. Maximum transferability. *J. Chem. Phys.* **2000**, *113*, 10886–10898.
- (61) Parr, R. G.; Ayers, P. W.; Nalewajski, R. F. What Is an Atom in a Molecule? *J. Phys. Chem. A* **2005**, *109*, 3957–3959.
- (62) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.;

- Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09 Revision D.01*; Gaussian Inc.: Wallingford CT, 2013.
- (63) Verstraelen, T.; Boguslawski, K.; Tecmer, P.; Heidar-Zadeh, F.; Chan, M.; Kim, T. D.; Zhao, Y.; Vandenbrande, S.; Yang, D.; González-Espinoza, C. E.; Limacher, P. A.; Berrocal, D.; Malek, A.; Ayers, P. W. *HORTON 2.0.0*; 2015; Available at <http://theochem.github.com/horton> (accessed June 14, 2016).
- (64) Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **1988**, *88*, 2547–2553.
- (65) Blöchl, P. E.; Först, C. J.; Schimpl, J. Projector augmented wave method: ab initio molecular dynamics with full wave functions. *Bull. Mater. Sci.* **2003**, *26*, 33–41.
- (66) Mortensen, J. J.; Hansen, L. B.; Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B* **2005**, *71*, 035109.
- (67) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dulak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. a.; Kristoffersen, H. H.; Kuisma, M.; Larsen, a. H.; Lehtovaara, L.; Ljungberg, M.; Lopez-Acevedo, O.; Moses, P. G.; Ojanen, J.; Olsen, T.; Petzold, V.; Romero, N. a.; Stausholm-Møller, J.; Strange, M.; Tritsaris, G. a.; Vanin, M.; Walter, M.; Hammer, B.; Häkkinen, H.; Madsen, G. K. H.; Nieminen, R. M.; Nørskov, J. K.; Puska, M.; Rantala, T. T.; Schiøtz, J.; Thygesen, K. S.; Jacobsen, K. W. Electronic structure calculations with GPAW: a real-

- space implementation of the projector augmented-wave method. *J. Physics.-Condens. Mat.* **2010**, *22*, 253202.
- (68) Bahn, S. R.; Jacobsen, K. W. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.* **2002**, *4*, 56–66.
- (69) Lee, L. P.; Cole, D. J.; Skylaris, C.-K.; Jorgensen, W. L.; Payne, M. C. Polarized Protein-Specific Charges from Atoms-in-Molecule Electron Density Partitioning. *J. Chem. Theory Comput.* **2013**, *9*, 2981–2991.
- (70) Hu, H.; Lu, Z.; Yang, W. Fitting Molecular Electrostatic Potentials from Quantum Mechanical Calculations. *J. Chem. Theory Comput.* **2007**, *3*, 1004–1013.
- (71) Gill, P. M. W. Extraction of Stewart atoms from electron densities. *J. Phys. Chem.* **1996**, *100*, 15421–15427.
- (72) Slater, J. C. Atomic shielding constants. *Phys. Rev.* **1930**, *36*, 57–64.
- (73) Wang, W.-P.; Parr, R. G. Statistical atomic models with piecewise exponentially decaying electron densities. *Phys. Rev. A* **1977**, *16*, 891–902.
- (74) Wang, W.-P. Fixed-shell statistical atomic models with piecewise exponentially decaying electron densities. *Phys. Rev. A* **1982**, *25*, 2901–2912.
- (75) Fernandez Pacios, L. A simplified representation of atomic electron densities and electrostatic potentials. *J. Phys. Chem.* **1991**, *95*, 10653–10658.
- (76) 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. USA* **2005**, *102*, 7829–7834.
- (77) Bultinck, P.; Cooper, D. L.; Van Neck, D. Comparison of the Hirshfeld-I and iterated stockholder atoms in molecules schemes. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3424–3429.

- (78) Baerends, E.; Ellis, D.; Ros, P. Self-consistent molecular Hartree—Fock—Slater calculations I. The computational procedure. *Chem, Phys.* **1973**, *2*, 41–51.
- (79) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On first-row diatomic molecules and local density models. *J. Chem. Phys.* **1979**, *71*, 4993.
- (80) Fonseca Guerra, C.; Snijders, J. G.; Te Velde, G.; Baerends, E. J. Towards an order-N DFT method. *Theor. Chem. Acc.* **1998**, *99*, 391–403.
- (81) Stewart, R. F. V. One-Electron Density Functions and Many-Centered Finite Multipole Expansions. *Israel J. Chem.* **1977**, *16*, 124–131.
- (82) Bučko, T.; Lebègue, S.; Hafner, J.; Ángyán, J. G. Improved Density Dependent Correction for the Description of London Dispersion Forces. *J. Chem. Theory Comput.* **2013**, *9*, 4293–4299.
- (83) Bučko, T.; Lebègue, S.; Ángyán, J. G.; Hafner, J. Extending the applicability of the Tkatchenko-Scheffler dispersion correction via iterative Hirshfeld partitioning. *J. Chem. Phys.* **2014**, *141*, 034114.
- (84) Gould, T.; Bucko, T. C6 coefficients and dipole polarizabilities for all atoms and many ions in rows 1-6 of the periodic table. *J. Chem. Theory Comput.* **2016**, *0*, null, 10.1021/acs.jctc.6b00361.
- (85) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.
- (86) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.
- (87) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark structures and binding energies

- of small water clusters with anharmonicity corrections. *J. Phys. Chem. A* **2011**, *115*, 12034–12046.
- (88) Hayward, J. S.; Reimers, J. R. Unit cells for the simulation of hexagonal ice. *J. Chem. Phys.* **1997**, *106*, 1518.
- (89) Lazzaro, A.; Gloeß, A.; Seitsonen, A. P.; Mundy, C.; Pousa, C.; Golze, D.; Mohamed, F.; Schiffmann, F.; Tabacchi, G.; Forbert, H.; Bani-Hashemian, H.; Bethune, I.; Kuo, I.-F. W.; Wilhelm, J.; VandeVondele, J.; Hutter, J.; Tong, L.; Walewski, L.; Schoenherr, M.; Guidon, M.; Mauri-Iannuzzi, M.; McGrath, M.; Krack, M.; Ben, M. D.; Ceriotti, M.; Schuett, O.; Seewald, P.; Andermatt, S.; Laino, T.; Chassaing, T.; Kuehne, T.; Borstnik, U.; Weber, V.; Rybkin, V. *CP2K 2.6.0*; 2014; Available at <http://www.cp2k.org> (accessed June 14, 2016).
- (90) VandeVondele, J.; Hutter, J. An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.* **2003**, *118*, 4365.
- (91) Vandevondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128.
- (92) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. CP2K: atomistic simulations of condensed matter systems. *WIREs Comput. Mol. Sci.* **2014**, *4*, 15–25.
- (93) 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.
- (94) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.



- (95) VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **2007**, *127*, 114105.
- (96) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703–1710.
- (97) Krack, M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.* **2005**, *114*, 145–152.
- (98) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (99) Horn, H.; Swope, W.; Pitera, J.; Madura, J.; Dick, T.; Hura, G.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.
- (100) Piquemal, J.-P.; Chelli, R.; Procacci, P.; Gresh, N. Key Role of the Polarization Anisotropy of Water in Modeling Classical Polarizable Force Fields. *J. Phys. Chem. A* **2007**, *111*, 8170–8176.
- (101) Burnham, C. J.; Anick, D. J.; Mankoo, P. K.; Reiter, G. F. The vibrational proton potential in bulk liquid water and ice. *J. Chem. Phys.* **2008**, *128*, 154519.
- (102) Yu, W.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Six-site polarizable model of water based on the classical Drude oscillator. *J. Chem. Phys.* **2013**, *138*, 34508.
- (103) Mei, Y.; Simmonett, A. C.; Pickard, F. C.; DiStasio, R. A.; Brooks, B. R.; Shao, Y. Numerical Study on the Partitioning of the Molecular Polarizability into Fluctuating Charge and Induced Atomic Dipole Contributions. *J. Phys. Chem. A* **2015**, *119*, 5865–5882.
- (104) 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.
- (105) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8*, 141–151.
- (106) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.
- (107) Férey, G.; Latroche, M.; Serre, C.; Millange, F.; Loiseau, T.; Percheron-Guégan, A. Hydrogen adsorption in the nanoporous metal-benzenedicarboxylate  $M(OH)(O_2C-C_6H_4-CO_2)$  ( $M = Al^{3+}, Cr^{3+}$ ), MIL-53. *Chem. Commun.* **2003**, 2976–2977.
- (108) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129–145.
- (109) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361–373.
- (110) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.
- (111) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833–1840.
- (112) Löwdin, P.-O. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365–375.

- (113) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural population analysis. *J. Chem. Phys.* **1985**, *83*, 735–746.
- (114) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.
- (115) Manz, T. A.; Limas, N. G. *Chargemol program for performing DDEC analysis, Version 09.15.2014*; 2014; Available at <http://ddec.sourceforge.net> (accessed June 14, 2016).
- (116) Keith, T. A. *AIMAll, Version 11.06.19*; TK Gristmill Software: Overland Park KS, USA, 2014; Available at <http://aim.tkgristmill.com> (accessed June 14, 2016).
- (117) Ibrahim, M. A. A. Molecular mechanical study of halogen bonding in drug discovery. *J. Comput. Chem.* **2011**, *32*, 2564–2574.
- (118) Bader, R.; Carroll, M.; Cheeseman, J.; Chang, C. Properties of atoms in molecules: atomic volumes. *J. Am. Chem. Soc.* **1987**, *109*, 7968–7979.
- (119) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497–4514.
- (120) Devereux, M.; Popelier, P. L. A.; McLay, I. M. Toward an ab initio fragment database for bioisosterism: Dependence of QCT properties on level of theory, conformation, and chemical environment. *J. Comput. Chem.* **2009**, *30*, 1300–1318.
- (121) Steinmann, S.; Corminboeuf, C. A System-Dependent Density-Based Dispersion Correction. *J. Chem. Theory Comput.* **2010**, *6*, 1990–2001.
- (122) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

- (123) Watson, R. E. Analytic Hartree-Fock Solutions for  $O^-$ . *Phys. Rev.* **1958**, *111*, 1108–1110.

# Graphical TOC Entry

