Article

PyRESP: A Program for Electrostatic Parameterizations of Additive and Induced Dipole Polarizable Force Fields

Shiji Zhao, Haixin Wei, Piotr Cieplak, Yong Duan,* and Ray Luo*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Molecular modeling at the atomic level has been applied in a wide range of biological systems. The widely adopted additive force fields typically use fixed atom-centered partial charges to model electrostatic interactions. However, the additive force fields cannot accurately model polarization effects, leading to unrealistic simulations in polarization-sensitive processes. Numerous efforts have been invested in developing induced dipole-based polarizable force fields. Whether additive atomic charge models or polarizable induced dipole models are used, proper parameterization of the electrostatic term plays a key role in the force field developments. In this work, we present a Python program called *PyRESP* for performing atomic multipole parameterizations by reproducing *ab initio* electrostatic potential (ESP) around molecules. *PyRESP* provides parameterization schemes for several electrostatic models, including the RESP model with atomic charges for the additive force fields and the RESP-ind and RESP-perm models with additional induced and permanent dipole moments for the polarizable force fields. *PyRESP* is a flexible and user-friendly program that can accommodate various needs during force field parameterizations for molecular modeling of any organic molecules.
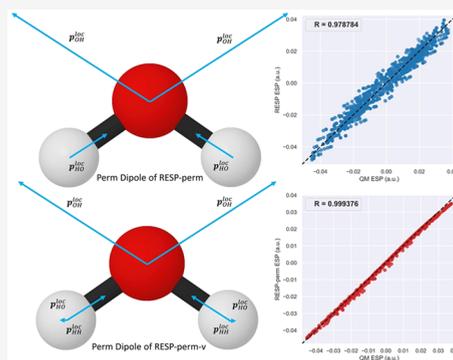
■ INTRODUCTION

Developing accurate force fields remains to be a great challenge for molecular modeling. One of the key components of force field development is the accurate modeling of atomic electrostatic interactions. The extensively used additive force fields apply fixed atom-centered partial charges to model electrostatic interactions, such as AMBER ff14SB,[1] ff19SB,[2] CHARMM,[3] and OPLS,[4] to name a few. One disadvantage of the additive force fields is that they are unable to model the atomic polarization effects, i.e., the redistribution of the atomic electron density due to the electric field produced by nearby atoms.[5] The importance of modeling polarization effects is well known. For example, during the protein folding process, amino acids forming a hydrophobic core must move from the hydrated environment to the more hydrophobic interior, experiencing considerably different dielectric environments.[6,7] Additive force fields are also considered to be unable to capture the important cation−π interactions between aromatic rings and charged amino acids, leading to unrealistic receptor−ligand interaction simulations.[8,9] Therefore, a great deal of effort has been directed to developing polarizable models, including the fluctuating charge models,[10,11] the Drude oscillator models,[12−16] and models incorporating induced dipoles[17,18] or continuum dielectric.[19,20]

The induced point dipole model is the most studied approach with a long history since the 1970s.[21,22] To date, it has been incorporated into several polarizable force fields, including AMOEBA,[23,24] AMBER ff02,[17] ff02pol.rl,[18] and ff12pol.[25−28] The original induced dipole model developed by Applequist et al. places the induced point dipole on each atom center, where the magnitude and direction of the induced dipole moment are determined by the isotropic polarizability of each atom and the electric field on this atom exerted by other atoms.[29] The induced dipole of atom i, subject to the external electric field E_i, is

$$\boldsymbol{\mu}_i = \alpha_i \left[\boldsymbol{E}_i - \sum_{j \neq i}^{n} \boldsymbol{T}_{ij} \boldsymbol{\mu}_j \right]$$

(1)

where α_i is the isotropic polarizability of atom i, and $\boldsymbol{T}_{ij}$ is the dipole field tensor with the matrix form

$$\boldsymbol{T}_{ij} = \frac{1}{r_{ij}{}^3} \boldsymbol{I} - \frac{3}{r_{ij}{}^5} \begin{bmatrix} x^2 & xy & xz \\ xy & y^2 & yz \\ xz & yz & z^2 \end{bmatrix}$$

(2)

where $\boldsymbol{I}$ is the identity matrix, and x, y, and z are the Cartesian components along the vector between atoms i and j at distance r_{ij}. However, this model suffers from the so-called "polarization

catastrophe" problem: the molecular polarizability diverges due to the cooperative interaction between induced dipoles at short distances.[5,29] One solution to this problem is to apply distance-dependent damping functions for interactions on short distances. Thole proposed several schemes by modeling the interaction using smeared charge distributions $\rho(u)$ instead of point charges, where $u = r_{ij}/(\alpha_i \alpha_j)^{1/6}$ is the effective distance. Here α_i and α_j are atomic polarizabilities of atoms i and j, and r_{ij} is the distance between them.[30,31] This modifies the dipole field tensor T_{ij} in such a way that it does not behave as r^{-3} at short distances. Among the proposed schemes, linear scheme (eq 3) and exponential scheme (eq 4) are shown to be the most effective

$$\rho(u) = \begin{cases} \dfrac{3}{\pi} \dfrac{(a-u)}{a^4} & u < a \\ 0 & u \geq a \end{cases} \tag{3}$$

and

$$\rho(u) = \frac{a^3}{8\pi} \exp(-au) \tag{4}$$

where a is the damping factor that controls the decay of the smeared charge distribution. Another Thole's scheme (eq 5) was adopted in the AMOEBA force field and implemented in the Tinker program,[23,24,32] which has the following form

$$\rho(u) = \frac{3a}{4\pi} \exp(-au^3) \tag{5}$$

The recently developed Thole scheme-based polarizable force field ff12pol has been shown to significantly reduce the root-mean-square errors of interaction energies with those calculated at the MP2/aug-cc-pVTZ level of theory, compared with additive force fields.[26]

About a decade ago, Elking et al. proposed a polarizable multipole model with Gaussian charge densities, which was later named the polarizable Gaussian multipole (pGM) model.[33] The nth-order Gaussian multipole at position r generated by an atom located at the coordinate R represented by the pGM model is defined as

$$\rho^{(n)}(r; R) = \Theta^{(n)} \cdot \nabla_R^{(n)} \left(\frac{\beta}{\sqrt{\pi}} \right)^3 e^{-\beta^2 |r-R|^2} \tag{6}$$

where $\Theta^{(n)}$ is the nth rank momentum tensor, $\nabla_R^{(n)}$ is the nth rank gradient operator, and β is a Gaussian exponent controlling the "radius" of the distribution with the following form

$$\beta = s \left(\frac{2\alpha}{3\sqrt{2\pi}} \right)^{-1/3} \tag{7}$$

where α is the atomic polarizability, and s is the screening factor. Although in the pGM model any order of multipoles can be modeled, only charges (zeroth-order multipole; eq 8) and dipoles (first-order multipole; eq 9) are retained in the current pGM model design

$$\rho^{(0)}(r; R) = q \left(\frac{\beta}{\sqrt{\pi}} \right)^3 e^{-\beta^2 |r-R|^2} \tag{8}$$

$$\rho^{(1)}(r; R) = p \cdot \nabla_R \left(\frac{\beta}{\sqrt{\pi}} \right)^3 e^{-\beta^2 |r-R|^2} \tag{9}$$

where q is the permanent charge and p is the permanent dipole. Wei et al. recently proposed a local frame for the permanent dipoles formed by covalent basis vectors (CBVs), which are unit vectors along the direction of covalent bonds or virtual bonds.[34,35] This design is based on the fact that atomic permanent moments mainly result from covalent bonding interactions. Replacing p with μ in eq 9 will give the pGM distribution of induced dipole, which has the same form as that of permanent dipole. A key advantage of the pGM model is that all short-range electrostatic interactions can be calculated analytically in a consistent manner, including the interactions of charge−charge, charge−dipole, charge−quadrupole, dipole−dipole, and so on. Consequently, it has been shown that the pGM model notably improves the prediction of molecular polarizability anisotropy compared with that of Thole models.[36]

Each of the four damping schemes discussed above requires parameterization of the atomic isotropic polarizabilities α and damping factors a (and s for the pGM model), which has been done by fitting experimental or *ab initio* molecular polarizability tensors using a genetic algorithm, as presented in our recent works.[25,36] In this work, we aim to take one step further toward the development of general and accurate polarizable force fields by developing a computer program for electrostatic parameterizations for the atomic charges and dipoles of various polarizable models.

For additive models, the atomic point partial charges are traditionally derived by performing least-squares fitting of the charges to reproduce the quantum mechanically (QM) determined electrostatic potential (ESP) at a large number of grid points lying outside the van der Waals distance of the molecule. Assuming a molecule with n atoms is being parameterized, and there are m ESP points lying outside the van der Waals distance of the molecule, then the least-squares fitting aims to minimize the objective function

$$\gamma = \sum_{j=1}^{m} (V_j^{QM} - V_j)^2 \tag{10}$$

where V_j^{QM} is the ESP value evaluated through QM calculations at point j, and V_j is the ESP value calculated from the fitting results. This method was initially used by Momany,[37] further refined by Cox et al.[38] An ESP point sampling scheme that uses points on molecular surfaces constructed using gradually increasing van der Waals radii for the atoms was proposed by Singh et al.[39,40] The CHELP algorithm initially employed a Lagrange multiplier method to perform constrained least-squares fitting, in which the Lagrange multiplier (λ) is multiplied by the constraining function (g) and added to the objective function γ to be minimized. In the context of charge fitting, the Lagrange multiplier method is mostly used to enforce the total charge constraints, i.e., the charge of all atoms of a molecule should sum to the total molecular charge. Alternatively, it can also be used to specify the total charge of molecular fragments. For example, during amino acid parameterizations, the *N*-acetyl (ACE) and *N*-methylamide (NME) groups are commonly used to cap amino acid dipeptides to mimic the chemical environment within a protein. Both capping fragments need to

be constrained to have a neutral charge to ensure the correct total charge of the amino acid fragments.[41,42]

In general, the ESP-based charge derivation methods perform very well in reproducing QM determined molecular multipole moments and intermolecular interaction energy. However, all methods discussed above suffer from the problem that the atomic charges are sensitive to molecular conformations, leading to a lack of transferability of the charges between identical molecules with different conformations, as well as between common functional groups in related molecules. Another problem of this approach is the poor determination of charges on buried atoms that are far from ESP points, which can fluctuate wildly to reach the optimal fitting to the ESP. Both problems have been addressed by the restrained electrostatic potential (RESP) method developed by Bayly et al., which employs restraints by adding a penalty function χ to the objective function during the fitting process.[43,44] Two types of penalty functions were proposed. The first is a simple harmonic function

$$\chi = a \sum_{i=1}^{n} q_i^2 \tag{11}$$

where a is the scale factor determining the restraining strength. The second penalty function is a hyperbolic function with the form

$$\chi = a \sum_{i=1}^{n} \left(\sqrt{q_i^2 + b^2} - b \right) \tag{12}$$

where a is again the scale factor that defines the restraining strength, and b determines the "tightness" of the hyperbola around its minimum. b has been recommended to be set to 0.1 by the original RESP work to make the restraint appropriately tight.[43] To this end, assuming there are w different Lagrange constraints imposed on the charges in a molecule, the objective function to be minimized becomes

$$z = \gamma + \lambda_1 g_1 + \lambda_2 g_2 + ... + \lambda_w g_w + \chi \tag{13}$$

To date, the computer program *RESP* has been applied in charge derivations of a variety of additive force fields[41,42] and is still being used actively for charge calculations for small organic molecules.[8,45,46] Following the idea of charge parameterization by reproducing ESPs, Cieplak et al. extended the RESP method for induced dipole electrostatic models, assuming that ESPs around molecules are determined by both permanent charges and atomic-induced dipoles. According to this method, atomic charges are iteratively fitted to the effective ESP, which is the difference between the QM-derived ESPs and the ESPs generated by induced dipoles. Iterations stop when the induced molecular dipole moment converges within a certain accuracy level.[5,17] A program named *i_RESP* has been developed to facilitate this iterative charge fitting procedure.

In this work, we further extended the RESP method for parameterizations of electrostatic models with induced point dipoles and permanent point dipoles. A Python program named *PyRESP* was designed and implemented based on its ancestor *RESP* program, providing the parameterization ability for three electrostatic models: (1) the additive RESP model; (2) the polarizable model with induced point dipoles only, named the RESP-ind model; and (3) the polarizable model with both induced point dipoles and permanent point dipoles, named the RESP-perm model. In the next section, we present the theory behind the parameterization strategies of the three models, as well as several other features provided by *PyRESP*. We have tested all three models using several representative molecules, and the parameterization results will be evaluated and discussed.

■ THEORY

In earlier works, the objective function z shown in eq 13 has been minimized using iterative gradient descent approaches, as were done by Momany et al. and Singh et al.[37,39] Similarly, the *i_RESP* program developed by Cieplak et al. parameterizes the induced dipole polarizable model iteratively by fitting charges to the differences between the QM-derived ESPs and the ESPs generated by induced dipoles.[5,17] In both cases, an initial guess on the atomic charges before the iteration process is required. On the other hand, iterative algorithms suffer from the problem that the convergence of iteration is sensitive to the specified accuracy level. In rare cases, the objective function might jump back and forth near the minimum, leading to a nonconvergence problem. Therefore, *PyRESP* takes a direct approach by solving the system of equation in the matrix form with the partial derivative of the objective function z against each parameter (permanent charges or dipoles) and each Lagrange multiplier λ set to be equal to zero, as were done in CHELP, CHELPG, and the original RESP works.[43,47−49] The advantage of the direct approach is that it gives the exact least-squares solution, so that the initial guess on the atomic charges and accuracy level is no longer needed. Another advantage of the direct approach is that the matrix form representations allow us to present each of the following electrostatic models in a consistent and elegant way.

RESP. The original RESP method performs charge fitting for additive electrostatic terms with the assumption that ESPs only come from permanent point charges.[43] For each ESP point j, the following equation needs to be solved

$$\sum_{i=1}^{n} \frac{q_i}{r_{ij}} = V_j^{\text{QM}} \tag{14}$$

In the matrix form

$$Xq = V \tag{15}$$

where X is m by n matrix for charge−ESP interactions between each ESP point j and atom i, q is the n-dimensional vector for the partial charge of each atom, and V is the m-dimensional vector for QM ESP. Typically, there are many ESP points sampled so that X becomes a rectangular matrix (tall and thin). Consequently, eq 15 is unlikely to have an exact solution. Therefore, we aim to find the least-squares solution by solving the following equation, the proof of which can be found in most linear algebra textbooks

$$X^T X q = X^T V \tag{16}$$

where $X^T X$ is a square matrix and is usually positive definite and invertible. The constraints on the charges could also be expressed in the following matrix form

$$Kq = L \tag{17}$$

where K is a w by n matrix with only 1 and 0 as elements indicating the presence or absence of each charge in each constraint, and L is the w-dimensional vector for the total charge in each constraint. The constrained least-squares fitting

has the following matrix form, whose solution gives constrained RESP fitting results

$$\begin{bmatrix} X^T X & K^T \\ K & 0 \end{bmatrix} \begin{bmatrix} q \\ \lambda \end{bmatrix} = \begin{bmatrix} X^T V \\ L \end{bmatrix} \tag{18}$$

where λ is the w-dimensional vector of all Lagrange multipliers. Finally, the penalty function χ could be applied to restrain fitted charges by adding its partial derivative only to the diagonal terms of the matrix in eq 18, and the reasoning can be found in the original RESP work.[43]

RESP-ind (RESP with Induced Point Dipole). Following Applequist et al.,[29] eq 1 may be rearranged into

$$\alpha_i^{-1} \mu_i + \sum_{j \neq i}^{n} T_{ij} \mu_j = E_i \tag{19}$$

which could be written in the following matrix form

$$A\mu = E \tag{20}$$

where A is a $3n$ by $3n$ matrix containing the information of polarizability and dipole field tensors, μ is a $3n$-dimensional vector of the induced dipole of each atom, and E is a $3n$-dimensional vector of the electric field at atom i.

The implicit assumption is that E_i is produced by permanent charges of all atoms other than i, and there are no additional applied external electric fields. Thus, we have

$$E_i = \sum_{j \neq i}^{n} \frac{q_j}{r_{ij}^3} r_{ji} \tag{21}$$

In the matrix form

$$E = Cq \tag{22}$$

where C is a $3n$ by n matrix of the charge-electric field coefficient between each atom pair. Combining eqs 20 and 22 gives

$$\mu = A^{-1} Cq \tag{23}$$

In contrast to the RESP model where the permanent charges are the only sources for ESPs, the RESP-ind model assumes that ESP comes from both permanent point charges and induced point dipoles. Therefore, for each ESP point j, we have the following equation

$$\sum_{i=1}^{n} \frac{q_i}{r_{ij}} + \sum_{i=1}^{n} \frac{\mu_i \cdot r_{ij}}{r_{ij}^3} = V_j^{QM} \tag{24}$$

In the matrix form

$$Xq + Y\mu = V \tag{25}$$

where Y is an m by $3n$ matrix for the dipole−ESP interactions between each ESP point and atom pair. Substitute eq 23 into eq 25 gives

$$(X + YA^{-1}C)q = V \tag{26}$$

Same as we did for the RESP model, solving the following equation gives the least-squares solution

$$(X + YA^{-1}C)^T (X + YA^{-1}C)q = (X + YA^{-1}C)^T V \tag{27}$$

and solving the following equation gives the constrained least-squares solution

$$\begin{bmatrix} (X + YA^{-1}C)^T (X + YA^{-1}C) & K^T \\ K & 0 \end{bmatrix} \begin{bmatrix} q \\ \lambda \end{bmatrix} = \begin{bmatrix} (X + YA^{-1}C)^T V \\ L \end{bmatrix} \tag{28}$$

Finally, the partial derivative of the penalty function χ can be applied to eq 28 to restrain atomic charges.

RESP-perm (RESP with Induced and Permanent Point Dipoles). RESP-perm is the electrostatic model with the highest degree of freedom implemented in *PyRESP*. It has one additional component compared to the RESP-ind model, the permanent point dipoles p_i of each atom i, which is a three-dimensional vector. Now, the electric field at atom i is produced by both permanent charges and permanent dipoles of all atoms other than i. Thus, we have

$$E_i = \sum_{j \neq i}^{n} \left(\frac{q_j}{r_{ij}^3} r_{ji} + T_{ij} p_j \right) \tag{29}$$

In the matrix form

$$E = Cq + Dp \tag{30}$$

where D is a $3n$ by $3n$ matrix of the dipole-electric field coefficients between each atom pair, and p is a $3n$-dimensional vector for the permanent dipole of each atom in the global frame. Therefore, the induced dipole vector μ becomes

$$\mu = A^{-1}(Cq + Dp) \tag{31}$$

Now, ESPs come from three sources: permanent point charges, permanent point dipoles, and induced point dipoles. That is

$$\sum_{i=1}^{n} \frac{q_i}{r_{ij}} + \sum_{i=1}^{n} \frac{(\mu_i + p_i) \cdot r_{ij}}{r_{ij}^3} = V_j^{QM} \tag{32}$$

In the matrix form

$$Xq + Y(\mu + p) = V \tag{33}$$

eq 31 can be plugged into eq 33 and rearranged to

$$(X + YA^{-1}C)q + Y(A^{-1}D + I)p = V \tag{34}$$

The RESP-perm model is designed to be compatible with the pGM model of Wei et al.,[34] where the permanent dipoles are defined in the local frame formed by CBVs. Assume that the molecule to be fitted has z CBVs, i.e., $z/2$ covalent bonds since covalent bonds are bidirectional, then the permanent dipoles in global frame p can be conveniently expressed in the local frame using a $3n$ by the z-dimensional conversion matrix F, with CBVs as its elements. The conversion has the simple matrix form

$$p = Fp^{loc} \tag{35}$$

where p^{loc} is a z-dimensional vector for permanent dipoles in the local frame. Therefore, the RESP-perm model in fact performs least-squares fitting on p^{loc} rather than on p, and eq 34 should be expressed as

$$(X + YA^{-1}C)q + Y(A^{-1}D + I)Fp^{loc} = V \tag{36}$$

One advantage of using matrix F is that the local frame can be easily extended to include noncovalent basis vectors. In the current *PyRESP* implementation, the "virtual" bonds of 1-3 interacting atom pairs are also enabled; all we need to do is to

increase the number of columns of F to contain both covalent basis vectors and 1-3 interaction basis vectors, and the number of rows of F will not change since the number of atoms stays the same. The RESP-perm model considering both 1-2 and 1-3 interacting atom pairs in the local frame is named RESP-perm-v, where v stands for "virtual".

To perform least-squares fitting on both q and p^{loc} directly, we construct a new vector Q, which is $(n + z)$-dimensional vector $\begin{bmatrix} q \\ p^{loc} \end{bmatrix}$, and a new matrix M, which is m by $(n + z)$ matrix $[(X + YA^{-1}C) \; Y(A^{-1}D + I)F]$. Then, we have

$$MQ = V \tag{37}$$

The least-squares solution of Q can be found by solving

$$M^T M Q = M^T V \tag{38}$$

and the constrained least-squares fitting has the matrix form

$$\begin{bmatrix} M^T M & K^T \\ K & 0 \end{bmatrix} \begin{bmatrix} Q \\ \lambda \end{bmatrix} = \begin{bmatrix} M^T V \\ L \end{bmatrix} \tag{39}$$

The current *PyRESP* implementation uses two separate restraining strengths for permanent charges and permanent dipoles, which can be set to different values according to users' preferences.

Intra- and Intermolecular Equivalences. A reliable force field would require atoms sharing equivalent chemical environments to have identical permanent charges and dipoles. Taking a methyl group as an example, all three hydrogens must have the same charge, and all permanent dipoles pointing from methyl carbon toward hydrogens (and those in reverse directions) must have the same magnitudes; otherwise, rotating the methyl to the three degenerate rotamers would give rise to different energies. Intramolecular equivalencing is applied for this symmetry purpose. One strategy examined by previous studies is averaging the charges of the equivalent atoms after the fitting, which were set free to change during the fitting process. However, this so-called *a posteriori* strategy was found to have an unsatisfying negative impact on the fitting quality and on the final molecular dipole moments.[43] Thus, the *PyRESP* program employs the improved approach proposed by the original RESP work that performs equivalencing during the fitting process. Depending on the specific electrostatic model selected, the preliminary matrices in eqs 18, 28, or 39 are generated as if there were no equivalent fitting centers. Then, the rows and columns of corresponding equivalent fitting centers were added up to form a single row and column, giving rise to smaller linear equation systems to be solved as usual.

In comparison, intermolecular equivalencing is often used for fitting one set of parameters for multiple conformations of the same molecule to further reduce the conformation-dependent problem, in addition to applying restraints. Alternatively, it can also be used for fitting the same chemical groups in different molecules. Both intra- and intermolecular charge equivalencing have already been implemented in the original RESP program.[43] In *PyRESP*, the equivalencing algorithm is extended so that both intra- and intermolecular equivalencing are enabled for permanent charges and dipoles in a consistent manner.

Polarization Catastrophe Avoidance. A well-known problem of the point dipole model discussed so far is that it may lead to infinite molecular polarizability by the cooperative interaction between two induced dipoles, known as "polarization catastrophe".[5,29] One way to avoid this problem is to turn off the polarization interactions between 1-2 and 1-3 interacting atom pairs, as were done in the AMBER ff02 and ff02pol.r1 force fields.[17,18] This can be easily achieved by setting corresponding elements in the charge-electric field coefficient matrix C and the dipole-electric field coefficient matrix D to zero. Alternatively, one can apply distance-dependent damping functions on interacting atom pairs, such as those developed by Thole[30,31] and the pGM scheme developed by Elking et al.,[33] which will lead to the damped dipole field tensor

$$T_{ij} = \frac{f_e}{r_{ij}^3} I - \frac{3f_t}{r_{ij}^5} \begin{bmatrix} x^2 & xy & xz \\ xy & y^2 & yz \\ xz & yz & z^2 \end{bmatrix} \tag{40}$$

with screening functions f_e and f_t. Consequently, the charge-electric field coefficient matrix C and the dipole-electric field coefficient matrix D will also contain elements damped by f_e and f_t correspondingly. It is easy to see that for the original undamped Applequist model, f_e and f_t are constants

$$f_e = 1.0; \; f_t = 1.0 \tag{41}$$

For the linear model, we have

$$v = u/a$$

$$f_e = \begin{cases} 4v^3 - 3v^4 & v < 1 \\ 1.0 & v \geq 1 \end{cases}$$

$$f_t = \begin{cases} v^4 & v < 1 \\ 1.0 & v \geq 1 \end{cases} \tag{42}$$

For the exponential model, we have

$$v = au$$

$$f_e = 1 - \left(\frac{v^2}{2} + v + 1 \right) \exp(-v)$$

$$f_t = 1 - \left(\frac{v^3}{6} + \frac{v^2}{2} + v + 1 \right) \exp(-v) \tag{43}$$

For the Tinker-exponential model, we have

$$v = au^3$$

$$f_e = 1 - \exp(-v)$$

$$f_t = 1 - (v + 1)\exp(-v) \tag{44}$$

For the pGM model, we have

$$S_{ij} = \frac{\beta_i \beta_j r_{ij}}{\sqrt{2(\beta_i^2 + \beta_j^2)}}$$

$$f_e = \text{erf}(S_{ij}) - \frac{2}{\sqrt{\pi}} S_{ij} \exp(-S_{ij}^2)$$

$$f_t = \text{erf}(S_{ij}) - \frac{2}{\sqrt{\pi}} S_{ij} \exp(-S_{ij}^2)\left(1 + \frac{2}{3} S_{ij}^2 \right)$$

$$f_0 = \text{erf}(S_{ij}) \tag{45}$$

Note that for the pGM model, the charge–ESP interaction matrix X and the dipole–ESP interaction matrix Y should be scaled by f_0 and f_e, respectively, in addition to modifying the dipole field tensor T_{ij}.

In the current *PyRESP* release, both polarization catastrophe avoidance strategies have been implemented, including turning off 1-2 and 1-3 interactions and the four damping schemes (linear, exponential, Tinker-exponential, and pGM schemes).

COMPUTATIONAL DETAILS

***Ab Initio* Calculations.** Several molecules were selected as candidates for testing the *PyRESP* program, including water, methanol (alcohol), ethane (aliphatic), benzene (aromatic), N-methyl acetamide (peptide backbone), dimethyl phosphate (nucleic acid backbone), adenine (nucleobase), alanine dipeptide (hydrophobic amino acid), serine dipeptide (polar amino acid), arginine dipeptide (positively charged amino acid), and aspartic acid dipeptide (negatively charged amino acid). For the seven non-amino acid molecules, single-conformation fittings were performed. For the four amino acid molecules, both single-conformation and double-conformation fittings were performed, with the main-chain torsion angles in ($\phi = 300°$, $\psi = 300°$) and ($\phi = 240°$, $\psi = 120°$), approximating α-helix and antiparallel β-sheet secondary structure conformations. The geometries of all molecules were optimized at the B3LYP/6-311++G(d,p) level of theory, with dihedral angle constraints applied to the corresponding amino acid molecules only.

QM ESP values were calculated at the MP2/aug-cc-pVTZ level of theory for a set of points fixed in space in the solvent-accessible region around each molecule. The points were generated using the method developed by Singh et al. on molecular surfaces (with a density of 6 points/Å²) at each of 1.4, 1.6, 1.8, and 2.0 times the van der Waals radii.[39,40] For small molecules such as water, approximately 1800 points were generated, while for large molecules such as arginine dipeptide, more than 9000 points were generated. All *ab initio* calculations were performed using Gaussian 09 software.[50]

Parameterizations. A two-stage parameterization procedure has been adopted as the standard approach for RESP parameterization.[43] We extended this procedure for all electrostatic models: RESP, RESP-ind, and RESP-perm (and RESP-perm-v for water molecule), where the hyperbolic function in eq 12 was applied in all parameterizations. In the first stage, all fitting centers (permanent charges for all models, and permanent dipoles for RESP-perm and RESP-perm-v) were set free to change, and a weak restraining strength of 0.0005 (a in eq 12) was applied to all fitting centers. In the second stage, intramolecular equivalencing was enforced on all fitting centers that share an identical chemical environment with others, such as methyl and methylene hydrogens. A stronger restraining strength of 0.001 was applied to those fitting centers, and all other fitting centers were set frozen to keep the values obtained from the first stage. The restraints were only applied to non-hydrogen heavy atoms. To get better fitting results, the only Lagrange constraint enforced during parameterization is the total charge constraint, without applying additional intramolecular charge constraints. Intermolecular equivalencing was enforced in both the first and the second stages for double-conformation fittings of amino acid molecules.

Previous studies have shown that in the polarizable models with Thole-like damping schemes, it is important to include all atomic pair interactions to have an anisotropic molecular response.[36,51] Therefore, for parameterizations of the RESP-ind, RESP-perm, and RESP-perm-v models, both 1-2 and 1-3 polarization interactions were included, and the pGM damping scheme was applied to all models to avoid polarization catastrophe.[33,34] The isotropic atomic polarizabilities derived in the previous work were employed for models considering polarization effects.[36]

The performance of each electrostatic model was evaluated based on the relative root-mean-square (RRMS) error,[38,43,49] given by

$$\text{RRMS} = \sqrt{\frac{\sum_{j=1}^{m} \left(V_j^{\text{QM}} - V_j\right)^2}{\sum_{j=1}^{m} V_j^{\text{QM2}}}} \tag{46}$$

The molecular dipole moments and quadrupole moments along the principal axes calculated with each electrostatic model were compared with those calculated using *ab initio* methods as an additional metric in evaluating parameterization results. The Pearson correlation analysis was performed using the Python package *Scipy*. The scatterplots for QM ESPs and ESPs calculated by electrostatic models are plotted using the Python package *Matplotlib*.

RESULTS

Water. The first molecule we tested is the water molecule. Table 1 shows the parameterization results, RRMS, and

Table 1. Parameterization Results, RRMS, and Molecular Dipole/Quadrupole Moments of Water Fitted with Four Electrostatic Models[a]

	RESP	RESP-ind	RESP-perm	RESP-perm-v	QM
Charges/a.u.					
H	0.3401	0.5182	0.7576	0.7441	
O	−0.6802	−1.0365	−1.5151	−1.4882	
Permanent Dipole Moments/a.u.					
H–O[a]			0.0753	0.0773	
O–H[a]			−0.2761[b]	−0.2577	
H–H[a]				−0.0121	
RRMS					
	0.2051	0.1244	0.0391	0.0404	
Dipole Moments/Debye					
μ[c]	1.9141	1.9417	1.8668	1.8660	1.8470
Quadrupole Moments/Debye Angstroms					
Q_{xx}[d]	1.0444	1.5151	1.8549	1.8803	1.8389
Q_{yy}[d]	−0.1858	−0.3198	−0.2467	−0.2841	−0.2418
Q_{zz}[d]	−0.8586	−1.1953	−1.6082	−1.5962	−1.5971

[a]Each permanent dipole moment $\boldsymbol{p}_{\text{AB}}^{\text{loc}}$ is named in the format A–B, corresponding to the CBV points from atom A to atom B. [b]Negative value indicates pointing in the reverse direction of CBV. [c]Dipole moment relative to the center of mass. [d]Quadrupole moments along the principal axes.

moments of the water molecule fitted with the RESP, RESP-ind, RESP-perm, and RESP-perm-v electrostatic models. All models fit permanent point charges on oxygen and hydrogen atoms. In addition, the RESP-perm and RESP-perm-v models also fit local frame permanent point dipole moments defined on CBVs, i.e., unit vectors along the direction of 1-2 interacting atom pairs (covalent bonds) or 1-3 interacting atom pairs (virtual bonds). For the RESP-perm model, a water molecule
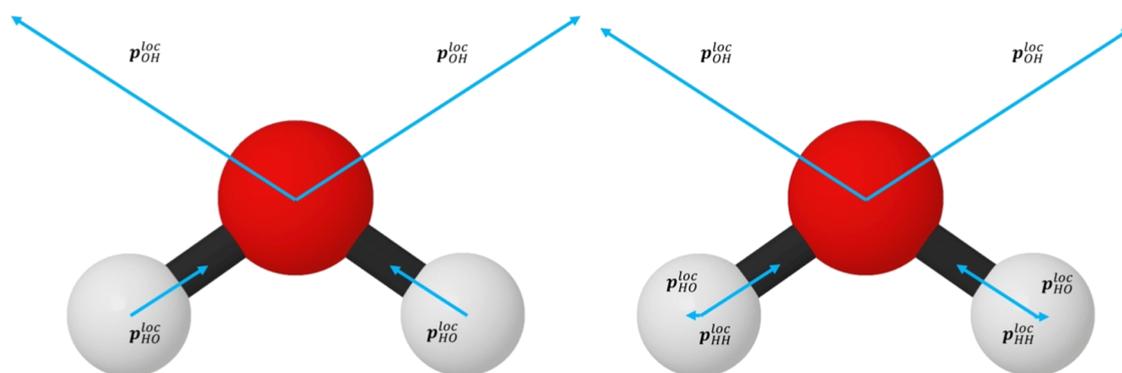
Figure 1. Schematic representation of local frame permanent dipole moments of water molecule fitted with RESP-perm (left) and RESP-perm-v (right) electrostatic models. The lengths of permanent dipole moments are shown in the scale of their magnitudes. Refer to the text for detailed descriptions.
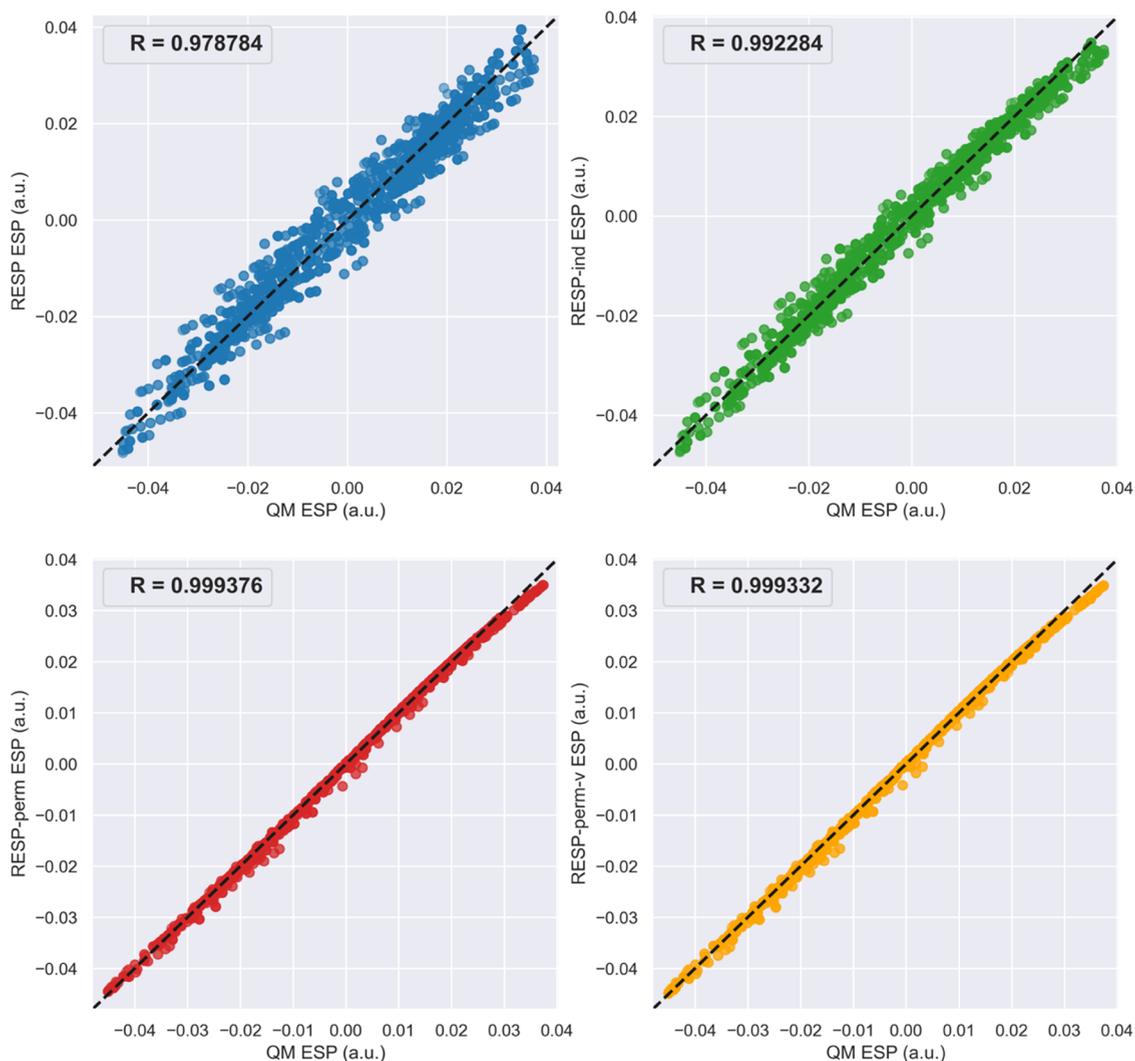
Figure 2. Correlation analysis of QM ESPs and ESPs calculated with various electrostatic models for a water molecule, which was fitted with 1874 ESP data points. The dashed line corresponds to a perfect correlation. R is the Pearson correlation coefficient.

has two types of permanent dipoles: p_{OH}^{loc} and p_{HO}^{loc}, while the RESP-perm-v model has one additional type of permanent dipole, p_{HH}^{loc}, corresponding to the virtual CBV between the two hydrogen atoms. The permanent dipoles p_{OH}^{loc} and p_{HH}^{loc} have negative values, which means they point in the opposite direction of corresponding CBVs. That is, p_{OH}^{loc} points from the

oxygen atom against the direction of the hydrogen atom, rather than the default CBV direction, which points from oxygen toward hydrogen. Similarly, p_{HH}^{loc} points from the hydrogen atom against the direction of the neighbor hydrogen atom, rather than the default CBV direction toward the neighbor hydrogen. Figure 1 gives a better illustration of the parameter-

ization results of local frame permanent dipole moments of a water molecule. It can be observed that the RESP-perm and RESP-perm-v models produce higher magnitudes of permanent charges than the RESP and RESP-ind models. That is, they assign values to the charge centers in a more aggressive way to reproduce QM ESPs. All models assign negative charges to the oxygen atom and positive charges to the hydrogen atom, and both the RESP-perm and RESP-perm-v models assign a large but negative value to permanent dipole moments p_{OH}^{loc}. This agrees with the fact that oxygen has a higher electronegativity than hydrogen.

The RESP-perm model produces the lowest RRMS, with its RRMS only 19% of that of the RESP model, a factor of more than 5-fold reduction. The RESP-perm and RESP-perm-v models also produce molecular dipole moments and quadrupole moments with better agreement with the QM moments. The scatterplots of QM ESPs versus calculated ESPs for water are shown in Figure 2. The Pearson correlation coefficients of the RESP-perm and RESP-perm-v models are the highest among all models, and the RESP-ind model comes next. We can therefore conclude that electrostatic models with induced dipoles and permanent dipoles perform better than the RESP model in terms of all metrics analyzed.

The current RESP-perm-v model enables the virtual bonds between 1−3 interacting atom pairs. In theory, we can also enable virtual bonds between 1-4, 1-5, and atom pairs with even longer distances using a consistent method, giving rise to higher-level RESP-perm-v models. However, as can be seen from Table 1 and Figure 2, the virtual bonds in the RESP-perm-v model do not improve the fitting quality of the water molecule. In fact, adding too many virtual bonds may lead to the overfitting problem and is expected to significantly increase the computational time for both parameterization and MD simulation processes. For these reasons, parameterization with the RESP-perm-v model will only be performed for the water molecule for illustration purposes, and other molecules will only be parameterized with the RESP, RESP-ind, and RESP-perm models.

Methanol, Ethane, and Benzene. We next extend our studies to the molecules methanol (CH_3OH), ethane (CH_3CH_3), and benzene (C_6H_6) to see how the parameterization results for these molecules differ from those for water. Methanol has lower symmetry than water, so it is of interest to see how electrostatic models parameterize this molecule. As shown in Table 2, all models assign large negative charges to the highly electronegative oxygen atom and produce low RRMS and high correlation coefficients (Figure 3). In terms of molecular dipole and quadrupole moments, the RESP-perm model yields the best agreement with QM calculations among all three models. The results of methanol show the importance of induced and permanent dipoles for modeling polar molecules.

In the case of ethane, all models assign positive charges to hydrogen and negative charges to carbon, as shown in Table 3. Among the three models, the RESP-ind model assigns charges with the highest magnitudes, and the RESP model assigns charges with the lowest magnitudes. Ethane is a nonpolar molecule, as reflected by the molecular dipole moments calculated by all three models as well as QM calculations. However, the RESP-perm model significantly outperforms the RESP and the RESP-ind models in terms of all other metrics, including RRMS, quadrupole moments, and correlation coefficients, making it the only model that gives reasonable

Table 2. Parameterization Results, RRMS, and Molecular Dipole/Quadrupole Moments of Methanol Fitted with Three Electrostatic Models[a]

	RESP	RESP-ind	RESP-perm	QM
Charges/a.u.				
C	0.1609	0.1008	−0.0763	
H (methyl)	0.0194	0.0770	0.1105	
O	−0.6002	−0.8841	−1.0075	
H (hydroxyl)	0.3812	0.5524	0.7524	
Permanent Dipole Moments/a.u.				
C−H (methyl)			−0.0141	
H (methyl)−C			−0.0068	
C−O			0.0158	
O−C			0.1071	
O−H (hydroxyl)			−0.2268	
H (hydroxyl)−O			0.0973	
RRMS				
	0.2519	0.1298	0.0801	
Dipole Moments/Debye				
μ	1.9558	1.7563	1.6786	1.6873
Quadrupole Moments/Debye Angstroms				
Q_{xx}	2.2197	2.5574	2.6684	2.6984
Q_{yy}	−0.7640	−0.7275	−0.6935	−0.8281
Q_{zz}	−1.4557	−1.8299	−1.9749	−1.8703

[a]See Table 1 for notation.

performance. As shown in Figure 3, the ESPs around the ethane molecule are very close to 0 a.u., with the range between −0.005 and 0.006 a.u., compared with that of polar molecules such as water (−0.045−0.04 a.u.) and methanol (−0.05−0.04 a.u.). The nonpolar nature of ethane makes it particularly difficult to parameterize, so that models with a high degree of freedom like RESP-perm perform significantly better than those with a low degree of freedom.

Table 4 shows the parameterization results, RRMS, and moments of benzene. Similar to the ethane molecule, benzene is also a nonpolar molecule, and the molecular dipole moment was successfully predicted by all three models. The RESP-ind model again fits charges most aggressively by assigning charges with the highest magnitudes, and the RESP model fits charges most conservatively by assigning charges with the lowest magnitudes. However, unlike the case of ethane, none of the models perform significantly better in terms of other metrics. The RESP model yields the lowest RRMS, but it is only 14% lower than the highest RRMS (given by the RESP-ind model). All models underestimate the molecular quadrupole moments, although those given by the RESP-ind model have better agreement with QM results than those of the other two models. As shown in Figure 3, the RESP-perm model has the highest correlation coefficient but is still lower than those for polar molecules such as water and methanol. Modeling aromatics such as benzene is therefore also a difficult task, possibly due to the existence of π orbitals that are located outside of the two-dimensional plane of the aromatics ring.

NMA, DMP, and Adenine. We next turn to N-methyl acetamide (NMA), dimethyl phosphate (DMP), and adenine base. These molecules are chosen as they are common model compounds for peptides and nucleic acids. Tables 5 and 6 show the charges, RRMS, and moments of NMA and DMP, respectively, and the permanent dipole moments fitted with the RESP-perm model are shown in Tables S1 and S2. All models produce charge sets with consistent signs for NMA.
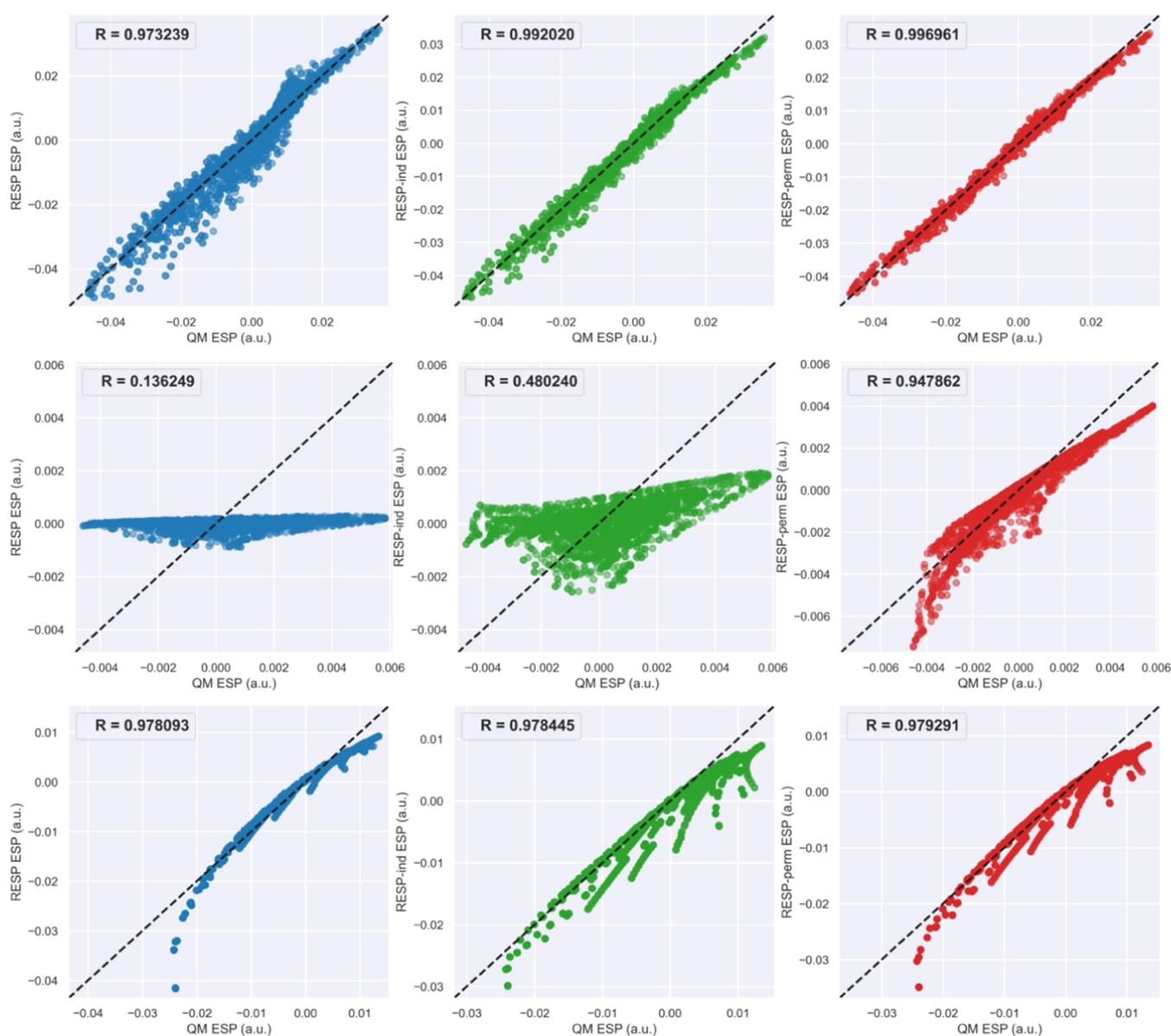
Figure 3. Correlation analysis of QM ESPs and ESPs calculated with various electrostatic models for methanol (upper panel), ethane (middle panel), and benzene (lower panel) molecules. Methanol, ethane, and benzene molecules were fitted with 2654, 2951, and 4130 ESP data points, respectively. The dashed lines correspond to a perfect correlation. R is the Pearson correlation coefficient.

Interestingly, there is significant variation in the atomic charges of DMP fitted by the three models. For example, the charges for the central phosphorus (P) range from −0.4188 to 1.1047 a.u. Low RRMS and high correlation coefficients (Figure 4) are yielded by all models. However, for both NMA and DMP molecules, the molecular dipole and quadrupole moments produced by the RESP-ind and RESP-perm models agree worse to the QM results than those of the RESP model, indicating the potential overfitting problem for the RESP-ind and RESP-perm models.

The charges, RRMS, and moments of the nucleic acid base adenine are shown in Table 7, and the permanent dipole moments fitted with the RESP-perm model are shown in Table S3. Among the three electrostatic models, RESP-ind assigns charges with the highest magnitude to most atoms but results in the worst RRMS, molecular dipole moment agreement, and correlation coefficient. On the other hand, the RESP-perm model yields the lowest RRMS, dipole, and quadrupole

moments with the best agreements and highest correlation coefficient (Figure 4). Therefore, permanent dipole moments are necessary components for modeling the adenine molecule.

Amino Acid Dipeptides. *PyRESP* was designed as the next-generation parameterization tool for polarizable force field development, with the aim to replace its ancestor *RESP* program.[43,44] Amino acids are key molecules for force field development for biomacromolecules, so we next tested the program on several amino acid dipeptides, all capped with *N*-acetyl (ACE) group at the N-terminal, and *N*-methylamide (NME) group at the C-terminal. Selected amino acids include alanine (hydrophobic amino acid), serine (polar amino acid), arginine (positively charged amino acid), and aspartic acid (negatively charged amino acid). Two conformations, approximating α-helix (ϕ = 300°, ψ = 300°) and antiparallel β-sheets (ϕ = 240°, ψ = 120°), were used for both single-conformation and double-conformation fittings. Double-conformation fittings were performed with intermolecular

Table 3. Parameterization Results, RRMS, and Molecular Dipole/Quadrupole Moments of Ethane Fitted with Three Electrostatic Models[a]

	RESP	RESP-ind	RESP-perm	QM
		Charges/a.u.		
C	−0.0254	−0.2148	−0.0723	
H	0.0085	0.0716	0.0241	
		Permanent Dipole Moments/a.u.		
C−H			−0.0201	
C−C			0.0645	
H−C			−0.0787	
		RRMS		
	0.9939	0.8808	0.3490	
		Dipole Moments/Debye		
μ	0.0000	0.0000	0.0000	0.0000
		Quadrupole Moments/Debye Angstroms		
Q_{xx}	0.0403	0.0457	−0.5761	−0.5050
Q_{yy}	−0.0201	−0.0229	0.2881	0.2525
Q_{zz}	−0.0201	−0.0229	0.2880	0.2524

[a]See Table 1 for notation.

Table 4. Parameterization Results, RRMS, and Molecular Dipole/Quadrupole Moments of Benzene Fitted with Three Electrostatic Models[a]

	RESP	RESP-ind	RESP-perm	QM
		Charges/a.u.		
C	−0.1123	−0.2464	−0.2227	
H	0.1123	0.2464	0.2227	
		Permanent Dipole Moments/a.u.		
C−H			0.0670	
H−C			0.0074	
C−C			−0.0290	
		RRMS		
	0.2203	0.2570	0.2432	
		Dipole Moments/Debye		
μ	0.0000	0.0000	0.0000	0.0000
		Quadrupole Moments/Debye Angstroms		
Q_{xx}	2.2657	2.3738	2.3203	2.6637
Q_{yy}	2.2655	2.3732	2.3199	2.6627
Q_{zz}	−4.5312	−4.7470	−4.6403	−5.3264

[a]See Table 1 for notation.

Table 5. Charges, RRMS, and Molecular Dipole/Quadrupole Moments of N-Methyl Acetamide (NMA) Fitted with Three Electrostatic Models[a]

	RESP	RESP-ind	RESP-perm	QM
		Charges/a.u.		
C1	−0.4202	−0.3524	−0.4778	
H1	0.1113	0.1422	0.1347	
C	0.6515	1.1283	1.0510	
O	−0.5297	−0.9953	−0.8081	
N	−0.4249	−1.1062	−0.5250	
H	0.2848	0.6127	0.2715	
C2	−0.3419	−0.1267	−0.1219	
H2	0.1488	0.1377	0.0687	
		RRMS		
	0.1029	0.0812	0.0786	
		Dipole Moments/Debye		
μ	3.8335	3.6657	3.6502	3.8004
		Quadrupole Moments/Debye Angstroms		
Q_{xx}	3.6515	3.1427	3.4849	3.6815
Q_{yy}	−0.7200	−0.3841	−0.6802	−0.7850
Q_{zz}	−2.9315	−2.7586	−2.8047	−2.8965

[a]See Table 1 for notation.

Table 6. Parameterization Results, RRMS, and Molecular Dipole/Quadrupole Moments of Dimethyl Phosphate (DMP) Fitted with Three Electrostatic Models[a]

	RESP	RESP-ind	RESP-perm	QM
		Charges/a.u.		
P	1.1047	0.5525	−0.4188	
O1 (O=)	−0.7411	−0.6776	−0.3424	
O2 (−O−)	−0.4399	−0.4920	−0.1201	
C	0.0553	0.0987	−0.2107	
H	0.0244	0.0982	0.1276	
		RRMS		
	0.0196	0.0161	0.0117	
		Dipole Moments/Debye		
μ	2.4333	2.4494	2.4635	2.5559
		Quadrupole Moments/Debye Angstroms		
Q_{xx}	9.2617	7.5526	8.3254	9.0420
Q_{yy}	−3.5225	−2.8853	−3.4178	−3.6665
Q_{zz}	−5.7392	−4.6673	−4.9076	−5.3755

[a]See Table 1 for notation.

equivalencing applied. For single-conformation fittings, we would like to examine both the differences and consistencies of the parameterizations between the two conformations, and we are interested in which electrostatic model can give the best performance in parameterizing each amino acid. For double-conformation fittings, it can be expected that they will show higher RRMS and lower correlation coefficients compared to single-conformation fittings since the double-conformation fitting needs to accommodate contributions from both conformations to reduce conformational dependence.

Tables 8−11 show the RRMS and moments of alanine dipeptide, serine dipeptide, arginine dipeptide, and aspartic acid dipeptide, respectively, fitted with both single-conformation and double-conformation fittings. The charges and permanent dipole moments are shown in Tables S4−S11. We first focus on the results for single-conformation fittings. For uncharged amino acids alanine and serine, the lowest RRMS is produced by the RESP-perm model for the α-helix conformation and by the RESP-ind model for the β-sheet

conformation. While for charged amino acids arginine and aspartic acid, the RRMS consistently decreases in the order of RESP, RESP-ind, and RESP-perm models for both α-helix and β-sheet conformations. In addition, most α-helix conformation fittings give lower RRMS than that of β-sheet conformation, which might be explained by the fact that amino acids in the α-helix conformation have higher polarity (larger dipole moment) than in the β-sheet conformation. A similar trend was observed in Figures 5 and S1−S3, where the correlation coefficients for the α-helix conformation are mostly higher than that of the β-sheet conformation. The correlation coefficients of single-conformation fittings consistently increase in the order of RESP, RESP-ind, and RESP-perm models for all amino acids in both conformations. The molecular dipole and quadrupole moments show interesting patterns. The RESP-ind model consistently yields the best agreement with QM moments for amino acids in the α-helix conformation. On the other hand, the RESP model yields the worst agreement for
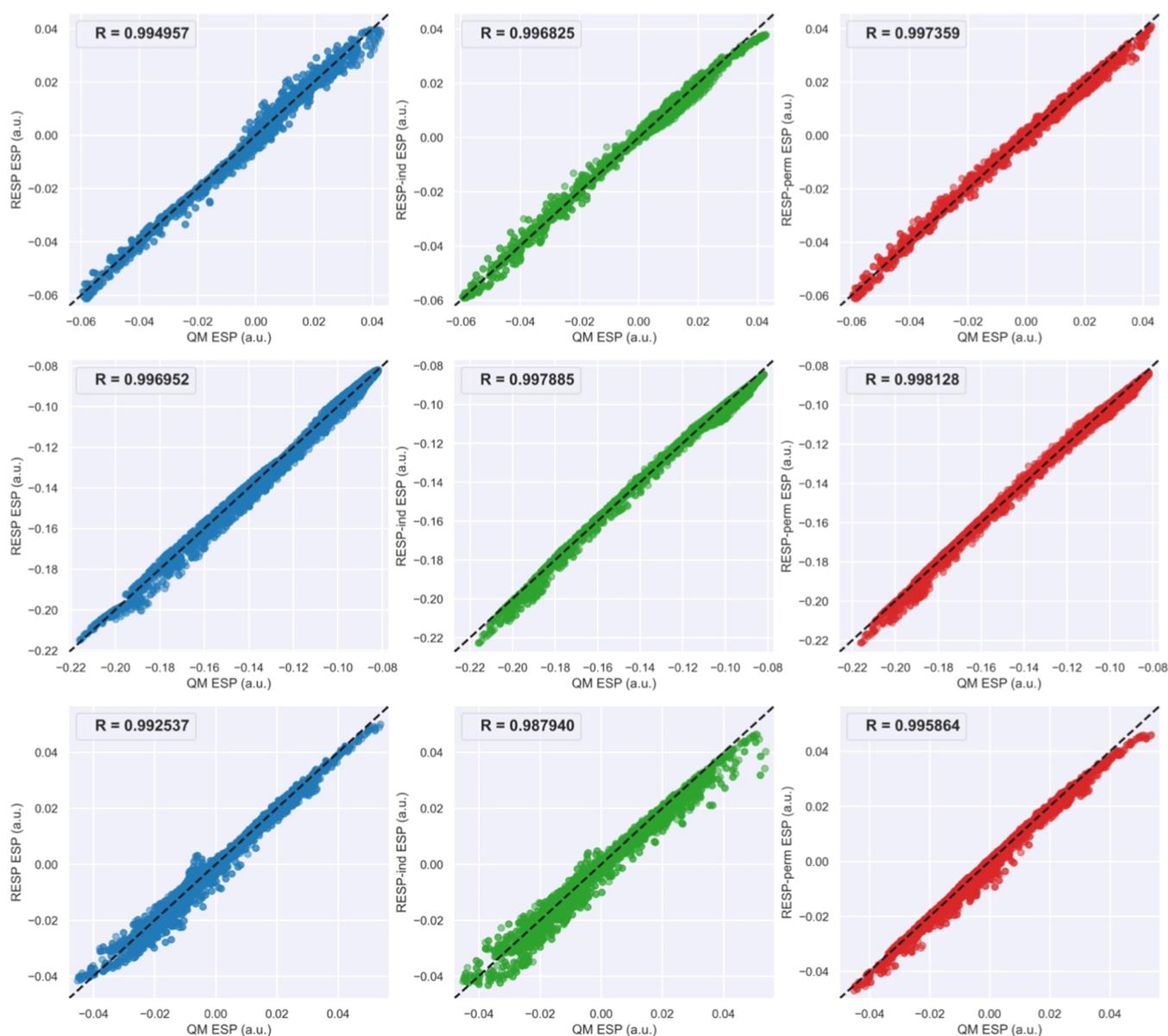
Figure 4. Correlation analysis of QM ESPs and ESPs calculated with various electrostatic models for *N*-methyl acetamide (NMA, upper panel), dimethyl phosphate (DMP, middle panel), and adenine (lower panel) molecules. NMA, DMP, and adenine molecules were fitted with 4159, 4847, and 5155 ESP data points, respectively. The dashed lines correspond to a perfect correlation. *R* is the Pearson correlation coefficient.

the α-helix conformation but yields the best agreement for the β-sheet conformation.

Next, we compare the results of double-conformation fittings with those of single-conformation fittings. Surprisingly, in contrast to the expectation that double-conformation fittings will always produce higher RRMS and lower correlation coefficients compared to single-conformation fittings, the double-conformation fittings of the RESP-perm model consistently give lower RRMS and higher correlation coefficients than those of single-conformation fittings for all amino acids in both conformations and so is the RESP model for amino acids in the α-helix conformation. Next, the molecular dipole and quadrupole moments of double- and single-conformation fittings are compared. Interestingly, most double-conformation fittings result in better agreement with the QM-calculated moments than those of single-conformation fittings for the α-helix conformation but result in worse

agreements for the β-sheet conformation. In particular, the RESP-perm model is the only model that improves the molecular moment qualities for all amino acids in both α-helix and β-sheet conformations.

■ DISCUSSION AND CONCLUSIONS

We have developed and implemented the *PyRESP* program for flexible force field parameterizations with four electrostatic models: RESP, RESP-ind, RESP-perm, and RESP-perm-v. The RESP model is a Python implementation of the original *RESP* program in the Fortran language.[43,44] Compared with previous ESP-based charge derivation methods,[37−39,47,48] the RESP model reduces the overall magnitude of the charges using a simple hyperbolic restraining function, which improves the transferability of fitted charges and reduces the conformational dependency problem. The RESP-ind, RESP-perm, and RESP-perm-v models were designed and implemented in a consistent

Table 7. Charges, RRMS, and Molecular Dipole/Quadrupole Moments of Adenine Fitted with Three Electrostatic Models[a]

	RESP	RESP-ind	RESP-perm	QM
		Charges/a.u.		
N1[b]	−0.7086	−2.0082	−0.0586	
C2[b]	0.4549	1.6084	−0.1038	
H2[b]	0.0770	0.3283	0.0701	
N3[b]	−0.7256	−2.5767	−0.1907	
C4[b]	0.6413	2.4364	0.1796	
C5[b]	0.0209	0.0431	0.1477	
C6[b]	0.6856	2.2390	0.4396	
N6[b]	−0.9046	−2.2041	−1.4981	
HN6[b]	0.4054	0.7019	0.5695	
N7[b]	−0.5608	−1.7370	−0.0397	
C8[b]	0.2693	1.2954	−0.0643	
H8[b]	0.1199	0.4007	−0.1734	
N9[b]	−0.5699	−1.9989	−0.3756	
HN9[b]	0.3898	0.7698	0.5283	
		RRMS		
	0.1263	0.1661	0.1043	
		Dipole Moments/Debye		
μ	2.5562	2.5856	2.4726	2.4994
		Quadrupole Moments/Debye Angstroms		
Q_{xx}	12.3287	12.0435	12.5849	12.7410
Q_{yy}	−5.7358	−6.0081	−5.6209	−6.0143
Q_{zz}	−6.5930	−6.0354	−6.9640	−6.7266

[a]See Table 1 for notation. [b]The atom names are from the adenine obtained from Protein Data Bank (ligand ID: ADE).

manner as the RESP model, with the additional modeling of atomic induced dipole moments, atomic permanent dipole moments, and atomic permanent virtual dipole moments, respectively. The Lagrange constraints as well as the intra- and intermolecular equivalencing schemes developed in the original RESP work were also implemented for the latter three models in *PyRESP*.

A variety of molecules were tested with various electrostatic models implemented in *PyRESP*. All molecules were parameterized using the standard two-stage approach proposed by the original RESP work.[43] The 1-2 and 1-3 interactions were included for all polarizable models, and the pGM damping function was applied to all electrostatic interactions both to avoid the polarization catastrophe and to achieve adequate anisotropic molecular response.[33,34,36] It can be observed that for each molecule, most charges fitted with the RESP-ind model have a higher magnitude than those of the RESP model. This is due to the polarization effect among atoms. Taking the water molecule as an example, the electric field at the position of the oxygen atom caused by the positively charged hydrogen atom points outside the molecule along the symmetric axis, which generates an induced dipole in the same direction. The dipole generates positive ESP at the outward direction of the oxygen atom, which cancels out certain amounts of ESP caused by the negatively charged oxygen atom. To compensate this effect, a negative charge with a higher magnitude was fitted to the oxygen atom. On the other hand, the magnitudes of charges fitted by the RESP-perm model do not show consistent trend when compared to those of the RESP-ind model. The charges with the RESP-perm model have higher magnitudes than those of the RESP-ind model for the water molecule, but the opposite is true for ethane and benzene molecules. The magnitude of charges with the RESP-perm model is directly affected by the directions of induced dipole moments and permanent dipole moments. If they point in the same direction, the charge magnitude will increase to compensate the combined effects of induced and permanent dipole moments. If they point in opposite directions, the cancel-off effect of polarization becomes weaker, leading to a lower magnitude of charges.

Among the molecules tested in this work, the parameterizations of the ethane molecule resulted in the highest RRMS and lowest correlation coefficients. This is not only because of its nonpolar nature but also because of the fact that it contains only weak electronegative elements carbon and hydrogen. Figure 3 shows that the ESPs around the ethane molecule are very close to 0 a.u., with a range between −0.005 and 0.006 a.u. The low magnitude of ESP makes the parameterization process sensitive to noise, so that models with a high degree of freedom like RESP-perm are needed to give reasonable fitting. Another molecule that none of the models gave satisfactory performances is benzene, also a nonpolar molecule. The difficulty in parameterizing benzene likely comes from the

Table 8. RRMS and Molecular Dipole/Quadrupole Moments of Alanine Dipeptide (Single and Double Conformations) Fitted with Three Electrostatic Models[a]

conformation		single-conformation fitting			double-conformation fitting			QM
		RESP	RESP-ind	RESP-perm	RESP	RESP-ind	RESP-perm	
				RRMS				
α-helix		0.0929	0.0551	0.0552	0.0854	0.0602	0.0432	
β-sheet		0.1210	0.0852	0.0870	0.1431	0.0939	0.0732	
				Dipole Moments/Debye				
α-helix	μ	7.2117	6.9530	6.8602	7.1200	6.9617	6.9060	7.0313
β-sheet	μ	0.7805	0.6641	0.6809	0.7759	0.6016	0.6166	0.6963
				Quadrupole Moments/Debye Angstroms				
α-helix	Q_{xx}	8.5529	7.5041	7.4211	8.3689	7.8608	7.9172	8.1763
	Q_{yy}	−0.8103	0.1479	0.7630	−0.3067	−0.2380	0.1786	0.1868
	Q_{zz}	−7.7425	−7.6519	−8.1841	−8.0622	−7.6229	−8.0958	−8.3630
β-sheet	Q_{xx}	14.6902	14.1491	14.1046	13.8200	14.0687	14.0705	14.9055
	Q_{yy}	3.9444	3.4097	3.1843	3.7454	3.4612	3.5775	3.4394
	Q_{zz}	−18.6346	−17.5588	−17.2889	−17.5654	−17.5299	−17.6481	−18.3449

[a]See Table 1 for notation.

Table 9. RRMS and Molecular Dipole/Quadrupole Moments of Serine Dipeptide (Single and Double Conformations) Fitted with Three Electrostatic Models[a]

conformation		single-conformation fitting			double-conformation fitting			QM
		RESP	RESP-ind	RESP-perm	RESP	RESP-ind	RESP-perm	
				RRMS				
α-helix		0.1092	0.0583	0.0544	0.1015	0.0638	0.0456	
β-sheet		0.1169	0.0719	0.0768	0.1283	0.0800	0.0627	
				Dipole Moments/Debye				
α-helix	μ	7.2984	7.0225	6.8966	7.1918	7.0907	6.9997	7.0311
β-sheet	μ	1.6728	1.7070	1.6607	1.6197	1.6176	1.6159	1.6838
				Quadrupole Moments/Debye Angstroms				
α-helix	Q_{xx}	4.9007	4.6936	5.0199	4.8222	4.4699	4.6349	4.5426
	Q_{yy}	3.1930	3.3143	3.2288	3.3371	3.6943	3.6715	3.9228
	Q_{zz}	−8.0937	−8.0079	−8.2487	−8.1593	−8.1642	−8.3065	−8.4653
β-sheet	Q_{xx}	14.1477	13.2178	13.0126	13.0852	13.1921	13.3675	14.0962
	Q_{yy}	6.7942	6.8604	6.8887	6.8816	6.7819	6.7311	6.5504
	Q_{zz}	−20.9419	−20.0782	−19.9014	−19.9668	−19.9740	−20.0986	−20.6466

[a]See Table 1 for notation.

Table 10. RRMS and Molecular Dipole/Quadrupole Moments of Arginine Dipeptide (Single and Double Conformations) Fitted with Three Electrostatic Models[a]

conformation		single-conformation fitting			double-conformation fitting			QM
		RESP	RESP-ind	RESP-perm	RESP	RESP-ind	RESP-perm	
				RRMS				
α-helix		0.0236	0.0163	0.0133	0.0256	0.0176	0.0129	
β-sheet		0.0185	0.0164	0.0148	0.0226	0.0177	0.0128	
				Dipole Moments/Debye				
α-helix	μ	24.6060	24.6416	24.5407	24.5512	24.5354	24.4555	24.4666
β-sheet	μ	17.0782	17.2996	17.2833	17.2463	17.4137	17.4077	17.0905
				Quadrupole Moments/Debye Angstroms				
α-helix	Q_{xx}	70.9219	71.3031	71.9070	70.4120	71.3074	71.3896	71.3708
	Q_{yy}	−26.1898	−25.8375	−26.0332	−25.4176	−25.5352	−25.5844	−25.8057
	Q_{zz}	−44.7321	−45.4656	−45.8738	−44.9944	−45.7722	−45.8052	−45.5651
β-sheet	Q_{xx}	79.1043	79.2100	79.3335	79.4498	79.1450	79.5268	79.4586
	Q_{yy}	−27.5596	−28.3617	−28.3538	−28.8008	−28.7891	−28.5127	−27.9418
	Q_{zz}	−51.5447	−50.8484	−50.9797	−50.6490	−50.3559	−51.0142	−51.5168

[a]See Table 1 for notation.

Table 11. RRMS and Molecular Dipole/Quadrupole Moments of Aspartic Acid Dipeptide (Single and Double Conformations) Fitted with Three Electrostatic Models[a]

conformation		single-conformation fitting			double-conformation fitting			QM
		RESP	RESP-ind	RESP-perm	RESP	RESP-ind	RESP-perm	
				RRMS				
α-helix		0.0238	0.0159	0.0126	0.0232	0.0164	0.0118	
β-sheet		0.0253	0.0156	0.0134	0.0259	0.0162	0.0125	
				Dipole Moments/Debye				
α-helix	μ	10.1754	9.8314	9.8469	10.0132	9.7958	9.8477	9.9939
β-sheet	μ	9.3896	9.2885	9.2847	9.5458	9.3331	9.2629	9.4228
				Quadrupole Moments/Debye Angstroms				
α-helix	Q_{xx}	21.5346	22.1268	22.4423	21.8971	22.0499	22.4577	22.6697
	Q_{yy}	16.9975	16.1860	15.7439	15.9289	16.2698	16.0475	16.3449
	Q_{zz}	−38.5321	−38.3128	−38.1862	−37.8260	−38.3197	−38.5053	−39.0145
β-sheet	Q_{xx}	27.8352	26.9809	26.9421	27.8597	26.9899	27.0867	27.8433
	Q_{yy}	−4.5477	−3.2223	−3.2344	−4.3845	−3.0961	−3.2715	−3.5950
	Q_{zz}	−23.2874	−23.7587	−23.7077	−23.4752	−23.8938	−23.8152	−24.2483

[a]See Table 1 for notation.

existence of the π orbital lying outside the ring plane, which cannot be modeled adequately even with the induced and permanent dipole moments, since they are both located on the two-dimensional plane. This is an inherent limitation of the
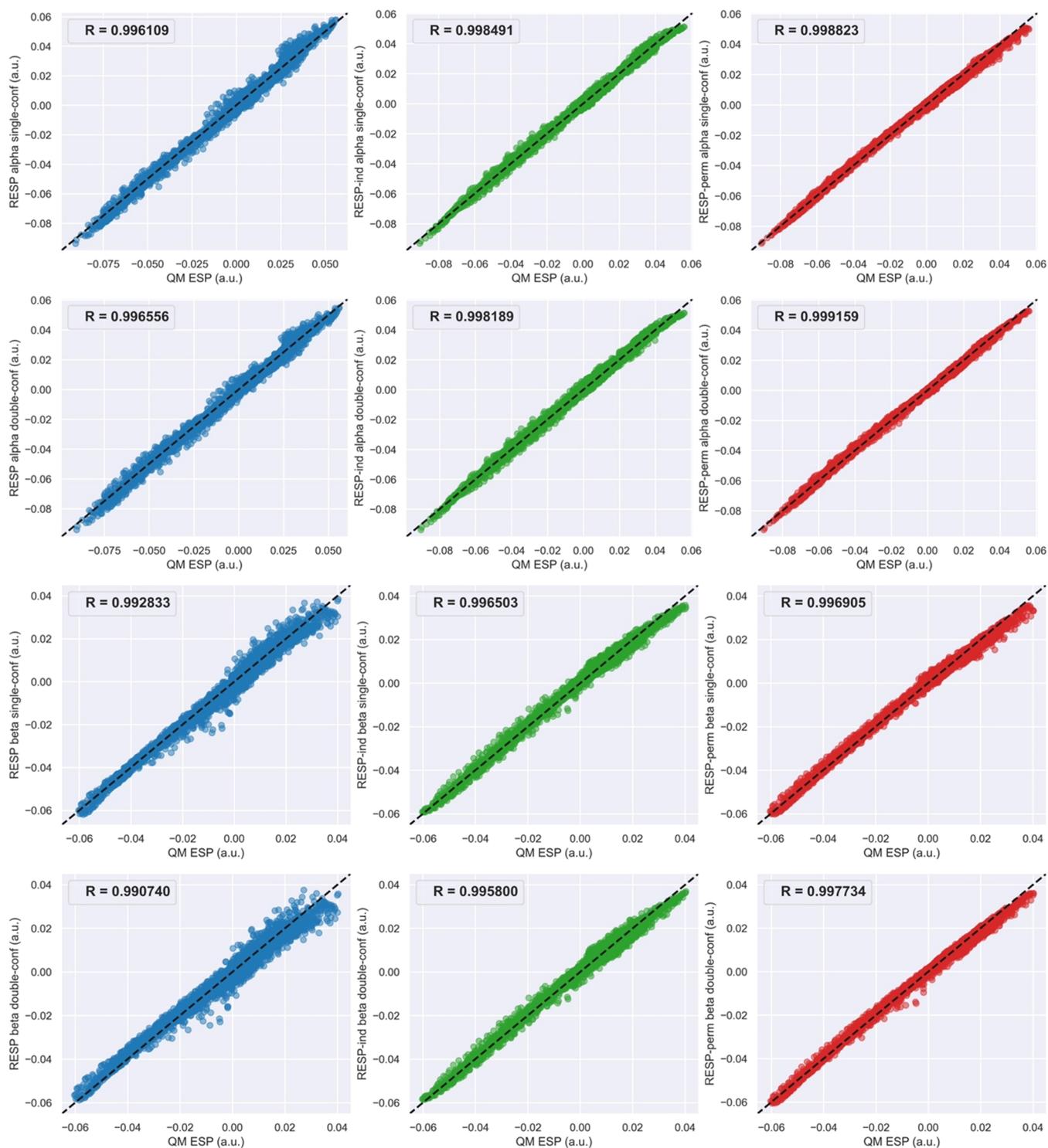
Figure 5. Correlation analysis of QM ESPs and ESPs calculated with various electrostatic models for alanine dipeptide using single- and double-conformation fittings. First row: α-helix conformation fitted with single conformation; second row: α-helix conformation fitted with double conformation; third row: β-sheet conformation fitted with single conformation; and fourth row: β-sheet conformation fitted with double conformation. The α-helix conformation was fitted with 6292 ESP data points, and the β-sheet conformation was fitted with 6460 ESP data points. The dashed lines correspond to a perfect correlation. R is the Pearson correlation coefficient.

current model, which may be improved by adding additional fitting centers outside the aromatic ring or by fitting permanent quadrupole moments in addition to permanent charges and dipoles. Therefore, modeling aromatic molecules remains a challenge even for polarizable force field developments.

The RESP-perm model has a higher degree of freedom than the RESP and RESP-ind models due to the addition of permanent dipole moments; the addition of virtual bonds increases the degree of freedom for the RESP-perm-v model even further. For most molecules tested here, the parameterizations with the RESP-perm/RESP-perm-v models resulted in

lower RRMS, higher correlation coefficients, and molecular moments agree better with QM calculations. However, the quadrupole moments of methanol, NMA, and DMP molecules fitted by the RESP-perm model clearly agree worse with QM results than those fitted by the RESP model. This raises the concern of the overfitting problem when the model degree of freedom is so high that noise starts to diminish fitting accuracy, leading to the deteriorated overall fitting quality. Among the metrics used here to evaluate models, the RRMS and correlation coefficients are highly correlated with the objective function to be minimized in eq 13, so that low RRMS and high correlation coefficients are not reliable enough to eliminate the concerns of overfitting. Therefore, while performing molecule parameterizations using electrostatic models with a high degree of freedom, it is critical to inspect the final molecular dipole and quadrupole moments to determine if the overfitting occurred.

We tested several amino acid dipeptide molecules using both single- and double-conformation fittings. The α-helix ($\phi = 300°$, $\psi = 300°$) and antiparallel β-sheet ($\phi = 240°$, $\psi = 120°$) conformations were selected since they are two of the most frequently found conformations for amino acids in proteins, and they represent considerably different electrostatic properties (e.g., notably different dipole moments). For single-conformation fittings, the RESP-ind model consistently yields the best agreement with QM moments for amino acids in the α-helix conformation, while the RESP model yields the best agreement for the β-sheet conformation. The RESP-perm model that has the highest degree of freedom shows the lowest RRMS and highest correlation coefficients but does not outperform other models in terms of reproducing QM molecular moments. Double-conformation fittings were expected to have poorer performances than those of single-conformation fittings. Surprisingly, double-conformation fittings with the RESP-perm model consistently show better overall performances than the single-conformation fittings for amino acids in both conformations, as illustrated by the lower RRMS, higher correlation coefficients, and moments agree better with QM results. This shows that the double-conformation fittings are necessary for amino acids fitted with the RESP-perm model. For future polarizable force field parameterizations, more conformations are expected to be included to further reduce conformational dependence of the parameters.

In conclusion, the *PyRESP* program developed here is a flexible, efficient, and user-friendly tool that is recommended for parameterizations of various additive and polarizable force fields. *PyRESP* has been released as an open-source software within AmberTools 2022 under the GNU General Public License, available for download from http://ambermd.org/[52]. Documentation and tutorials will also be made available on the Amber website. Alternatively, the standalone version of *PyRESP* with the latest updates is available through https://github.com/ShijiZ/PyRESP.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00230.

> Permanent dipole moments (a.u.) of *N*-methyl acet-amide (NMA) fitted with the RESP-perm electrostatic model (Table S1); permanent dipole moments (a.u.) of dimethyl phosphate (DMP) fitted with the RESP-perm electrostatic model (Table S2); permanent dipole moments (a.u.) of adenine fitted with the RESP-perm electrostatic model (Table S3); charges (a.u.) of alanine dipeptide (single and double conformations) fitted with three electrostatic models (Table S4); permanent dipole moments (a.u.) of alanine dipeptide (single and double conformations) fitted with the RESP-perm electrostatic model (Table S5); charges (a.u.) of serine dipeptide (single and double conformations) fitted with three electrostatic models (Table S6); permanent dipole moments (a.u.) of serine dipeptide (single and double conformations) fitted with the RESP-perm electrostatic model (Table S7); correlation analysis of QM ESPs and ESPs calculated with various electrostatic models for serine dipeptide using single- and double-conformation fittings (Figure S1); charges (a.u.) of arginine dipeptide (single and double conformations) fitted with three electrostatic models (Table S8); permanent dipole moments (a.u.) of arginine dipeptide (single and double conformations) fitted with the RESP-perm electrostatic model (Table S9); correlation analysis of QM ESPs and ESPs calculated with various electrostatic models for arginine dipeptide using single- and double-conformation fittings (Figure S2); charges (a.u.) of aspartic acid dipeptide (single and double conformations) fitted with three electrostatic models (Table S10); permanent dipole moments (a.u.) of aspartic acid dipeptide (single and double conformations) fitted with the RESP-perm electrostatic model (Table S11); and correlation analysis of QM ESPs and ESPs calculated with various electrostatic models for aspartic acid dipeptide using single- and double-conformation fittings (Figure S3) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Yong Duan − *UC Davis Genome Center and Department of Biomedical Engineering, University of California, Davis, Davis, California 95616, United States;* ⓞ orcid.org/0000-0003-3793-5099; Email: duan@ucdavis.edu

Ray Luo − *Departments of Molecular Biology and Biochemistry, Chemical and Biomolecular Engineering, Materials Science and Engineering, and Biomedical Engineering, University of California, Irvine, Irvine, California 92697, United States;* ⓞ orcid.org/0000-0002-6346-8271; Email: rluo@uci.edu

Authors

Shiji Zhao − *Departments of Molecular Biology and Biochemistry, Chemical and Biomolecular Engineering, Materials Science and Engineering, and Biomedical Engineering, University of California, Irvine, Irvine, California 92697, United States;* ⓞ orcid.org/0000-0002-4514-0897

Haixin Wei − *Departments of Molecular Biology and Biochemistry, Chemical and Biomolecular Engineering, Materials Science and Engineering, and Biomedical Engineering, University of California, Irvine, Irvine, California 92697, United States;* ⓞ orcid.org/0000-0002-0954-4560

Piotr Cieplak − *SBP Medical Discovery Institute, La Jolla, California 92037, United States;* ⊙ orcid.org/0000-0003-0700-5691

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.2c00230

Author Contributions

S.Z. designed and implemented the *PyRESP* program, performed *ab initio* calculations, parameterizations, and data analysis, and wrote the original manuscript. H.W. participated in the *PyRESP* program design. P.C., D.Y., and R.L. directed the project and supported the manuscript revision. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.

(2) Tian, C.; Kasavajhala, K.; Belfon, K. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **2020**, *16*, 528−552.

(3) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(4) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(5) Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *J. Phys.: Condens. Matter* **2009**, *21*, No. 333102.

(6) Dill, K. A.; Bromberg, S.; Yue, K.; Chan, H. S.; Fiebig, K. M.; Yee, D. P.; Thomas, P. D. Principles of protein folding—a perspective from simple exact models. *Protein Sci.* **2008**, *4*, 561−602.

(7) Fitch, C. A.; Karp, D. A.; Lee, K. K.; Stites, W. E.; Lattman, E. E.; García-Moreno, E. B. Experimental pKa values of buried residues: analysis with continuum methods and role of water penetration. *Biophys. J.* **2002**, *82*, 3289−3304.

(8) Zhao, S.; Schaub, A. J.; Tsai, S.-C.; Luo, R. Development of a Pantetheine Force Field Library for Molecular Modeling. *J. Chem. Inf. Model.* **2021**, *61*, 856−868.

(9) Lin, F. Y.; MacKerell, A. D., Jr. Improved Modeling of Cation-π and Anion-Ring Interactions Using the Drude Polarizable Empirical Force Field for Proteins. *J. Comput. Chem.* **2020**, *41*, 439−448.

(10) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X. X.; Murphy, R. B.; Zhou, R. H.; Halgren, T. A. Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *J. Comput. Chem.* **2002**, *23*, 1515−1531.

(11) Friesner, R. A. Modeling Polarization in Proteins and Protein−ligand Complexes: Methods and Preliminary Results. In *Advances in Protein Chemistry*, Academic Press, 2005; Vol. 72, pp 79−104.

(12) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 245−249.

(13) Lopes, P. E. M.; Lamoureux, G.; Roux, B.; MacKerell, A. D. Polarizable empirical force field for aromatic compounds based on the classical drude oscillator. *J. Phys. Chem. B* **2007**, *111*, 2873−2885.

(14) Patel, S.; Mackerell, A. D.; Brooks, C. L. CHARMM fluctuating charge force field for proteins: II - Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **2004**, *25*, 1504−1514.

(15) Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D., Jr.; Schulten, K.; Roux, B. High-Performance Scalable Molecular Dynamics Simulations of a Polarizable Force Field Based on Classical Drude Oscillators in NAMD. *J. Phys. Chem. Lett.* **2011**, *2*, 87−92.

(16) Kumar, A.; Pandey, P.; Chatterjee, P.; MacKerell, A. D., Jr. Deep Neural Network Model to Predict the Electrostatic Parameters in the Polarizable Classical Drude Oscillator Force Field. *J. Chem. Theory Comput.* **2022**, *18*, 1711−1725.

(17) Cieplak, P.; Caldwell, J.; Kollman, P. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J. Comput. Chem.* **2001**, *22*, 1048−1057.

(18) Wang, Z. X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. Strike a balance: optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J. Comput. Chem.* **2006**, *27*, 781−790.

(19) Tan, Y.-H.; Luo, R. Continuum treatment of electronic polarization effect. *J. Chem. Phys.* **2007**, *126*, No. 094103.

(20) Tan, Y.-H.; Tan, C.; Wang, J.; Luo, R. Continuum polarizable force field within the Poisson-Boltzmann framework. *J. Phys. Chem. B* **2008**, *112*, 7675−7688.

(21) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(22) Vesely, F. J. N-particle dynamics of polarizable Stockmayer-type molecules. *J. Comput. Phys.* **1977**, *24*, 361−371.

(23) Ren, P.; Ponder, J. W. Consistent treatment of inter-and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.* **2002**, *23*, 1497−1506.

(24) Ren, P.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003**, *107*, 5933−5947.

(25) 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*, 3091−3099.

(26) 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*, 3100−3111.

(27) Wang, J.; Cieplak, P.; Cai, Q.; Hsieh, M.-J.; Wang, J.; Duan, Y.; Luo, R. Development of polarizable models for molecular mechanical calculations. 3. Polarizable water models conforming to Thole polarization screening schemes. *J. Phys. Chem. B* **2012**, *116*, 7999−8008.

(28) Wang, J.; Cieplak, P.; Li, J.; Cai, Q.; Hsieh, M.-J.; Luo, R.; Duan, Y. Development of polarizable models for molecular mechanical calculations. 4. van der Waals parametrization. *J. Phys. Chem. B* **2012**, *116*, 7088−7101.

(29) Applequist, J.; Carl, J. R.; Fung, K.-K. Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *J. Am. Chem. Soc.* **1972**, *94*, 2952−2960.

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

(31) Van Duijnen, P. T.; Swart, M. Molecular and atomic polarizabilities: Thole's model revisited. *J. Phys. Chem. A* **1998**, *102*, 2399−2407.

(32) Ponder, J. W. *TINKER: Software Tools for Molecular Design*; Washington University School of Medicine: Saint Louis, MO, 2004, Vol. 3.

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

(34) Wei, H.; Qi, R.; Wang, J.; Cieplak, P.; Duan, Y.; Luo, R. Efficient formulation of polarizable Gaussian multipole electrostatics for biomolecular simulations. *J. Chem. Phys.* **2020**, *153*, No. 114116.

(35) Wei, H.; Cieplak, P.; Duan, Y.; Luo, R. Stress Tensor and Constant Pressure Simulation for Polarizable Gaussian Multipole Model. *J. Chem. Phys.* **2022**, *156*, 114114.

(36) Wang, J.; Cieplak, P.; Luo, R.; Duan, Y. Development of Polarizable Gaussian Model for Molecular Mechanical Calculations I: Atomic Polarizability Parameterization To Reproduce ab Initio Anisotropy. *J. Chem. Theory Comput.* **2019**, *15*, 1146−1158.

(37) Momany, F. A. Determination of partial atomic charges from ab initio molecular electrostatic potentials. Application to formamide, methanol, and formic acid. *J. Phys. Chem. A* **1978**, *82*, 592−601.

(38) Cox, S. R.; Williams, D. Representation of the molecular electrostatic potential by a net atomic charge model. *J. Comput. Chem.* **1981**, *2*, 304−323.

(39) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.

(40) Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548−558.

(41) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995**, *16*, 1357−1377.

(42) 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*, 5179−5197.

(43) 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. B* **1993**, *97*, 10269−10280.

(44) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.

(45) Konishi, S.; Kashiwagi, Y.; Watanabe, G.; Osaki, M.; Katashima, T.; Urakawa, O.; Inoue, T.; Yamaguchi, H.; Harada, A.; Takashima, Y. Design and mechanical properties of supramolecular polymeric materials based on host−guest interactions: the relation between relaxation time and fracture energy. *Polym. Chem.* **2020**, *11*, 6811−6820.

(46) Wang, X.; Gao, J. Atomic partial charge predictions for furanoses by random forest regression with atom type symmetry function. *RSC Adv.* **2020**, *10*, 666−673.

(47) Chirlian, L. E.; Francl, M. M. Atomic charges derived from electrostatic potentials: A detailed study. *J. Comput. Chem.* **1987**, *8*, 894−905.

(48) 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.

(49) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990**, *11*, 431−439.

(50) 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. J.; 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, N. J.; 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, 2009.

(51) Xie, W.; Pu, J.; Gao, J. A coupled polarization-matrix inversion and iteration approach for accelerating the dipole convergence in a polarizable potential function. *J. Phys. Chem. A* **2009**, *113*, 2109−2116.

(52) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *Amber 2021*; University of California: San Francisco, 2021.

📖 Recommended by ACS