

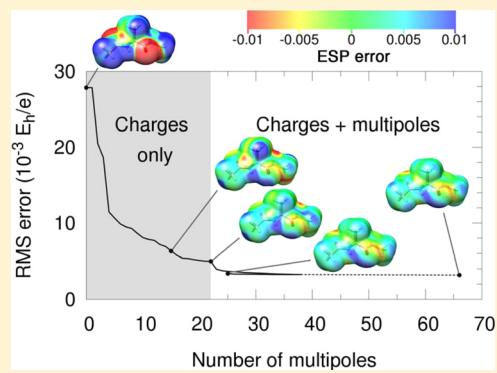
Systematic Improvement of Potential-Derived Atomic Multipoles and Redundancy of the Electrostatic Parameter Space

Sofie Jakobsen and Frank Jensen*

Department of Chemistry, Aarhus University, DK-8000 Aarhus, Denmark

Supporting Information

ABSTRACT: We assess the accuracy of force field (FF) electrostatics at several levels of approximation from the standard model using fixed partial charges to conformational specific multipole fits including up to quadrupole moments. Potential-derived point charges and multipoles are calculated using least-squares methods for a total of ~ 1000 different conformations of the 20 natural amino acids. Opposed to standard charge fitting schemes the procedure presented in the current work employs fitting points placed on a single isodensity surface, since the electrostatic potential (ESP) on such a surface determines the ESP at all points outside this surface. We find that the effect of multipoles beyond partial atomic charges is of the same magnitude as the effect due to neglecting conformational dependency (i.e., polarizability), suggesting that the two effects should be included at the same level in FF development. The redundancy at both the partial charge and multipole levels of approximation is quantified. We present an algorithm which stepwise reduces or increases the dimensionality of the charge or multipole parameter space and provides an upper limit of the ESP error that can be obtained at a given truncation level. Thereby, we can identify a reduced set of multipole moments corresponding to $\sim 40\%$ of the total number of multipoles. This subset of parameters provides a significant improvement in the representation of the ESP compared to the simple point charge model and close to the accuracy obtained using the complete multipole parameter space. The selection of the $\sim 40\%$ most important multipole sites is highly transferable among different conformations, and we find that quadrupoles are of high importance for atoms involved in π -bonding, since the anisotropic electric field generated in such regions requires a large degree of flexibility.



INTRODUCTION

A force field (FF) description of the potential energy surface is essential for simulations of large biomolecules over long time scales. The FF energy is divided into a number of bonded and nonbonded interaction terms, which are parametrized to fit experimental or ab initio results.^{1,2} The nonbonded electrostatic energy usually dominates the intermolecular interactions, especially for polar systems. Most commonly used FFs, such as AMBER,^{3,4} CHARMM,⁵ GROMOS,⁶ and OPLS,^{7,8} employ a fixed point charge model for the electrostatic energy, where both direct polarization and higher order multipole effects are neglected. Higher order atomic multipoles provide a rigorous and systematic path for improving the electrostatic energy for a fixed molecular structure, while polarization effects are important for modeling the conformational dependence and intermolecular interactions. Development of the next generation of FFs has focused on implementing explicit polarization, and several reviews^{9–12} discuss the details. These models can be divided into three classes: the induced dipole model (e.g., AMOEBA,¹³ AMBER ff02^{14,15}), the Drude oscillator model (e.g., CHARMM-PIPF,¹⁶ CHARMM-Drude^{17,18}), and the fluctuating charge model (e.g., SIBFA¹⁹ and the POLS^{20,21} water model). The Drude oscillator and fluctuating charge models retain a partial charge description and are therefore easy to implement in current software packages. Polarizable FFs

have the attractive feature that the parameters can be assigned directly from ab initio results,²² while fixed charge FFs employ artificially increased partial charges in order to incorporate the missing polarization in an average fashion.⁹

Increasing the number of parameters for describing the electrostatic interaction (static or induced) can with careful parametrization improve the accuracy of the FF energy, but it also leads to an increase in the computational requirements. It is therefore of interest to investigate how the representation of the electrostatic interaction can be improved in a systematic fashion with the smallest computational cost. This is closely related to the problem of redundancy of the parameter space. Adding higher orders of multipoles at all atomic sites may include parameters that describe the same physical effect, as for example an atomic dipole moment may generate nearly the same electrostatic potential (ESP) as two monopoles. It is well-known that a model with (only) partial charges at all atomic sites contains many near-redundant parameters even for medium sized molecules,^{23–28} and while adding higher order multipoles at a first glance will improve the accuracy, it will also significantly increase the near-redundancy problem. Fitting parameters that

Received: September 3, 2014

Published: October 20, 2014

are near-redundant is numerically unstable and also undesirable in terms of computational efficiency.

In the present paper, we investigate the problem of representing a molecular ESP by atomic multipoles, with special focus on identifying and eliminating redundancy in the parameter space. The longer aim is to design new FFs capable of providing well-defined accuracies and remain computationally efficient. Our reference data set consists of ab initio ESPs for ~1000 geometry optimized conformations of the 20 neutral natural amino acids capped with acetyl and methyl amide at the N- and C-terminals, respectively. Analysis of fitting atomic monopole, dipole, and quadrupole moments indicate that roughly one-third of the partial charges and almost two-thirds of the total number of multipoles moments can be removed without significantly affecting the results.

METHODS

The molecular ESP can be calculated by electronic structure methods and is fairly insensitive toward the level of theory. In the present case we have employed the ω B97X-D²⁹/aug-cc-pVQZ³⁰ method, but the fitting and parameter analysis can be employed for any reference method, and the conclusions are unlikely to change upon improvements in the methods for calculating the ESP. Trial molecular structures for the 20 neutral natural amino acids capped with acetyl and methyl amide group was generated by FF methods. The structures are fully optimized at the B3LYP^{31,32}/aug-cc-pVQZ level as described in ref 33 and subsequently reoptimized at the ω B97X-D/aug-cc-pVQZ level using Gaussian09,³⁴ generating ~1000 unique conformations. While this by no means is an exhaustive search, it provides a representative sampling of the phase space available for these peptide models. Charged amino acids have not been considered, but there is no reason to expect them to behave different from the neutral ones. The goal in FF methods is to represent the ESP in a computationally efficient form by employing electrostatic multipole parameters. We focus in the present case on representing the static ESP for a given molecular structure, while the inclusion of polarization will be considered in future work.

Point Charge Fit. Point charges can be fitted to reproduce the ab initio ESP, and several fitting schemes have been described in the literature. They are all based on the same standard procedure, including the MK/MKS charges developed by Merz, Kollman, and Singh,^{35,36} the RESP charges developed by Bayly et al.,^{23,24} the CHELP charges developed by Chirlian and Franci,³⁷ and the CHELPG charge developed by Breneman and Wiberg.³⁸ These schemes differ primarily in the selection of sampling points, where the MK, RESP, and CHELP use surface-based points, while CHELPG uses grid-based points. Other methods place points at random positions around the molecules^{39,40} or based on their Boltzmann weight from the van der Waals energy term.⁴⁰ The sensitivity of the fitted atomic charges toward the different schemes for selecting sampling points has been the subject of several investigations.^{26,41} Tsiper and Burke have shown, however, that the ESP on an isodensity surface that includes essentially all the electron density *uniquely* determines the ESP at *all* points outside the surface.⁴² This implies that the discussion of sampling points is largely irrelevant, and different fitting schemes thus primarily differ in terms of the incomplete sampling of the proper isodensity surface and associated numerical instabilities.

In the current work, the fitting points are placed on a single isodensity surface corresponding to $\rho = 10^{-3}$ e/ a_0^3 , which reflects a typical van der Waals molecular surface. The WFA

(Wave Function Analysis) program developed by Bulat et al.⁴³ is used to represent this isosurface as a collection of polygons created by the Marching Tetrahedra algorithm, based on a cubic grid of the electron density. The cubegen facility in Gaussian09³⁴ is used for computing the electron densities and ESPs. The convergence of the ESP surface integral is quantified for a single conformation of alanine using different grid spacing in the range from $0.10a_0$ to $1.50a_0$, as reported in Figure S1.1 in Supporting Information. A grid separation of $0.20a_0$ (~0.11 Å) is small enough that the surface integral of the ESP is converged to within 10^{-4} e_h/e, and is used throughout this work. It is important to emphasize that the present surface integral approach in connection with a sufficient small grid separation ensures that *all* the information contained within the electronic structure calculation is taken into account and that no further improvement can be obtained by adding more sampling points. It furthermore provides a unique way of quantifying the error arising from approximate representations of the ESP by atomic multipoles.

The standard least-squares method for fitting point charges to the ESP minimizes the target function, χ^2 , defined in eq 1, where the isodensity surface integral⁴³ is approximated as a sum over points on the surface.

$$\begin{aligned} \chi^2 &= \frac{1}{S} \oint_S [\phi^{\text{QM}}(\mathbf{r}) - \phi^{\text{app}}(\mathbf{r})]^2 d\mathbf{s} \\ &\approx \frac{1}{P} \sum_{p=1}^P [\phi^{\text{QM}}(\mathbf{r}_p) - \phi^{\text{app}}(\mathbf{r}_p)]^2 \end{aligned} \quad (1)$$

$$\phi^{\text{app}}(\mathbf{r}_p) = \sum_{k=1}^{N_A} \frac{q_k}{|\mathbf{r}_{kp}|} \quad (2)$$

ϕ^{QM} is the quantum mechanical ESP determined in P fitting points and ϕ^{app} is the ESP approximated by N_A point charges as shown in eq 2 in atomic units. q_k is the charge of atom k , and the shorthand notation $\mathbf{r}_{kp} = \mathbf{r}_p - \mathbf{r}_k$ is used. The set of point charges that minimizes χ^2 can be found by analytical differentiation of eq 1 and solving the derived linear equations.

Different constraints or restraints are often used to determine chemically more relevant point charges. Chirlian and Franci³⁷ introduced the use of Lagrangian multipliers to constrain the sum of all charges to be equal to the total molecular charge. This increases the number of parameters by one, which tends to aggravate numerical instabilities due to rank deficiency, as discussed in the following. Hinsen and Roux⁴¹ suggested another route by eliminating constraints that are linear in the atomic charges, thus reducing the number of parameters by one. Elimination is used in the present work to constrain the sum of atomic charges to the total molecular charge, and the number of unknowns in the fitting procedure is thereby $N_A - 1$, as the last charge, q_n , is determined by the constraint.

Potential-derived point charges are often further restrained in order to provide ‘chemically reasonable charges’. Charges of buried atoms only affect the ESP slightly and can therefore lead to situations where large and opposite charges on atoms near each other are obtained by the fitting process. This has been countered by adding penalty terms to the fitting process, such that only atoms that make significant contributions are allowed to have charges significantly different from zero.^{23,24} Other strategies to avoid artifacts from buried atoms include fitting to the molecular dipole or higher order molecular moments. Thole and van Duijnen have described a population

analysis that preserve the molecular dipole moment,⁴⁶ and based on this model, Zhang et al. have developed the dipole preserving and polarization consistent charges.⁴⁷ Truhlar and Cramer et al. have presented a set of charge models, which also reproduce the molecular dipole moment, where the partial charges are computed by a parametrized mapping procedure from different population methods.^{44,45} Including higher order molecular moments as restraints in the fitting process only makes small changes, as a near-exhaustive fitting of the ESP reproduces molecular moments closely without any restraints. The problem with ‘unreasonable’ atomic charges just reflects the near-redundancy of the parameter space, which is the focus of the present work, and it should be recognized that atomic charges are not physical observables; only the molecular ESP has physical significance. Addition of user-defined restraints will increase the fitting error and thus require more parameters for a given accuracy, while we focus on eliminating the parameter redundancy a priori. Our fitting surface is essentially the van der Waals surface, and it is therefore near the limit for how close molecules can approach each other. The electrostatic interaction is the only important long-range interaction, and that is fully accounted for by the present procedure. The real interactions at van der Waals distances and shorter will require dispersion and perhaps also charge penetration terms, in addition to the electrostatic interaction.

The $N_A - 1$ charges are obtained by solving the normal equations of the linear least-squares problem in eq 3.

$$\mathbf{A}\mathbf{q} = \mathbf{b} \quad (3)$$

Here, \mathbf{A} is a symmetric square matrix with dimensionality equal to the number of fitting parameters and with elements shown in eq 4, where n denotes the index of the eliminated charge. The \mathbf{q} vector contains the $N_A - 1$ unconstrained atomic point charges (q_k), while the \mathbf{b} vector contains all information about the reference ESP with elements given in eq 5.

$$A_{mk} = \sum_{p=1}^P (|\mathbf{r}_{kp}|^{-1} - |\mathbf{r}_{np}|^{-1}) (|\mathbf{r}_{mp}|^{-1} - |\mathbf{r}_{np}|^{-1}) \quad (4)$$

$$b_m = \sum_{p=1}^P \phi^{\text{QM}}(\mathbf{r}_p) (|\mathbf{r}_{mp}|^{-1} - |\mathbf{r}_{np}|^{-1}) \quad (5)$$

Solving eq 3 is trivial if \mathbf{A} is nonsingular. If \mathbf{A} is rank deficient, however, the fitting procedure is ill conditioned and not all charges can be statistically valid assigned. This is a result of the fitting points not containing enough diverse data to determine all charges. Increasing the number of fitting points will not have any effect after a certain limit, since the additional points will only contribute redundant data. It has been shown that 500–1000 points per atom is sufficient,²⁶ and the $0.20\text{-}a_0$ cubic grid that is used in the current work to generate the surface points results in roughly 3500 points per atom, which is significantly above this suggestion. The present formulation in terms of a surface integral ensures strict control over the error due to sampling points and that we capture all the available information.

Multipole Fit. The molecular ESP from an ab initio calculation can be reproduced exactly by a finite set of distributed multipole moments placed at a number of points located between atom pairs, as shown by Stone.⁴⁸ For computational efficiency, however, the points are usually restricted to atomic

positions, in which case the expansion in terms of electric multipole moments becomes infinite. It has been found that including up to quadrupole moments on all atomic positions usually provide a quite accurate representation. Terminating the distributed multipole moments calculated by the Stone procedure at the quadrupole level leads to errors arising from neglect of higher order moments, but some of these effects can be accounted for by fitting atomic multipole moments up to a given order to the reference ESP.

The standard charge fitting procedure (eqs 1–5) can be extended to include higher order multipoles. In this work Cartesian dipole and quadrupoles are employed, resulting in an approximated ESP given by eq 6.

$$\phi^{\text{app}}(\mathbf{r}_p) = \sum_{k=1}^{N_A} \frac{q_k}{|\mathbf{r}_{kp}|} + \frac{\boldsymbol{\mu}_k \cdot \mathbf{r}_{kp}}{|\mathbf{r}_{kp}|^3} + \frac{\mathbf{r}_{kp} \cdot \boldsymbol{\Theta}_k \cdot \mathbf{r}_{kp}}{|\mathbf{r}_{kp}|^5} \quad (6)$$

Here, the k^{th} atom is parametrized using a charge, q_k , a dipole tensor, $\boldsymbol{\mu}_k$ and a Cartesian quadrupole tensor, $\boldsymbol{\Theta}_k$. The dipole tensor has three components and the quadrupole tensor five independent components due to symmetry and the quadrupole moment being traceless. As a result, nine parameters are fitted per atom. The sum of all charges is still constrained to the molecular charge, resulting in a total number of fitting parameters of $(9N_A - 1)$. The normal equations are given in eq 7, where the \mathbf{A} matrix has a block structure with monopole, dipole, and quadrupole blocks, respectively. The elements of the different blocks of \mathbf{A} and \mathbf{b} are given in Table A1 of Appendix A.

$$\begin{bmatrix} A_{qq} & A_{q\mu} & A_{q\Theta} \\ A_{\mu q} & A_{\mu\mu} & A_{\mu\Theta} \\ A_{\Theta q} & A_{\Theta\mu} & A_{\Theta\Theta} \end{bmatrix} \begin{bmatrix} \mathbf{q} \\ \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_z \\ \boldsymbol{\Theta}_{xx} \\ \boldsymbol{\Theta}_{xy} \\ \boldsymbol{\Theta}_{xz} \\ \boldsymbol{\Theta}_{yy} \\ \boldsymbol{\Theta}_{yz} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \end{bmatrix} \quad (7)$$

Fitting multipole components up to quadrupoles requires inversion of a square matrix of dimensionality $(9N_A - 1)$. Since this is rather time-consuming, a computational cheaper fit has also been attempted based on Stone’s Distributed Multipole Analysis (DMA).⁴⁹ The DMA dipole and quadrupole moments are generated from the ω B97X-D/aug-pc-1 electronic density and normalized, in the sense that a dipole tensor is divided by its \mathbb{R}^3 -norm, while a quadrupole tensor is decomposed into a vector in \mathbb{R}^5 using only its independent components, and normalized correspondingly. Each normalized multipole is scaled by a single fitting parameter, and we will refer to these fits as DMA-scaled multipole fits. The DMA-scaling procedure constrains the direction of dipoles and the relative composition of quadrupole components, while the monopoles are unrestricted. The number of parameters is in this approach reduced to $(3N_A - 1)$, and the elements of \mathbf{A} and \mathbf{b} for this fit are given in Table A2 of Appendix A.

RESULTS

Limitations of Electrostatic Models. Most FF studies employ point charge electrostatics, and it is of interest to investigate the absolute limits for the performance of this model, as well as models including higher order multipole moments. The point charge fitting procedure described in the previous

section provides the set of atomic charges that best reproduces the reference ESP for each conformation, and the (approximate) ESP generated from such charges serves as an upper limit for the accuracy of the partial charge model in FF applications. Standard FFs rely on the assumption of transferability, meaning that parameters should be transferable between different conformations and perhaps also between different molecules, in terms of an atom type. This implies that the partial charges are calculated as a suitable average over structures and will therefore provide an ESP of lower accuracy compared to that from charges fitted to a specific conformation. Five levels of approximations are introduced here for the electrostatic model: two of them employ conformational averaged parameters, (1) standard FF point charges and (2) point charges fitted to all conformations simultaneously, while three of them employ parameters that are individually fitted to each molecular structure: (3) charges, (4) charges and dipoles, and (5) charges, dipoles, and quadrupoles. The CHARMM36 charges⁵⁰ have been used as level 1 parameters, but other additive protein FFs use very similar parameters.

The results are exemplified for a single structure of the alanine dipeptide, shown in Figure 1a. Figure 2a displays the

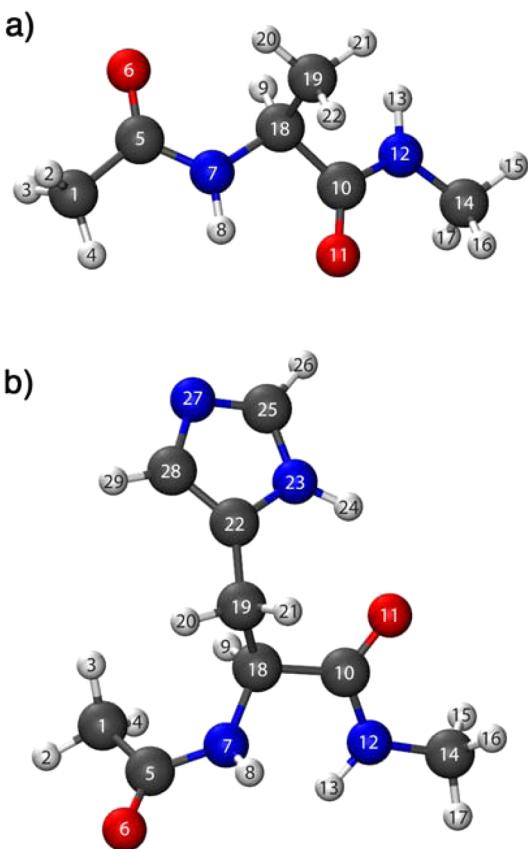


Figure 1. Atom numbering of (a) alanine and (b) histidine systems.

total ω B97X-D/aug-cc-pVQZ ESP for this conformation, while Figure 2b–f show the ESP difference between the ω B97X-D reference and the five levels of approximations (note the difference in the color scales between Figure 2a and Figure 2b–f). The corresponding Root Mean Square-values (RMS), $\sqrt{\chi^2}$, between the fitted potentials and the ω B97X-D/aug-cc-pVQZ reference are 6.46, 5.90, 4.98, 3.32, and $3.17 \times 10^{-3} \text{ E}_h/\text{e}$ for the five levels, respectively. Even though there is some visible error

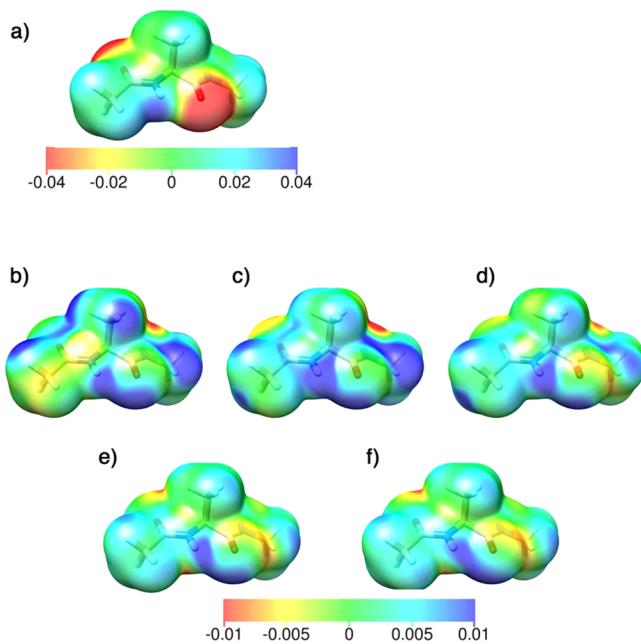


Figure 2. Electrostatic potential for a representative conformation of alanine plotted on the ω B97X-D/aug-cc-pVQZ isodensity surface corresponding to $\rho = 10^{-3} \text{ e}/\text{a}_0^3$. (a) Total ω B97X-D ESP and (b–f) ESP difference between ω B97X-D and approximated potentials: (b) CHARMM36 partial charges, (c) partial charges fitted to all 28 conformations simultaneously, (d) partial charges fitted to the specific conformation, (e) up to dipoles, and (f) up to quadrupoles.

left in the ESP at level 5 using the current color scale, the figures clearly indicate that the error is reduced with the levels of theory, and especially the addition of dipoles seems to give a significant improvement. This does not show that dipoles necessarily are the superior type of parameters, but merely an indication that going beyond the partial charge approximation results in a significant improvement. In general, the ESP errors are independent of the absolute value in the reference ESP.

Figure 3 plots the RMS-values between the approximate potentials and the ω B97X-D reference averaged over the set of conformations for alanine, glutamine, histidine, threonine, and tryptophan. These are chosen since they are representative and include both hydrophobic, polar, and aromatic side chains. Equivalent plots for the remaining 15 amino acids can be found in Figure SI.2 in Supporting Information. The reduction in RMS upon going from level 2 to level 3 indicates the importance of polarization, since the specific molecular geometry is taken into account. By including dipoles and quadrupoles (level 4 and 5), the RMS is reduced further, and it can be concluded that the effect of polarization and the need for higher order multipoles are of the same order of importance, since they give similar reductions in the RMS.

The reduction in the RMS-value by addition of quadrupole moments (from level 4 to level 5) is minor, suggesting that the convergence with respect to increase in multipole order will be slow. Stone has shown that a finite number of multipoles can represent the molecular ESP exactly when the expansion points are not restricted to nuclei centers.⁴⁹ We have performed multipole fits including up to quadrupole moments at all nuclei and additionally all bond-midpoints. These fits provide unphysical electrostatics parameters due to the issue of rank deficiency but without significantly lowering the RMS error. Considering also the residual error displayed in Figure 2f, this

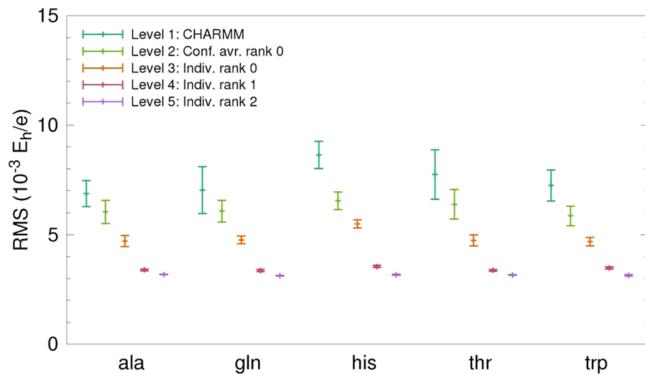


Figure 3. RMS between the approximated ESP and ω B97X-D reference for alanine (ala), glutamine (gln), histidine (his), threonine (thr), and tryptophan (trp) using the five levels of approximations: (1) CHARMM point charges, (2) point charges fitted to all conformations simultaneously, level 3–5 are fitted to individual conformations, where (3) uses point charges only, (4) uses charges and dipoles, and (5) uses charges, dipoles, and quadrupoles. RMS-values are averaged over the set of conformations and shown with standard deviation. Corresponding plot for the remaining 15 amino acids are shown in Supporting Information Figure SI.1.

suggests that a further reduction of the RMS-value will require high multipole orders or expansion points placed outside the bonding framework of the molecule. Popelier et al. have studied the convergence of a nuclei-centered multipole expansion for several systems, where the multipole moments are derived using the quantum chemical topology method,^{51,52} by partitioning the electron density into atomic basins and integrating over the atomic volumes. Convergence behaviors have been quantified using the interaction rank, $L = l_A + l_B + 1$, where l_A and l_B are the multipole ranks of atoms A and B, respectively. It has been shown that $L = 5$ is sufficient both for obtaining the correct behavior of the pair correlation function in simulations of liquid water⁵³ and for converging the electrostatic energy for 1,4-interactions in glycine.⁵⁴ An interaction rank of $L = 5$ requires up to hexadecapoles ($l = 4$), and even though the multipoles are derived by a different strategy than in the current work, this suggests that multipoles up to quadrupole level is not sufficient to converge the molecular ESP representation.

Redundancy among Electrostatic Parameters. It is well-known that partial charges for atoms near the molecular surface are more important for representing the ESP than atoms buried within the molecule, as is also evident from eq 4. Even for medium sized molecules, the equations for fitting partial charges become near-singular,^{25,26} which implies that the charge parameters are near-redundant. In this section, we investigate the level of redundancy for both a point-charge fit and a multipole fit including up to quadrupoles moments.

Redundancy Among Charges. For the point-charge fit, we employ a top-down approach to quantify the importance of each partial charge. Starting from the full set of N_A charges, the least important charge is identified as the one leading to the smallest increase in the χ^2 -function, when all remaining charges are refitted. This procedure is used recursively and requires t fitting equation to be solved at level t where point charges are fitted to (only) t atoms. The algorithm produces $N_A - 1$ different truncation levels, going from $t = N_A$ where the entire set of charges are populated, to $t = 2$ using only two charges. If $t = 1$ is used, the single charge would be constrained to have a value of 0, and this situation would therefore be identical to $t = 0$.

There is no reason to expect that the top-down approach will lead to the globally best set of parameters at a given truncation level. The only way to find these is to perform an exhaustive search, and this was done at the 10-charge truncation level for the conformation of alanine shown in Figure 1a. The global search is a combinatorial problem of selecting and fitting 10 charges at 22 sites, resulting in $22!/(10!(22-10)!) = 646646$ point-charge fits, of which the best one has a RMS ($\sqrt{\chi^2}$) of $6.6 \times 10^{-3} E_h/e$. The 10-charge fit using the top-down approach has RMS = $8.4 \times 10^{-3} E_h/e$, corresponding to a ranking of 1652 or within the top 0.3% of all combinations. While other cases may perform differently, we believe that the top-down approach will generate a subset of charges that is a near-optimum reduction of the full set at each truncation level, and in the following, it will be used as an upper bound for the accuracy at a given truncation level. A global optimization method could be employed to search for parameter sets closer to the global best, and this could potentially lead to even larger parameter reductions for a given accuracy in terms of the RMS, but that will only strengthen the conclusions.

Figure 4 shows how the RMS-value increases as the number of parameters is reduced (from right to left in the plot) by the

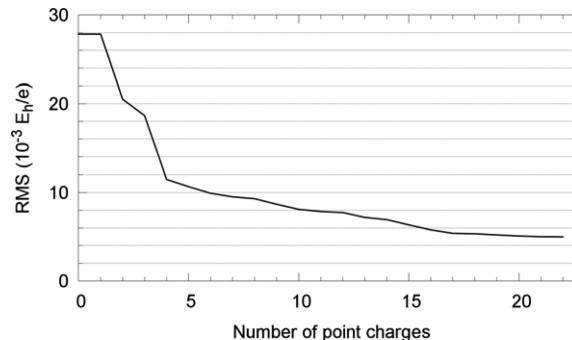


Figure 4. RMS as a function of the number of point charges used in the fit for alanine (structure shown in Figure 1a).

top-down procedure for the alanine structure in Figure 1a. The flat slope in the right-hand side of the plot implies that the set of point charges is near-redundant and that a number of charges can be eliminated without significantly affecting the ESP accuracy.

The parametrization using two-thirds of the charges will be used to further quantify the redundancy. This corresponds to 15 of 22 charges for alanine, which is a reduction in the parameter space that causes the RMS to increase from $5.0 \times 10^{-3} E_h/e$ to $6.3 \times 10^{-3} E_h/e$. Figure 5 compares the RMS of the full point-charge fit (red) with the 67%-point-charge fit (purple) averaged over the set of conformations. The RMS-values are highly system dependent and differ by up to 33% between different systems, as a consequence of the incomplete parameter space. Also, the extent to which the parameter space can be reduced varies for the amino acids and seems uncorrelated to the accuracy of the full point-charge fit.

Redundancy Among Multipoles. The 20 amino acid models contain from 19 to 36 atoms, and the full component multipole fits have 170–323 fitting parameters, while the number of parameters is 56–107 in the DMA-scaled fits. The same conformation of alanine (Figure 1a) is again used as a representative example for the general trends of these fits.

The large number of parameters in both the full component and the DMA-scaled multipole fits causes increased redundancy

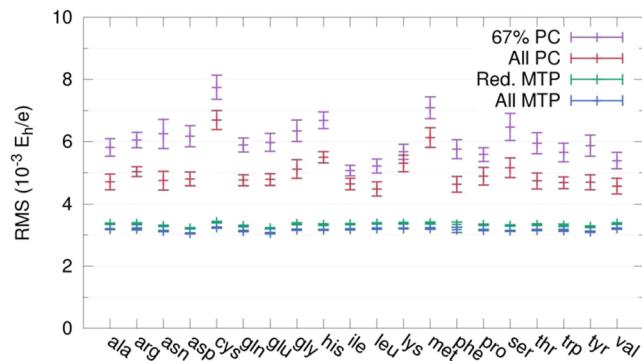


Figure 5. RMS-values averaged over all conformations and shown with standard deviations for parametrizations using either 66% of the point charges (purple), all point charges (red), a reduced set of multipoles (green), or all multipoles (blue).

compared to the point-charge fit. The redundancy is easily visible when attempting to use the top-down approach for ranking the importance of the multipole parameters. Starting from a full set of atomic multipoles, any multipole moment can be removed in the first few steps without affecting the quality of the ESP fit, since the remaining set of refitted parameters can compensate for the missing parameter(s). The top-down approach therefore fails to identify an order of importance and instead produces an almost random ranking of the multipoles, and we have instead employed a two-stage approach to reduce the number of parameters. Since atomic monopoles are considered to be more important than higher order moments, we start from a model employing one charge parameter per atom. Higher order multipoles are added one at a time using a recursive procedure, where the new multipole is selected as the one leading to the largest decrease in χ^2 , when all parameters are refitted. The addition of multipoles continues until χ^2 is lower than a preset threshold, for which we have used a value corresponding to 5% above the best χ^2 -value (corresponding to a nonreduced parameter set). In the second stage, the parameter set composed of N_A atomic point charges and the atomic multipoles selected in the first stage, is reduced by the top-down approach, until a second threshold is fulfilled. The value must be larger than the first threshold, and we employ a value of 10% above the limiting value. This procedure identifies a selected set of atomic multipoles, which is capable of representing the ESP to within $\sim 10\%$ of the accuracy possible by employing a full set of all multipole moments. We will refer to this set of multipoles as the reduced multipole level.

Figure 6a and b shows the RMS as a function of the number of multipoles in each step of the algorithm for alanine using the full component and the DMA-scaled procedures, respectively. Note that the horizontal axis is the total number of multipole moments, where each moment in Figure 6a employs one, three, or five components depending on the rank, but always one parameter in Figure 6b. The blue 'x' corresponds to a non-reduced parameter set, while the orange 'x' corresponds to a generalized reduced set of multipoles that will be discussed in the following section. In Figure 4, the RMS-value is reduced from $\sim 27.8 \times 10^{-3} E_h/e$ (no parameters, i.e., a zero-ESP) to $\sim 5.0 \times 10^{-3} E_h/e$ (full point-charge fit), and Figure 6 can be considered as the extension of Figure 4, starting from this point-charge fit and reducing the RMS-value further by addition of higher order moments. The multipole parameter space is highly redundant using both the full component and the

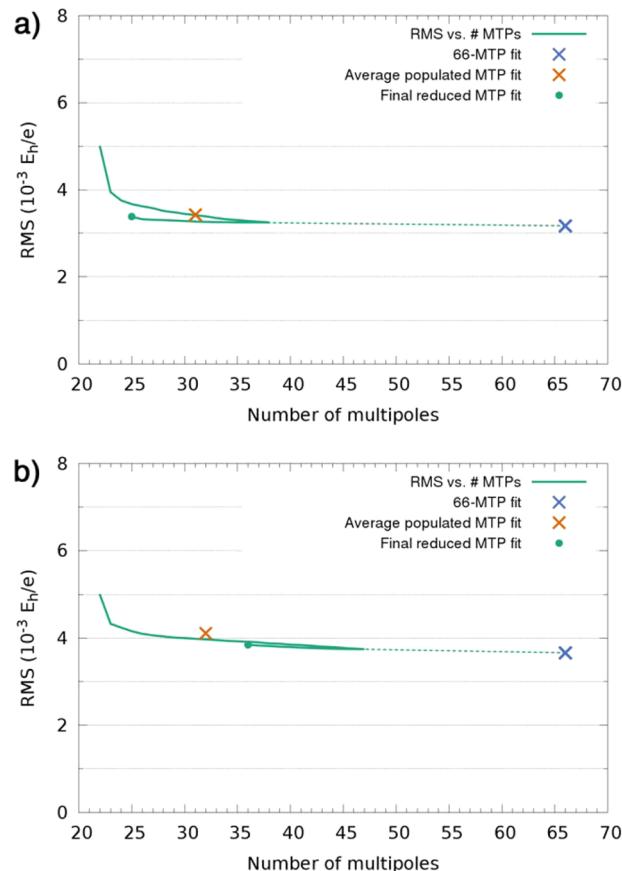


Figure 6. RMS as a function of the number of multipole (MTP) moments using the two-stage algorithm for a single conformation of alanine (green). The final reduced fit is shown with a dot. The blue 'x' corresponds to the RMS-value that can be obtained when all 66 multipole moments (up to quadrupoles) are included in the fit. The orange 'x' shows the performance of a fit where the selected multipoles are the average most populated based on Figure 9a and 10a population plots. (a) Full component fit and (b) DMA-scaled fit.

DMA-scaled approaches. The transition from stage 1 to stage 2 is found at 38 multipoles in Figure 6a and at 47 multipoles in Figure 6b, where the corresponding fit provides a RMS-value within 5% of the nonreduced fit. Practically no information is gained by expanding the parameter space beyond this point, as represented by the dashed line. In particular, for the full component fit (Figure 6a), the final reduced set of parameters provides an accurate representation of the ESP for only a small increase in the number of multipoles over the number of charges in a point-charge fit (25 vs 22). On the other hand, the DMA-scaled multipole fit (Figure 6b) results in a larger RMS-value as a result of the constrained multipole moments.

Figure 7 visualizes the ESP difference between a full 66-multipole fit and the reduced multipole fit where only 25 multipoles are populated, using the same color scale as Figure 2b–f. It is clearly evident that the error arising from ignoring 41 of the multipoles is very minor.

In Figure 5, the conformational averaged RMS-values of both the reduced and the nonreduced multipole fits are compared to the point-charge fits. While the RMS-values are system dependent for the point-charge fits, the accuracy that can be obtained using a multipole fit is essentially system independent having a RMS of roughly $3.0 \times 10^{-3} E_h/e$ for all amino acids. This convergence among systems further supports that we have

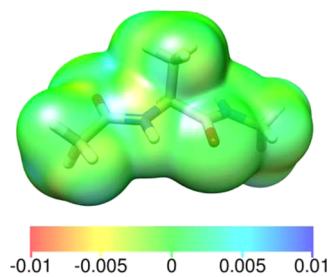


Figure 7. Difference between the full ESP generated from all 66 multipoles and that using only 25 multipoles, of which 10 are charges, 4 are dipoles, and 11 are quadrupoles. The surface corresponds to $\rho = 10^{-3} \text{ e}/a_0^3$.

reached the limit of ESP accuracy that can be obtained by atomic-centered multipoles.

Figure 8 plots the RMS as a function of the number of multipole moments for three selected fits; the full point charge

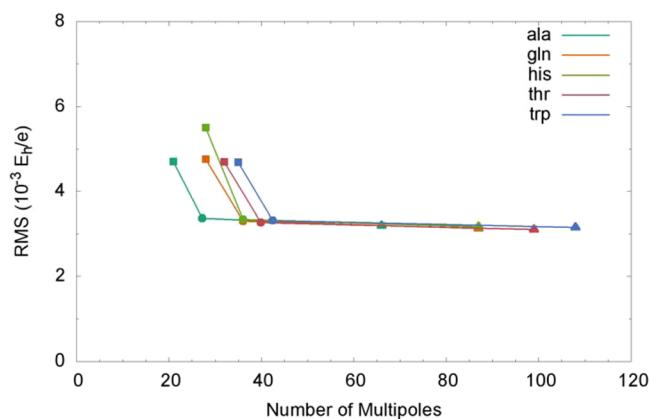


Figure 8. RMS-values averaged over all conformations for three different multipole fits, using partial charges only (squares), a reduced set of charges and multipoles (circles), and all multipoles up to quadrupoles (triangles).

model (squares), the reduced multipole model (circles), and the full multipole model (triangles). A significant number of multipoles can be removed for a negligible increase in the RMS upon going from the full to the reduced multipole fit. On average, $\sim 40\%$ of the multipoles are populated in the reduced multipole model, of which 59% are monopoles, 8% dipoles, and 33% quadrupoles. The relative high ratio of monopoles is a consequence of starting from the point charge model and does not imply that monopoles are more important compared to higher order moments. This is validated by additional recursive fits where the initial model employs all dipoles or all quadrupoles, respectively. The relative composition of the multipole rank for alanine during the two-stage procedure is plotted in Figure 9a–c, demonstrating that the composition of the final reduced fit is biased by the initial model. The final set of multipole moments is clearly dependent on the initial selection of parameters, and the models starting from dipoles or quadrupoles result in parameter sets with no charges populated. When all charges are eliminated they will not be populated again by the employed procedure, since they are added one-by-one and the first charge will have a value of zero due to the molecular charge restraint. Figure 9 shows that there is considerable freedom in choosing a mixture of charge, dipole, and quadrupole parameters to provide a fit of a given accuracy.

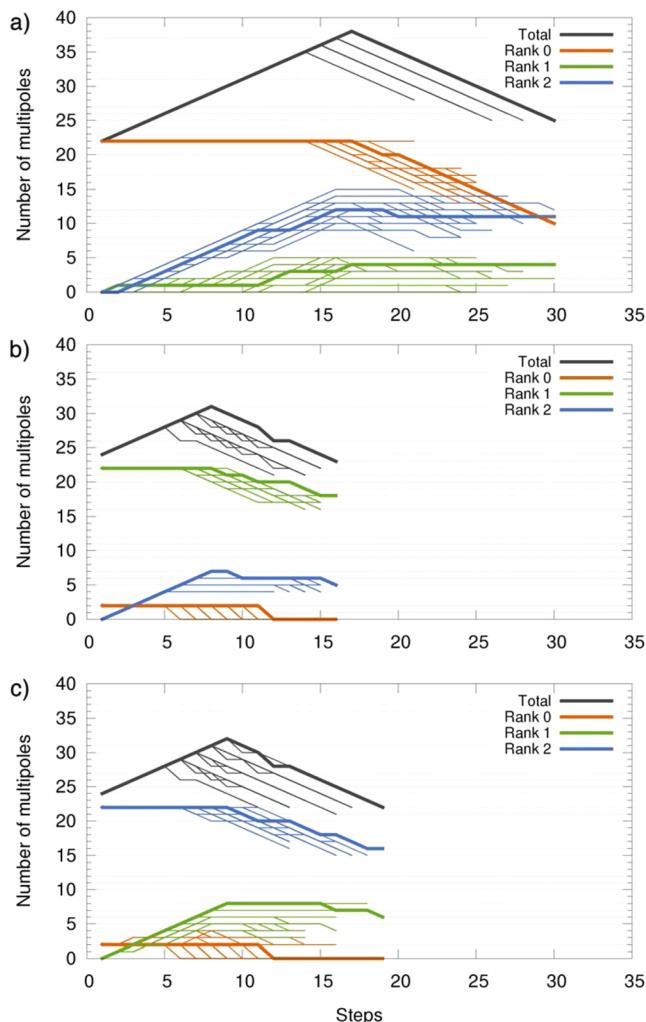


Figure 9. Number of monopoles, dipoles, quadrupoles, and all multipoles shown as a function of the number of steps in the two-stage procedure for the conformations of alanine. Part a corresponds to starting from all monopoles (as described in the paper), while b and c use an initial model of all dipoles or all quadrupoles, respectively. Bold lines correspond to the conformation of alanine shown in Figure 1a, while thin lines represent the remaining 27 conformations (several of them are overlaying). Monopoles are treated different compared to dipoles and quadrupoles, as the sum of all monopoles is constrained to the total molecular charge. At least two monopoles are consequently needed in the fit, and the fits starting from dipoles and quadrupoles do therefore additionally employ charges for C_α and H_α for the initial fit.

The apparent importance of a given type of parameter therefore depends on the exact procedure used for adding (bottom-up) or eliminating (top-down) parameters. Computational considerations favor low order multipole moments (monopoles), and this motivated the present approach starting from a full set of point charges. By comparing the full point-charge fit to the reduced multipole fit in Figure 8, it is evident that a significant gain in accuracy can be obtained by only increasing the number of multipole moments slightly. This is possible because the inclusion of higher order moments provides an expansion of the parameter space, and thus, the reduced set of multipole parameters is larger but still less redundant than the full point charge model.

While the 5% and 10% criteria employed for terminating the first and second stage fits are somewhat arbitrary, the results clearly

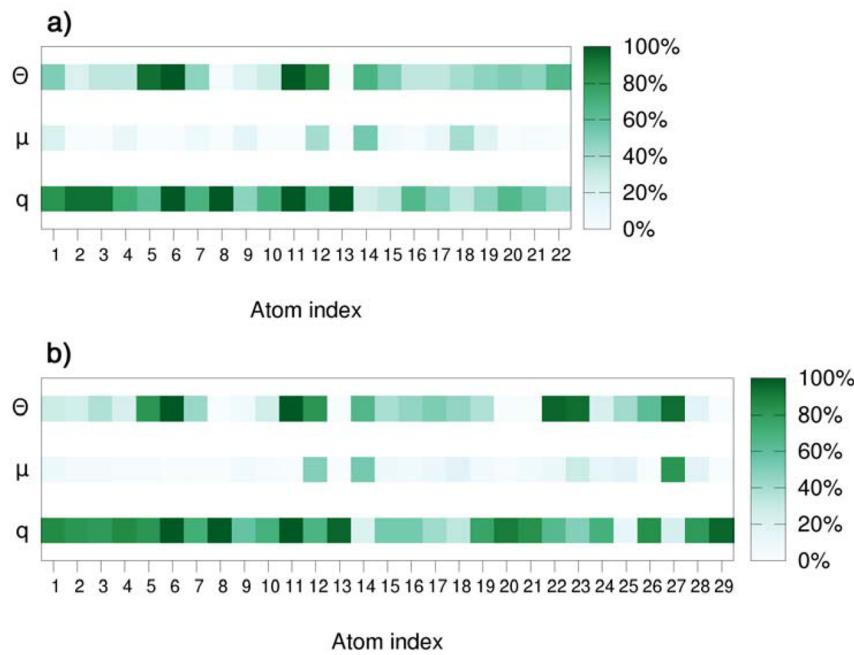


Figure 10. Full component population plot for (a) alanine and (b) histidine, showing the percent-wise populated multipole sites at the reduced multipole level averaged over all conformations.

show that ~60% of the higher order multipole moments are near-redundant and can be removed without affecting the resulting ESP significantly. We emphasize that the reduced set of multipoles is not guaranteed to be the global optimum set of parameters, and the ~60% therefore corresponds to a lower bound.

The reduction in the parameter space presented above is performed by removing single charges or multipole moments. For the point-charge fit we have also attempted a different strategy, where a linear combination of charges is eliminated in each step of the top-down approach. We have employed multivariable data analysis methods such as principal component analysis and partial least-squares to identify a set of statistically important combinations of charges, which subsequently are used for the elimination. However, it was found that the simple approach of truncating the direct charge parameter space was superior to eliminations in the rotated vectors spaces with regard to producing a low χ^2 -value at a given truncation level.

Conformational Transferability. The previous sections have quantified the limits for how accurate a given set of atomic multipoles can represent the ESP for a given geometry. However, one of the central FF concepts is the assumption of transferability, meaning that the ESP representation must be capable of modeling a range of geometries with a common set of parameters. It is therefore of interest to probe how similar a reduced set of multipoles are among different conformations, or equivalently, whether it is the same atomic multipole moments that are important for different conformations. If the reduction in the parameter set should be applied to FFs, the general assumption of transferability of parameters should still be valid. Since dipoles and quadrupoles are directionally dependent, their values can only be compared after rotating to a local coordinate system, which has not been done in the present case. We have instead focused on to what extend the same multipole sites are populated for the different molecular structures, as this is a prerequisite requirement for selecting a conformational consistent set of reduced multipoles.

The phrase ‘multipole site’ is used for either a monopole, dipole, or quadrupole at a given atomic center.

A population plot is generated for all amino acids to demonstrate how populated a given multipole site is on average over the set of conformations at the reduced multipole level described above. Figure 10a and b demonstrate two plots for alanine and histidine, respectively. A dark green spot corresponds to a multipole site that is widely populated among the different conformations and indicates high importance with respect to reproducing the ESP, independent of the molecular geometry. The atom numbering is shown in Figure 1, and it should be noted that the backbone atoms (1–18) are numbered consistently for the different amino acids. Figure 10 shows that monopoles seem to be the most important multipole type, and this is in agreement with results for the other 18 amino acids. The fact that monopoles are on average most populated is a result of the parameter reduction method, as mentioned above and shown in Figure 9. Some multipole sites are highly populated on average while others only rarely populated, and we can therefore conclude that the algorithm is able to identify a systematic set of parameters fairly independent of the geometry. If the selection was purely random, one would expect most multipoles to have a population of ~40% (42% for alanine, 41% for histidine), since this is how many of the $3N_A$ multipole moments that are populated on average at the reduced multipole level.

Among the backbone atoms of the alanine and histidine models, quadrupole moments are widely selected for the N-methyl amide carbon and oxygen (atoms 5 and 6) and for the C-acetyl oxygen and nitrogen (atoms 11 and 12), indicating the importance of higher order moments at these atoms. Furthermore, most histidine conformations carry quadrupole moments at one of the carbons and the two nitrogens in the imidazole ring (atoms 22, 23, and 27), and the N_e (atom 27) also has a highly populated dipole moment. The importance of higher order moments is in general observed on atoms participating in π -orbitals, which generate an anisotropic electric field that is poorly described by monopoles or dipoles.

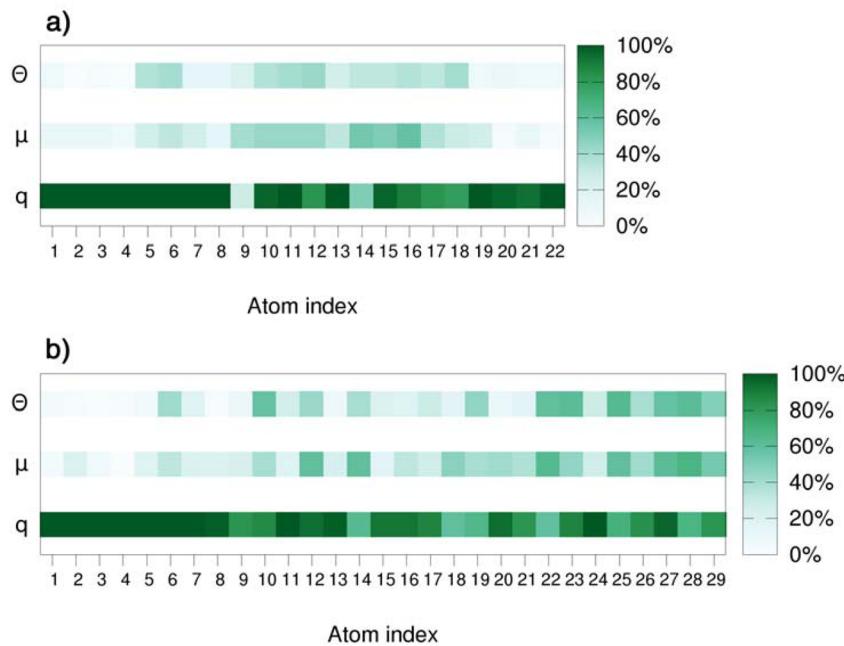


Figure 11. DMA-scaled population plots for (a) alanine and (b) histidine, showing the percent-wise populated multipole sites at the reduced multipole level averaged over all conformations.

Figure 11a and b show the corresponding population plots for the DMA-scaled fit of alanine and histidine, respectively. Contrary to the full component fit, the DMA-scaled fits predict that only monopole sites are widely selected across the set of conformations. The fact that dipoles, and especially quadrupoles, are more populated in the full component fit is a result of an inappropriate locked combination of individual components used in the DMA-scaled fit. The monopoles are the only parameters not restricted in the DMA-scaled fit, which make them more transferable.

Figures 10 and 11 signify the importance to avoid restriction in the combination of individual multipole components when reducing the multipole parameter space, and especially that scaled DMA dipoles and quadrupoles involve a restriction, which results in a poor degree of transferability. Combined with the increased RMS-values for individual DMA-scaled fit observed in Figure 6b, this emphasizes the importance of fitting individual components of the dipole and quadrupole moments. It also implies that dipole and perhaps also quadrupole polarizability will be important for modeling the conformational dependence of multipole moments.

Figure 6a and b compares RMS-values of the reduced multipole fit, where the multipoles are selected for the specific conformation (green), to a fit employing a similar amount of multipoles selected as the average most populated (orange 'x') based on the population plots in Figures 10 and 11. For the full component fit, the average selection of sites provides an ESP accuracy at the same level as individual multipole selections, suggesting a large degree of transferability for the important multipole sites. On the other hand, the DMA-scaled fit in Figure 6b shows a larger deviation between the generalized and individual multipole site selections. These trends apply for all amino acids in this study.

The number of multipoles that is needed in the reduced multipole fit to keep χ^2 within $\sim 10\%$ compared to a non-reduced fit is dependent on the specific conformation, and its average is taken as a threshold, $\langle N_{MTP,10\%} \rangle$. The $\langle N_{MTP,10\%} \rangle$ most populated multipole sites are subsequently selected based on the population plots. However, this cannot necessarily be

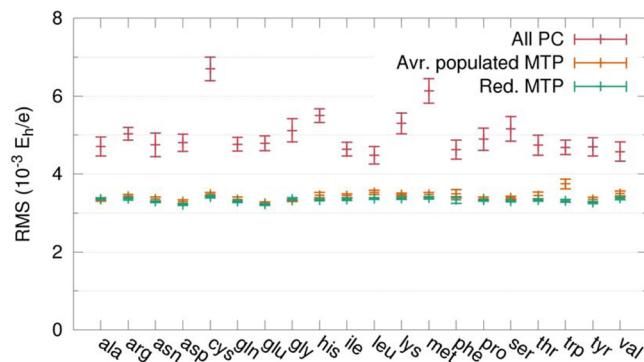


Figure 12. Averaged RMS-values for point charge (PC) and full component multipole (MTP) fits. 'Red. MTPs' is the reduced multipole fit obtained by the two-stage procedure. 'Avr. populated MTPs' refers to a fit populating roughly the same amount of parameters as 'Red. MTPs', but the selection of multipole sites is taken as the average most populated multipoles.

done uniquely, since several multipole sites might be equally populated. The reduced multipole fit for alanine on average populates 27 of the 66 multipoles. However, the 25th to 31st most used sites are equally populated (46.4% corresponding to 13 of the 28 conformations). In this case, $\langle N_{MTP,10\%} \rangle$ is increased from 27 to 31 in order to make a unique selection, as also seen in Figure 6a where the average populated multipole fit uses 31 moments. At the reduced level $\sim 40\%$ of the multipole sites are populated on average, while $\langle N_{MTP,10\%} \rangle$ is $\sim 44\%$ of the number of atomic multipole moments.

Figure 12 compares the RMS-values for the fits employing the $\langle N_{MTP,10\%} \rangle$ average most populated multipole sites (orange) to the point-charge fits (red) and the reduced multipole fits (green), respectively, where the latter two were also presented in Figure 5. The individually selected multipole sites only performs slightly better compared to the sites chosen from the average population, implying that the same multipole sites are important independent of the conformation.

CONCLUSION

Analysis of point charge and multipole fits to the molecular ESP for ~1000 conformations for the 20 natural amino acids show that the point charge representation is inferior to mathematical more flexible electrostatic models, but paradoxically, the parameter space is also near-redundant, and one-third of the charges can be eliminated for only a small decrease in accuracy with respect to reproducing the reference ESP.

Extending the electrostatic parameters to include dipoles and quadrupoles increases the accuracy of the ESP, as well as the number of near-redundant parameters. We have shown that a subset of multipoles only slightly larger (~20%) than the number of atoms represents the ESP much more accurately than the conventional point charge model. Including all multipoles up to quadrupoles results in a highly redundant set of parameters, which is not computationally efficient. At least 60% of the multipole moments can be removed for only a negligible reduction in the ESP accuracy. Since this result is not based on an exhaustive search of the reduced parameter space, it serves as a lower bound, and the same accuracy might be reachable employing an even smaller set of parameters. An atom-centered multipole expansion up to quadrupoles implies that the RMS with respect to the reference ESP converges to $\sim 3.0 \times 10^{-3} E_h/e$. The atom-centered multipole expansion is slowly convergent and higher order multipoles or nonatomic expansion centers may be required to improve the ESP accuracy further.

The reduction in ESP error upon going from a point-charge fit where all conformations are fitted simultaneously to an individual fit is of the same order of magnitude as going from a point-charge to a multipole fit. This is consistent with previous studies⁵⁵ and suggests that polarization and higher order multipoles should be introduced at the same level in FF developments.

Multipole moments calculated by Stone's DMA approach have been rescaled to better fit the ab initio ESP; however, the DMA-constraint results in a larger deviation to the reference ESP compared to a full component fit. Truncation in the DMA-scaled parameter space eliminates mainly dipoles and quadrupoles, since these are the directionally constrained parameters. A reduced set of DMA-scaled multipoles, therefore, primarily consists of monopoles, and hence providing a poor representation of the reference ESP. The fact that higher order moments are less transferable within the DMA-constraint reflects the need for dipole and perhaps also quadrupole polarization.

For the full component fit, we find that all orders of multipoles are transferable. Monopoles are, in general, most important, but quadrupoles are also important, particularly for atoms involved in π -bonding, since the anisotropic electric field generated from such orbitals are poorly described by isotropic point charges.

In the current work, we have treated each atom individually, even though the convention in the FF community is to use the concept of atom types. However, Popelier et al. have shown that atom types are more or less arbitrary defined,⁵⁶ and for example, the AMBER atom types are ill balanced by sometimes being under-differentiated and other times being overdifferentiated. A protein FF only needs parameters for the 20 natural amino acids in their different protonation states, which is a limited number of atoms that easily can have assigned individual parameters, and there is no reason to restrict the parameter space by introducing atom types.

APPENDIX A

The elements of the **A** matrix and the **b** vector are found by analytical differentiation of eq 1 in the paper, where the approximated ESP, $\phi^{app}(\mathbf{r}_p)$, is given by eq 6. Tables A1 and A2

Table A1. For the Full Component Fit the Monopole Blocks Have $N_A - 1$ Rows/Columns, the Dipole Blocks Have $3N_A$ Rows/Columns (for the x , y , and z Components), and the Quadrupole Blocks Have $5N_A$ Rows/Columns Corresponding to the Independent xx , xy , xz , yy , and yz Components^a

R_m	R_k	A	b
0	0	$A_{m,k} = \sum_p^P \left(\frac{1}{\mathbf{r}_{kp}} - \frac{1}{\mathbf{r}_{np}} \right) \left(\frac{1}{\mathbf{r}_{mp}} - \frac{1}{\mathbf{r}_{np}} \right)$	$b_m = \sum_p^P \phi^{QM}(\mathbf{r}_p) \left(\frac{1}{\mathbf{r}_{mp}} - \frac{1}{\mathbf{r}_{np}} \right)$
0	1	$A_{m,k\alpha} = \sum_p^P \frac{\mathbf{r}_{kp}^\alpha}{ \mathbf{r}_{kp} ^3} \left(\frac{1}{\mathbf{r}_{mp}} - \frac{1}{\mathbf{r}_{np}} \right)$	
0	2	$A_{m,k\alpha\beta} = \sum_p^P \frac{f(r_{kp}^\alpha r_{kp}^\beta)}{ \mathbf{r}_{kp} ^5} \left(\frac{1}{\mathbf{r}_{mp}} - \frac{1}{\mathbf{r}_{np}} \right)$	
1	1	$A_{ma,k\beta} = \sum_p^P \frac{\mathbf{r}_{kp}^\beta}{ \mathbf{r}_{kp} ^3} \frac{\mathbf{r}_{mp}^\alpha}{ \mathbf{r}_{mp} ^3}$	$b_{m\alpha} = \sum_p^P \phi^{QM}(\mathbf{r}_p) \frac{\mathbf{r}_{mp}^\alpha}{ \mathbf{r}_{mp} ^3}$
1	2	$A_{ma,k\beta\gamma} = \sum_p^P \frac{f(r_{kp}^\beta r_{kp}^\gamma)}{ \mathbf{r}_{kp} ^5} \frac{\mathbf{r}_{mp}^\alpha}{ \mathbf{r}_{mp} ^3}$	
2	2	$A_{maf\beta,k\gamma\delta} = \sum_p^P \frac{f(r_{kp}^\gamma r_{kp}^\delta)}{ \mathbf{r}_{kp} ^5} \frac{f(r_{mp}^\alpha r_{mp}^\beta)}{ \mathbf{r}_{mp} ^5}$	$b_{maf\beta} = \sum_p^P \phi^{QM}(\mathbf{r}_p) \frac{r_{mp}^\alpha r_{mp}^\beta}{ \mathbf{r}_{mp} ^5}$

^aThe indices α , β , γ , and δ are used as Cartesian indices running over x , y , z . $f(r^\alpha r^\beta)$ is a function taking the symmetry of the quadrupole tensor into account and is defined as

$$f(r^\alpha r^\beta) = \begin{cases} 2r^\alpha r^\beta & \text{for } \alpha \neq \beta \\ r^\alpha r^\beta - (r^z)^2 & \text{for } \alpha = \beta \end{cases}$$

Table A2. For the DMA-Scaled Fits the Monopole Blocks Have $N_A - 1$ Rows/Columns, While the Dipole and Quadrupole Blocks Both Have N_A Rows/Columns

R_m	R_k	A	b
0	0	$A_{m,k} = \sum_p^p \left(\frac{1}{\mathbf{r}_{kp}} - \frac{1}{\mathbf{r}_{np}} \right) \left(\frac{1}{\mathbf{r}_{mp}} - \frac{1}{\mathbf{r}_{np}} \right)$	$b_m = \sum_p^p \phi^{QM}(\mathbf{r}_p) \left(\frac{1}{\mathbf{r}_{mp}} - \frac{1}{\mathbf{r}_{np}} \right)$
0	1	$A_{m,k} = \sum_p^p \frac{\mathbf{r}_{kp}\boldsymbol{\mu}_k}{ \mathbf{r}_{kp} ^3} \left(\frac{1}{\mathbf{r}_{mp}} - \frac{1}{\mathbf{r}_{np}} \right)$	
0	2	$A_{m,k} = \sum_p^p \frac{\mathbf{r}_{kp}\Theta_k\mathbf{r}_{kp}}{ \mathbf{r}_{kp} ^5} \left(\frac{1}{\mathbf{r}_{mp}} - \frac{1}{\mathbf{r}_{np}} \right)$	
1	1	$A_{m,k} = \sum_p^p \frac{\mathbf{r}_{kp}\boldsymbol{\mu}_k}{ \mathbf{r}_{kp} ^3} \frac{\mathbf{r}_{mp}\boldsymbol{\mu}_m}{ \mathbf{r}_{mp} ^3}$	$b_m = \sum_p^p \phi^{QM}(\mathbf{r}_p) \frac{\mathbf{r}_{mp}\boldsymbol{\mu}_m}{ \mathbf{r}_{mp} ^3}$
1	2	$A_{m,k} = \sum_p^p \frac{\mathbf{r}_{kp}\Theta_k\mathbf{r}_{kp}}{ \mathbf{r}_{kp} ^5} \frac{\mathbf{r}_{mp}\boldsymbol{\mu}_m}{ \mathbf{r}_{mp} ^3}$	
2	2	$A_{m,k} = \sum_p^p \frac{\mathbf{r}_{kp}\Theta_k\mathbf{r}_{kp}}{ \mathbf{r}_{kp} ^5} \frac{\mathbf{r}_{mp}\Theta_m\mathbf{r}_{mp}}{ \mathbf{r}_{mp} ^5}$	$b_m = \sum_p^p \phi^{QM}(\mathbf{r}_p) \frac{\mathbf{r}_{mp}\Theta_m\mathbf{r}_{mp}}{ \mathbf{r}_{mp} ^5}$

provide elements for the full component and the DMA-scaled fits, respectively. R_m and R_k are the rank of the m and k indices, where a rank of 0 corresponds to a monopole, a rank of 1 is a dipole, and a rank of 2 is a quadrupole.

ASSOCIATED CONTENT

Supporting Information

Convergence of the electrostatic potential surface integral and performance of the five levels of approximations for all amino acids. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: frj@chem.au.dk.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Support from the Danish Center for Scientific Computation, Danish Natural Science Research Council, is gratefully acknowledged.

REFERENCES

- (1) Jensen, F. *Introduction to Computational Chemistry*; Wiley: New York, 2007; p 22–77.
- (2) Cramer, C. J. *Essentials of Computational Chemistry*. John Wiley & Sons, Ltd: New York, 2004; p 17–67.
- (3) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197.
- (4) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65* (3), 712–725.
- (5) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; seek, J. D. E.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prothom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (6) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25* (13), 1656–1676.
- (7) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- (8) Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. OPLS all-atom force field for carbohydrates. *J. Comput. Chem.* **1997**, *18* (16), 1955–1970.
- (9) Halgren, T. A.; Damm, W. Polarizable force fields. *Curr. Op. in Struct. Biol.* **2001**, *11*, 236–242.
- (10) Rick, S. W.; Stuart, S. J. Potentials and algorithms for incorporating polarizability in computer simulations. *Rev. Comput. Chem.* **2003**, *18*, 89–146.
- (11) Ponder, J. W.; Case, D. A. Force Fields for Protein Simulations. *Protein Simulations* **2003**, *66*, 27–85.
- (12) MacKerell, A. D.; Lopes, P. E. M.; Roux, B. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: Theory and applications. *Theor. Chem. Acc.* **2009**, *124*, 11–28.
- (13) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046–4063.
- (14) Wang, J.; Cieplak, P.; Li, J.; Hou, T.; Luo, R.; Duan, Y. Development of polarizable models for molecular mechanical calculations i: parameterization of atomic polarizability. *J. Phys. Chem. B* **2011**, *115* (12), 3091–3099.
- (15) Wang, J.; Cieplak, P.; Li, J.; Wang, J.; Cai, Q.; Hsieh, M.; Lei, H.; Luo, R.; Duan, Y. Development of polarizable models for molecular mechanical calculations ii: induced dipole models significantly improve accuracy of intermolecular interaction energies. *J. Phys. Chem. B* **2011**, *115* (12), 3100–3111.
- (16) Xie, W.; Pu, J.; MacKerell, A. D.; Gao, J. Development of a polarizable intermolecular potential function (PIPF) for liquid amides and alkanes. *J. Chem. Theory Comput.* **2007**, *3* (6), 1878–1889.
- (17) MacKerell, A. D.; Baker, C. M.; Anisimov, V. M. Development of CHARMM polarizable force field for nucleic acid bases based on the classical drude oscillator model. *J. Phys. Chem. B* **2011**, *115*, 580–596.
- (18) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. Polarizable Force Field for Peptides and Proteins

- Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013**, *9* (12), 5430–5449.
- (19) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand–Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3* (6), 1960–1986.
- (20) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.* **1994**, *101* (7), 6141–6156.
- (21) Stern, H. A.; Rittner, F.; Berne, B. J.; Friesner, R. A. Combined fluctuating charge and polarizable dipole models: Application to a five-site water potential function. *J. Chem. Phys.* **2001**, *115* (5), 2237–2251.
- (22) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* **2014**, *10* (10), 4497–4514.
- (23) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97* (40), 10269–10280.
- (24) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **1993**, *115* (21), 9620–9631.
- (25) Franc, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M. Charges fit to electrostatic potentials. II. Can atomic charges be unambiguously fit to electrostatic potentials? *J. Comput. Chem.* **1996**, *17* (3), 367–383.
- (26) Franch, M. M.; Chirlian, L. E. The pluses and minuses of mapping atomic charges to electrostatic potentials. *Rev. Comput. Chem.* **2000**, *14*, 1–31.
- (27) Zeng, J.; Duan, L.; Zhang, J. Z. H.; Mei, Y. A numerically stable restrained electrostatic potential charge fitting method. *J. Comput. Chem.* **2013**, *34* (10), 847–853.
- (28) Wang, B.; Truhlar, D. G. Partial atomic charges and screened charge models of the electrostatic potential. *J. Chem. Theory Comput.* **2012**, *8* (6), 1989–1998.
- (29) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
- (30) Jensen, F. Polarization consistent basis sets. III. The importance of diffuse functions. *J. Chem. Phys.* **2002**, *117* (20), 9234–9240.
- (31) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.
- (32) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98* (45), 11623–11627.
- (33) Popelier, P. L. A.; Jensen, F.; Mills, M. J. L.; Yuan, Y. Comprehensive analysis of energy minima of the twenty natural amino acids. *J. Phys. Chem. A* **2014**, *118* (36), 7876–7891.
- (34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian09, Revision A.02; Gaussian, Inc.: Wallingford, CT, 2002.
- (35) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5* (2), 129–145.
- (36) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11* (4), 431–439.
- (37) Chirlian, L. E.; Franc, M. M. Atomic charges derived from electrostatic potentials: A detailed study. *J. Comput. Chem.* **1987**, *8* (6), 894–905.
- (38) 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* (3), 361–373.
- (39) Woods, R. J.; Khalil, M.; Pell, W.; Moffat, S. H.; Smith, V. H. Derivation of net atomic charges from molecular electrostatic potentials. *J. Comput. Chem.* **1990**, *11* (3), 297–310.
- (40) White, J. C.; Davidson, E. R. Comparison of ab initio and multipole determinations of the electrostatic interaction of acetamide dimers. *J. Mol. Struct. THEOCHEM* **1993**, *282* (1–2), 19–31.
- (41) Hinsen, K.; Roux, B. A potential function for computer simulation studies of proton transfer in acetylacetone. *J. Comput. Chem.* **1997**, *18* (3), 368–380.
- (42) Tsiper, E. V.; Burke, K. Rules for minimal atomic multipole expansion of molecular fields. *J. Chem. Phys.* **2004**, *120*, 1153–1156.
- (43) Bulat, F. A.; Toro-Labbé, A.; Brinck, T.; Murray, J. S.; Politzer, P. Quantitative analysis of molecular surfaces: Areas, volumes, electrostatic potentials, and average local ionization energies. *J. Mol. Model.* **2010**, *16*, 1679–1691.
- (44) Storer, J.; Giesen, D.; Cramer, C.; Truhlar, D. Class IV charge models: A new semiempirical approach in quantum chemistry. *J. Comput.-Aided Mol. Des.* **1995**, *9* (1), 87–110.
- (45) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. New Class IV charge model for extracting accurate partial charges from wave functions. *J. Phys. Chem. A* **1998**, *102* (10), 1820–1831.
- (46) Thole, B.; van Duijnen, P. A general population analysis preserving the dipole moment. *Theor. Chim. Acta* **1983**, *63* (3), 209–221.
- (47) Zhang, P.; Bao, P.; Gao, J. Dipole preserving and polarization consistent charges. *J. Comput. Chem.* **2011**, *32* (10), 2127–2139.
- (48) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: 2002; p 105–119.
- (49) Stone, A. J.; Alderton, M. Distributed multipole analysis methods and applications. *Mol. Phys.* **1985**, *100* (1), 221–233.
- (50) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257–3273.
- (51) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Oxford Univ. Press: Oxford, U.K., 1990.
- (52) Popelier, P. L. A. *Atoms in Molecules. An Introduction*; Pearson Education: London, 2000.
- (53) Liem, S. Y.; Popelier, P. L. A.; Leslie, M. Simulation of liquid water using a high-rank quantum topological electrostatic potential. *Int. J. Quantum Chem.* **2004**, *99* (5), 685–694.
- (54) Yuan, Y.; Mills, M. J. L.; Popelier, P. L. A. Multipolar electrostatics for proteins: Atom–atom electrostatic energies in crambin. *J. Comput. Chem.* **2014**, *35* (5), 343–359.
- (55) Rasmussen, T. D.; Ren, P.; Ponder, J. W.; Jensen, F. Force field modeling of conformational energies: Importance of multipole moments and intramolecular polarization. *Int. J. Quantum Chem.* **2007**, *107* (6), 1390–1395.
- (56) Popelier, P. L. A.; Aicken, F. M. Atomic properties of amino acids: Computed atom types as a guide for future force-field design. *ChemPhysChem* **2003**, *4* (8), 824–829.