

UC Irvine

UC Irvine Previously Published Works

Title

Refinement of Atomic Polarizabilities for a Polarizable Gaussian Multipole Force Field with Simultaneous Considerations of Both Molecular Polarizability Tensors and In-Solution Electrostatic Potentials.

Permalink

<https://escholarship.org/uc/item/4709k5ht>

Journal

Journal of chemical information and computer sciences, 65(3)

Authors

Duan, Yong

Wang, Junmei

Cieplak, Piotr

et al.

Publication Date

2025-02-10

DOI

10.1021/acs.jcim.4c02175

Peer reviewed

Refinement of Atomic Polarizabilities for a Polarizable Gaussian Multipole Force Field with Simultaneous Considerations of Both Molecular Polarizability Tensors and In-Solution Electrostatic Potentials

Yong Duan,* Junmei Wang, Piotr Cieplak, and Ray Luo



Cite This: *J. Chem. Inf. Model.* 2025, 65, 1428–1440



Read Online

ACCESS |



Metrics & More



Article Recommendations



Supporting Information

ABSTRACT: Atomic polarizabilities are considered to be fundamental parameters in polarizable molecular mechanical force fields that play pivotal roles in determining model transferability across different electrostatic environments. In an earlier work, the atomic polarizabilities were obtained by fitting them to the B3LYP/aug-cc-pvtz molecular polarizability tensors of mainly small molecules. Taking advantage of the recent PCMRESPPOL method, we refine the atomic polarizabilities for condensed-phase simulations using a polarizable Gaussian Multipole (pGM) force field. Departing from earlier works, in this work,

we incorporated polarizability tensors of a large number of dimers and electrostatic potentials (ESPs) in multiple solvents. We calculated 1565×4 ESPs of small molecule monomers and dimers of noble gas and small molecules and 4742×4 ESPs of small molecule dimers in four solvents (diethyl ether, $\epsilon = 4.24$, dichloroethane, $\epsilon = 10.13$, acetone, $\epsilon = 20.49$, and water, $\epsilon = 78.36$). For the gas-phase polarizability tensors, we supplemented the molecule set that was used in our earlier work by adding both the 4252 monomer and dimer sets studied by Shaw and co-workers and the 7211 small molecule monomers listed in the QM7b database to a combined total of 13,523 molecular polarizability tensors of monomers and dimers. The QM7b polarizability set was obtained from quantum-machine.org and was calculated at the LR-CCSD/d-aug-cc-pVDZ level of theory. All other polarizability tensors and all ESPs were calculated at the ω B97X-D/aug-cc-pVTZ level of theory. The atomic polarizabilities were developed using all polarizability tensors and the 1565×4 ESPs of small molecule monomers and were then assessed by comparing them to the 4742×4 ab initio ESPs of small molecule dimers. The predicted dimer ESPs had an average relative root-mean-square error (RRMSE) of 9.30%, which was only slightly larger than the average fitting RRMSE of 9.15% of the monomer ESPs. The transferability of the polarizability set was further evaluated by comparing the ESPs calculated using parameters developed in another dielectric environment for both tetrapeptide and DES monomer data sets. It was observed that the polarizabilities of this work retained or slightly improved the transferability over the one discussed in earlier work even though the number of parameters in the present set is about half of that in the earlier set. Excluding the gas-phase data, for the DES monomer set, the average transfer RRMSEs were 16.25% and 10.83% for pGM-ind and pGM-perm methods, respectively, comparable to the average fitting RRMSEs of 16.03% and 10.54%; for tetrapeptides, the average transfer RRMSEs were 5.62% and 3.95% for pGM-ind and pGM-perm methods, respectively, slightly larger than 5.41% and 3.61% of the fitting RRMSEs. Therefore, we conclude that the pGM methods with updated polarizabilities achieved remarkable transferability from monomer to dimer and from one solvent to another.

	GAS	ETH	DCE	ACT	WAT	AVE
GAS	4.45	5.97	5.40	4.55	4.79	5.18
ETH	6.75	3.94	3.76	4.07	3.81	4.60
DCE	6.48	4.01	3.68	3.84	3.52	4.46
ACT	5.52	4.44	3.93	3.60	3.50	4.35
WAT	6.47	4.59	3.99	3.88	3.21	4.73

Average RRMSE (%) color scale: 4 (blue) to 6 (red)

INTRODUCTION

A variety of methods have been explored to accurately account for the contributions from atomic polarization, including induced point dipole (IPD) models,^{1–4} Drude oscillator models,⁵ and fluctuating charge models⁶ and the ACKS2⁷ implementation and explicit electron approach⁸ in the reactive force field ReaxFF.⁹ In the IPD polarizable models, the atomic polarization is modeled as an IPD in response to an electrostatic field. Examples of the IPD models include those in AMBER,^{1,2} AMOEBA,³ and SIBFA.⁴ In the Thole-style¹⁰ IPD models, the interactions between induced dipoles are attenuated at short

range for stability to circumvent the divergence problem associated with point-induced dipoles.¹¹ To this end, a significant departure from the IPD framework is the polarizable Gaussian Multipole (pGM) model^{12–20} that consistently treats

Received: November 27, 2024

Revised: January 11, 2025

Accepted: January 14, 2025

Published: January 27, 2025



all monopoles and multipoles by Gaussian distribution functions.

The induction energy, also called polarization energy, plays critical roles in the accurate representation of molecular potential energy surfaces. Because molecular conformation fluctuation leads to changes of the electrostatic environments, lacking the induction term, as in the nonpolarizable fixed charge models, can significantly reduce the transferability and accuracy, as demonstrated earlier.^{17,19,20} Furthermore, because induction energy is nonadditive, such deficiency is difficult to compensate by other (additive) terms. In polarizable molecular mechanics force fields, atomic polarizabilities and the associated atomic radii are considered fundamental parameters upon which all other electrostatic parameters, including charges and multipoles, are developed. In practice, polarizable force field development often starts from determining atomic polarizabilities and radii (or the screening constant in the case of Thole-style models). Charges and multipoles that determine the electrostatic components of the force field are obtained thereafter. On the other hand, because the electrostatic components account for the long-range forces and are often approximated by the contributions up to atomic quadrupoles, they can potentially be one of the least transferable parts of a force field. Therefore, accuracy of the atomic polarizabilities and radii can affect the fidelity of reproducing the ab initio electrostatic potential (ESP) as well as the transferability of a polarizable force field across multiple solvation environments.

Because of their importance, efforts have been made by a number of groups, including our own, to develop accurate atomic polarizabilities. Earlier approaches often involved fitting to the molecular polarizabilities as we did earlier.^{2,14} Interestingly, Bosque and Sales²¹ were able to obtain an average fitting error of 2.31% using a simple additive model to fit experimental molecular polarizabilities without considering the intramolecular polarization effect. Their results were rather counterintuitive. Such an excellent agreement and the fact that their model did not consider the intramolecular polarization at all suggest that, perhaps, molecular polarizability alone is insufficient to determine the atomic polarizabilities. Alternatively, one may obtain the polarizabilities by placing probes outside the molecule and perform appropriate fitting to the changes of the ESPs as Elking et al.¹² and Litman et al.²² have done. Here, we propose to use the ESPs generated in multiple solvents by the polarizable continuum model (PCM)²³ for its realism to mimic the solvation environment. Multiple solvent ESPs also provide a natural avenue to evaluate the accuracy and transferability of the model that are intimately related to polarizabilities.

Another important factor that has been overlooked might be the difference between gas-phase and condensed-phase polarizabilities. Because atomic polarizabilities are intimately related to the local population of electron densities, it is possible that effects such as charge transfer that can take place in the condensed phase would be difficult to observe in the gas phase. To alleviate the problem, we now include a set of large number of dimers in our polarizability tensor data set. The combination of a large number of dimer polarizability tensors and multisolvant ESPs is expected to help make the model better suited for condensed-phase simulations.

We have investigated the pGM model^{14–20} in a series of recent studies that was developed based on the work of Elking et al.^{12,13} We found that pGM can model the cooperative effects of the peptide main chain hydrogen bonds accurately¹⁷ and has

excellent transferability across oligomeric states, peptide conformations, and different sequences with varied lengths¹⁹ of amino acid sequences. When we tested the transferability across multiple solvents,²⁰ we found that pGM is highly transferable, making it a promising framework to develop a polarizable force field that can model the ESPs accurately across multiple solvents and conformations, a key impetus of polarizable force field development.

The pGM model is an induced dipole polarizable model. Although pGM can include higher order atomic multipoles, we limit our effort, in this iteration, to develop a force field with atomic charges and permanent and induced atomic dipoles and leave more complex models such as atomic quadrupoles for future efforts. Induced dipole modeling is one of the most extensively studied methods in polarizable force field development. In this method, the induction energy due to polarization of atomic dipoles is

$$V_{\text{ind}} = -\frac{1}{2}\alpha_i \mathbf{E}_i^2 \quad (1)$$

Here, V_{ind} is the potential energy due to induction, α_i is the polarizability, and \mathbf{E}_i is the static external electrostatic field exerted on atom i from the permanent multipoles (charges, permanent dipoles, and, if applicable, higher order multipoles). Note that V_{ind} is many-body in nature and nonlinear to the electrostatic field. Therefore, the potential energy surface of polarizable force fields modeled by induced dipoles are nonadditive. A consequence is that all contributions from monopoles and multipoles need to be accounted for in a consistent manner. This includes short-range terms such as polarization between bonded atoms and long-range terms such as non-bonded electrostatic interactions. We showed in our previous work¹⁴ that the short-range interactions are critical to reproduce molecular anisotropy largely because of the nonlinear term V_{ind} . Furthermore, because of the need to include the short-range static fields, the fields from nearby static monopoles, and multipoles, the short-range fields also need to be attenuated when atoms are close to each other. This includes all of the electrostatic interactions. Failure to do so may lead to unphysically large induced dipoles, akin to what has been termed as “polarization catastrophe”.¹⁰ Partially due to this reason, some of the existing polarizable models have to ignore the short-range interactions, leading to unrealistic molecular anisotropy and inconsistency between intra- and intermolecular interactions. Fortunately, pGM provides a unique property that allows for consistent attenuation for all short-range interactions and fields.

In the induced dipole models, the atomic induced dipoles are calculated by

$$\boldsymbol{\mu}_i = \alpha_i [\mathbf{E}_i - \sum_{j \neq i}^n \mathbf{T}_{ij} \boldsymbol{\mu}_j] \quad (2)$$

$$\mathbf{T}_{ij} = \frac{f_{e,ij}}{r_{ij}^3} \mathbf{I} - \frac{3f_{t,ij}}{r_{ij}^5} \begin{bmatrix} x^2 & xy & xz \\ xy & y^2 & yz \\ xz & yz & z^2 \end{bmatrix} \quad (3)$$

where $\boldsymbol{\mu}_i$ represents the induced dipole of atom i , α_i is its polarizability tensor, \mathbf{E}_i is the electrostatic field from the permanent multipoles acting on atom i , and \mathbf{T}_{ij} is the dipole field tensor. Here, \mathbf{I} is the identity matrix, and x , y , and z are the Cartesian components of the distance vector \mathbf{r}_{ij} between atoms i

and j . f_e and f_t are distance-dependent Thole¹⁰ damping functions that attenuate T_{ij} , preventing the “polarization catastrophe” phenomenon observed in classical Applequist point-induced dipole models¹¹ when interatomic distances r_{ij} become small, where induced dipoles can mutually reinforce each other and impede convergence. However, it is important to note that the strict Thole models only attenuate the induced dipole interactions while still treating other electrostatic terms as interactions between point monopoles and multipoles without attenuation. This can potentially lead to unphysically large fields from short-range point charges and point multipoles when two atoms are in close contact (e.g., ions). As we stated earlier, because the nonlinear induction energy requires a full account of the electrostatic fields, including even the mostly static fields from bonded atoms, reconciling the short-range and long-range contributions in the Thole models remains a challenging task without damping terms other than the induced dipole interactions. Therefore, in practice, one often needs to exclude the short-range interaction terms such as the so-called 1–2 and 1–3 terms. Such ad hoc treatment of the 1–2 and 1–3 terms further reduces the consistency of the model. For example, when 1–2 and 1–3 terms are ignored, the critically important solvent water model would have no intramolecular interactions, making it difficult to delineate permanent and induced dipoles. Fortunately, pGM consistently allows attenuation of all short-range interactions consistently. Therefore, all 1–2 and 1–3 interactions can be explicitly accounted for in the pGM. This is a critical advantage of the pGM over other types of polarizable models.

The pGM model offers a comprehensive framework for coherent modeling of electrostatic interactions. In the pGM model, all multipoles are represented by Gaussian distributions^{12,13} with the n th-order multipole defined by eq 4

$$\rho^{(n)}(\mathbf{r}; \mathbf{R}) = \Theta^{(n)} \cdot \nabla_{\mathbf{R}}^{(n)} \left(\frac{\beta}{\sqrt{\pi}} \right)^3 e^{-\beta^2 |\mathbf{r} - \mathbf{R}|^2} \quad (4)$$

Here, $\rho^{(n)}$ is the distribution of the n th-order multipole, \mathbf{r} is the atomic coordinate, \mathbf{R} is the distance from the atom, $\Theta^{(n)}$ is the n th-order multipole, $\nabla_{\mathbf{R}}^{(n)}$ is the n th-order gradient, and β is the Gaussian factor (see β_{ij} in eq 7). This formulation provides a unified treatment of all multipoles and effectively eliminates the point charges and point multipoles that are the root causes of potential singularities on the energy surface, while also coherently addresses the charge-penetration effect. With this framework, the damping functions are defined as follows

$$f_{e,ij} = \operatorname{erf}(S_{ij}) - \frac{2}{\sqrt{\pi}} S_{ij} \exp(-S_{ij}^2) \quad (5)$$

$$f_{t,ij} = \operatorname{erf}(S_{ij}) - \frac{2}{\sqrt{\pi}} S_{ij} \exp(-S_{ij}^2) \left(1 + \frac{2}{3} S_{ij}^2 \right) \quad (6)$$

$$S_{ij} = \beta_{ij} r_{ij} = \frac{r_{ij}}{\sqrt{2(R_i^2 + R_j^2)}} \quad (7)$$

In these equations, r_{ij} is the distance between atoms i and j , and R_i and R_j are their respective pGM radii of the Gaussian functions.

To calculate the molecular polarizability tensors from the atomic polarizability tensors, we need the total molecular induced dipole, which is a sum of all atomic induced dipoles of the molecule.

$$\boldsymbol{\mu}_{\text{mol}} = \sum_{i=1}^n \boldsymbol{\mu}_i = \boldsymbol{\alpha}_{\text{mol}} \mathbf{E} \quad (8)$$

where $\boldsymbol{\mu}_{\text{mol}}$ is the total induced molecular dipole, $\boldsymbol{\mu}_i$ is the induced dipole of atom i , $\boldsymbol{\alpha}_{\text{mol}}$ is the polarizability tensor of the molecule, and \mathbf{E} is the external static field. Since atomic induced dipoles can be calculated in eq 2 which can be rewritten as

$$\boldsymbol{\alpha}_i^{-1} \boldsymbol{\mu}_i + \sum_{j \neq i}^n \mathbf{T}_{ij} \boldsymbol{\mu}_j = \mathbf{E}_i \quad (9)$$

In matrix form

$$\mathbf{A} = \begin{cases} \mathbf{A}_{ii} = \boldsymbol{\alpha}_i^{-1} \\ \mathbf{A}_{ij} = \mathbf{T}_{ij} \end{cases} \quad \mathbf{A} \boldsymbol{\mu}_{\text{mol}} = \mathbf{E} \quad \boldsymbol{\mu}_{\text{mol}} = \mathbf{A}^{-1} \mathbf{E} \quad (10)$$

The molecular polarizability tensor is

$$\boldsymbol{\alpha}_{\text{mol}} = \sum_{i,j=1}^n \{ \mathbf{A}^{-1} \}_{i,j} \quad (11)$$

Therefore, given a set of molecular polarizability tensors, $\boldsymbol{\alpha}_{\text{mol}}$, the atomic polarizability tensors $\boldsymbol{\alpha}_i$ can be determined. Since \mathbf{T}_{ij} (therefore, \mathbf{A}) depends on the screening functions $f_{e,ij}$ and $f_{t,ij}$, polarizability must depend on the specific screening functions. For example, the polarizabilities derived for Applequist point-induced dipoles that have constant $f_{e,ij} = 1$ and $f_{t,ij} = 1$ are expected to be different from those developed for Thole-type induced dipole polarizable models, which have nonconstant screening functions. Therefore, it is unrealistic to expect a set of universal polarizabilities that are equally applicable to all induced dipole polarizable force fields. In fact, the atomic polarizabilities would need to be developed for each type of $f_{e,ij}$ and $f_{t,ij}$ in a specific polarizable force field, even though with an aim to reproduce the same molecular polarizability tensors.

In recent developments, a python program, PyRESP,¹⁶ has been introduced for electrostatic parameterization for both polarizable and nonpolarizable force fields and a companion tool, PyRESP_GEN,¹⁸ was developed for generating the input files for PyRESP.¹⁶ PCMRESP²⁰ has been developed for explicit consideration of solvent polarizations in electrostatic parameterization using PCM.²⁴ A key departure from the traditional RESP (and PyRESP) is that PCMRESP considers the polarizations due to surface charges explicitly. In this study, we apply PCMRESP to evaluate the transferability of electrostatic parameters across multiple solvents. We also use this method in the fitting of the polarizabilities.

Methods for the development of ESP parameters, including atomic polarizabilities, are based on the molecular properties. Depending on the approach, the methods can be broadly categorized as energy- and density-based decomposition or fitting to molecular ESPs and polarizability tensors. Ideally, the results of these two types of approaches should be in close agreement. However, a large discrepancy can often be observed. For example, in the work of Montilla et al.,²⁵ the atomic polarizability of oxygen in H₂O was 0.723 (Å³), less than twice of hydrogen (0.365 Å³) in H₂O. This was somewhat surprising given the high electronegativity and the presence of the lone electron pairs of oxygen. Thus, we chose to develop the atomic polarizabilities based on both gas-phase molecular polarizability tensors and multisolvant ESPs. A key factor in our reasoning to make such a choice is the observation that atomic polarizabilities play pivotal roles in the transferability of electrostatic parameters

(charges and multipoles) across solvents. In addition, the methods we use in this study will allow for full representation of intramolecular polarization that is a key feature of the pGM framework and has been shown to play critical roles in determining molecular anisotropy in our earlier study.

In an earlier work, we developed a set of atomic polarizabilities² to reproduce the experimental²¹ and ab initio molecular polarizabilities for a 420-molecule data set for a variety of Thole-style¹⁰ IPD models. Subsequently, we refined the atomic polarizabilities¹⁴ based on ab initio polarizability tensors of 4842 molecules and dimers, including amino acid analogues, dipeptides, tetrapeptides, nucleic acid bases, and water clusters. In this work, we combine more than 13,000 molecular polarizability tensors and more than 6000×4 molecular ESPs in four different solvents. We chose to use a least-squares fitting process in which the first derivatives were calculated numerically to guide the iterative search. The developed polarizabilities were subjected to evaluations in two stages with a focus on the transferability. In the first stage, the four-solvent ESPs of dimers were calculated using the parameters obtained from monomer ESPs and compared to the ab initio ESPs. This test examined the monomer to dimer transferability. In the second stage, four solvents and gas-phase ESPs of small molecule monomers and tetrapeptides were calculated using the parameters obtained from the ESPs in a different solvent and compared them to the ab initio ESPs. This test examined cross-solvent transferability.

METHODS

Polarizability Tensor Data Sets. In this study, we utilized three data sets: (1) the JMW set, the molecules and dimers used in the development of an earlier polarizability set¹⁴ plus additional dimers with noble gas atoms; (2) the DES set, comprising 418 small molecules, including single-atom species, and their 3867 dimers,^{26,27} and (3) the QM7b set, 7211 small molecules of up to seven heavy atoms. The initial coordinates of the molecules in the DES data set were sourced from Shaw and co-workers.²⁶ Geometries of molecules and dimers in both JMW and DES data sets were optimized at the MP2/6-311++G(d, p) level of theory for consistency, and the polarizability tensors were calculated at the ω B97X-D/aug-cc-pVTZ level.²⁸ The QM7b polarizability tensors and coordinates were sourced from the data set of Yang et al. and Wilkins et al.,²⁹ and their molecular polarizabilities were calculated at the LR-CCSD/d-aug-cc-pVDZ level.

In this work, the polarizability tensors of JMW and DES monomer sets were calculated by five methods, including B3LYP/aug-cc-pVDZ (model 1), B3LYP/aug-cc-pVTZ (model 2), MP2/aug-cc-pVDZ (model 3), MP2/aug-cc-pVTZ (model 4), and ω B97X-D/aug-cc-pVTZ (model 5). Among them, only ω B97X-D/aug-cc-pVTZ (model 5) polarizability tensors were used for the development, and the polarizability tensors from all other four methods (models 1–4) were included for comparisons. In addition, we chose to include the QM7b polarizability tensors at the LR-CCSD/d-aug-cc-pVDZ level in the development rather than performing calculations at the ω B97X-D/aug-cc-pVTZ level. We compared a subset of randomly selected molecules between these two levels and found that the agreement was excellent.

Electrostatic Potential Data Sets. The ESPs were calculated for the JMW and DES sets. All ESPs were calculated in four solvents: diethyl ether (ETH, $\epsilon = 4.24$), dichloroethane (DCE, $\epsilon = 10.13$), acetone (ACT, $\epsilon = 20.49$), and water (WAT,

$\epsilon = 78.36$). To account for solvent polarization effects, we employed the PCM continuum solvent.²⁴ The surface was generated using Lebedev–Laikov grids, applying SMD–Coulomb atomic radii developed by Truhlar and co-workers.³⁰ The grids were on the molecular surface with a density of approximately 5 points/Å² and further smoothed using the York–Karplus algorithm.³¹ The surface polarization charges were represented as spherical Gaussians. For comparison with our earlier work, we also used the ESP set of DES monomers that we used in our previous work²⁰ which was calculated at the MP2/aug-cc-pVTZ//MP2/6-311++G(d, p) level with PCM in the above-mentioned four solvents and in the gas phase.

In addition, we also used the ESPs of the tetrapeptides that has been reported earlier.²⁰ Here, we briefly summarize the key details. Each tetrapeptide was terminated by the N-acetyl (ACE) and N-methylamide (NME) groups and was modeled in five distinct conformations: antiparallel β -sheet ($a\beta$), right-handed α -helix (αR), left-handed α -helix (αL), β -sheet (β), and polyproline type II (pII) conformations. The initial coordinates of the 100 tetrapeptide conformers were obtained from the work of Jiang et al.³² The coordinates were subsequently optimized at the MP2/6-311++G(d, p) level of theory, with fixed mainchain torsional angles (ϕ , ψ) set to $(-140^\circ, 135^\circ)$, $(-57^\circ, -47^\circ)$, $(57^\circ, 47^\circ)$, $(-119^\circ, 113^\circ)$, and $(-79^\circ, 150^\circ)$, corresponding to the five conformations.

Except for the DES monomers, all ESPs were calculated at the ω B97X-D/aug-cc-pVTZ level of theory with solvent polarization effects being represented by PCM. We used two ESP sets for DES monomers in this study at different stages of the present work: one set at the MP2/aug-cc-pVTZ level of theory that was used in our earlier work and the other set at the ω B97X-D/aug-cc-pVTZ level of theory. The MP2 ESPs were used only for cross-solvent transferability assessments, whereas the ω B97X-D ESPs were used in the polarizability fitting as well as monomer–dimer transferability assessments. In all ESP calculations, grid points were generated based on the method developed by Singh et al.³³ These grid points were located at distances of 1.4, 1.6, 1.8, and 2.0 times the van der Waals radii, with a grid density of 6 points per Å². All QM calculations were performed using the Gaussian 16 package.³⁴

Parameter Development and Test. The main departure of this refinement from the traditional approach is the utilization of both molecular polarizability tensors and solution-phase ESPs in the fitting process. This has been made possible by the recently developed PCMRESP method²⁰ that facilitates induced dipole polarizable force field parameterization in multiple solvents. Therefore, in our fitting procedure, we fit both the molecular polarizability tensors and the in-solution ESPs simultaneously. The molecular polarizability tensors were calculated using eq 11. They were iteratively fit to 13,523 ab initio polarizability tensors. The atomic polarizabilities were applied to fit the 1565×4 monomer ESPs in four solvents. Numerical derivatives were calculated by varying the atomic polarizabilities (and radii or screening constant, if applicable) by a small amount. A steepest descent procedure was performed to iteratively search the minimal weighted-sum-of-square error (WSSE) of both ESPs and molecular polarizability tensors.

$$\text{WSSE} = w_p \sum_{i=1}^{N_{\text{pol}}} \left(\frac{\alpha_i^{\text{calc}} - \alpha_i^{\text{QM}}}{\|\alpha_i^{\text{QM}}\|} \right)^2 + w_{\text{ESP}} \sum_{j=1}^{N_{\text{ESP}}} \sum_{l=1}^{N_{\text{solv}}} \left(\frac{\text{ESP}_{j,l}^{\text{calc}} - \text{ESP}_{j,l}^{\text{QM}}}{\|\text{ESP}_{j,l}^{\text{QM}}\|} \right)^2 \quad (12)$$

where WSSE is the weighted sum of square errors, α_i^{calc} and α_i^{QM} are the pGM and ab initio polarizability tensors, respectively, of the i th molecule, $\text{ESP}_{j,l}^{\text{calc}}$ and $\text{ESP}_{j,l}^{\text{QM}}$ are the pGM and ab initio ESPs, respectively, of the j th molecule in the l th solvent, $\|\alpha_i^{\text{QM}}\|$ is the isotropic molecular polarizability, $\|\text{ESP}_{j,l}^{\text{QM}}\|$ is the root-mean-square ESP, and w_p and w_{ESP} are two adjustable weighting factors. In our case, $w_p = 1$ and $w_{\text{ESP}} = 1$ were chosen.

Tests were then performed on the atomic polarizabilities by comparing the ESPs against those calculated quantum mechanically using three ESP data sets: (1) the 4785×4 DES and amino acid analogue dimer set for the monomer to dimer transferability test; (2) the 403×5 DES monomer set for the cross-solvent transferability test; and (3) the $20 \times 5 \times 5$ TET-pep set for the cross-solvent transferability test. Two sets of ESP parameters were developed for the DES monomer data sets, including atomic monopoles and permanent dipoles, and were applied to calculate either the dimer ESPs or ESPs in different solvents. The parameters that were applied to evaluate the monomer to dimer transferability were developed by combining ESP DES monomers in four solvents (ETH, DCE, ACT, and WAT). The DES monomer parameters were then applied to calculate the ESPs of the dimers. For the parameters used to evaluate the cross-solvent transferability, only the ESPs in a chosen solvent were used in parameterization.

The charges and permanent dipoles were obtained by fitting to the in-solution ESPs using the PCMRSP method.²⁰ The surface charges, coordinates, and weighting factors were taken directly from Gaussian 16³⁴ outputs. Except those of the TET-pep data set, all ESP parameters were developed using a two-stage fitting procedure that involved iteratively fitting the ESPs, as extensively detailed in our prior publications.^{16,18} In the first stage, the initial monopoles were set to zero, and for the pGM-perm model, the initial permanent dipoles were also set to zero. During this stage, chemically equivalent atoms, except those in the $-\text{CH}_2-$ and $-\text{CH}_3$ groups, were constrained to have identical parameters. In the second stage, only the $-\text{CH}_2-$ and $-\text{CH}_3$ groups underwent fitting with appropriate chemical equivalencing applied, and all other parameters, including monopoles and permanent dipoles, retained values obtained from the first stage of fitting. The surface charges and ESPs for multisolvent fittings were obtained by combining those of the same molecule. The parameters (charges and dipoles) in different solvents were equivalenced.

For the TET-pep data set that was used in cross-solvent transferability tests, parameters were developed for each peptide by combining the ESPs in the designated solvent of all five conformations in a single-stage procedure. In this process, chemical equivalence was enforced for all atoms, except the methyl groups of the terminal residues. The chemical equivalence in the fitting process is expected to lead to some degree of deterioration in the fitting quality because of the reduced number of degrees of freedom. However, because many of these groups can rotate freely, chemical equivalence effectively accounts for the average effects.

The primary objective of our transferability test was to assess to which extent the electrostatic parameters obtained in one medium or monomeric state could be applied to other media or a dimeric state. We selected the gas phase, diethyl ether, acetone, dichloroethane, and water as the test media, encompassing a range of dielectric constants from 1.0 to 78.36. All the solution media were implicitly described using PCM as implemented in Gaussian 16³⁴ software.

As measures of errors, we calculated the root-mean-square error (RMSE), root-mean-square difference (RMSD), and relative RMSE (RRMSE) between calculated and ab initio ESPs as well as between calculated and ab initio molecular polarizabilities. We also calculated the mean-absolute-percent-difference (MAPD).

RESULTS AND DISCUSSION

Comparing Ab Initio Polarizabilities to Experimental Measurements. In addition to the polarizabilities reported earlier that were calculated using four different methods,^{2,14} we have now calculated the polarizability tensors using the $\omega\text{B97X-D/aug-cc-pVTZ}$ method for a diverse set of molecules. All polarizability calculations were performed in the gas phase. Therefore, the polarizabilities are what are commonly referred to as “static polarizabilities”. Among them, a subset of approximately 400 molecules were calculated by all five methods (B3LYP/aug-cc-pVDZ, B3LYP/aug-cc-pVTZ, MP2/aug-cc-pVDZ, MP2/aug-cc-pVTZ, and $\omega\text{B97X-D/aug-cc-pVTZ}$) and were experimentally measured. The remaining approximately 300 molecules with experimental data were too large, and calculation of the polarizability tensors at the MP2/aug-cc-pVTZ level was too expensive. Therefore, calculations on these 300 molecules were performed by using the DFT methods only. For unbiased comparison of the methods, data in Tables 1 and 2 and Figure 1 were collected from the set of 400 molecules that were calculated by all five methods.

Table 1. Comparison between Calculated and Experimental²¹ Isotropic Polarizabilities of the Subset of Molecules That Were Calculated by All Five Methods and Have Experimental Data

	model 1 ^a	model 2 ^b	model 3 ^c	model 4 ^d	model 5 ^e
slope	1.0176	1.0051	0.9854	0.9896	0.9909
R ²	0.9995	0.9995	0.9994	0.9994	0.9994
RMSE	0.34	0.27	0.33	0.31	0.30
MAPD	1.98%	1.59%	2.52%	2.19%	2.06%

^aModel 1: B3LYP/aug-cc-pVDZ. ^bModel 2: B3LYP/aug-cc-pVTZ. ^cModel 3: MP2/aug-cc-pVDZ. ^dModel 4: MP2/aug-cc-pVTZ. ^eModel 5: $\omega\text{B97X-D/aug-cc-pVTZ}$. Models 1–4 were obtained from ref 14.

Table 2. RMSD (in Å³) between the Isotropic Polarizabilities Calculated by the Models^a

	model 1	model 2	model 3	model 4
model 2	0.154			
model 3	0.418	0.299		
model 4	0.353	0.225	0.095	
model 5	0.329	0.194	0.150	0.080

^aSee Table 1 caption for definition of models.

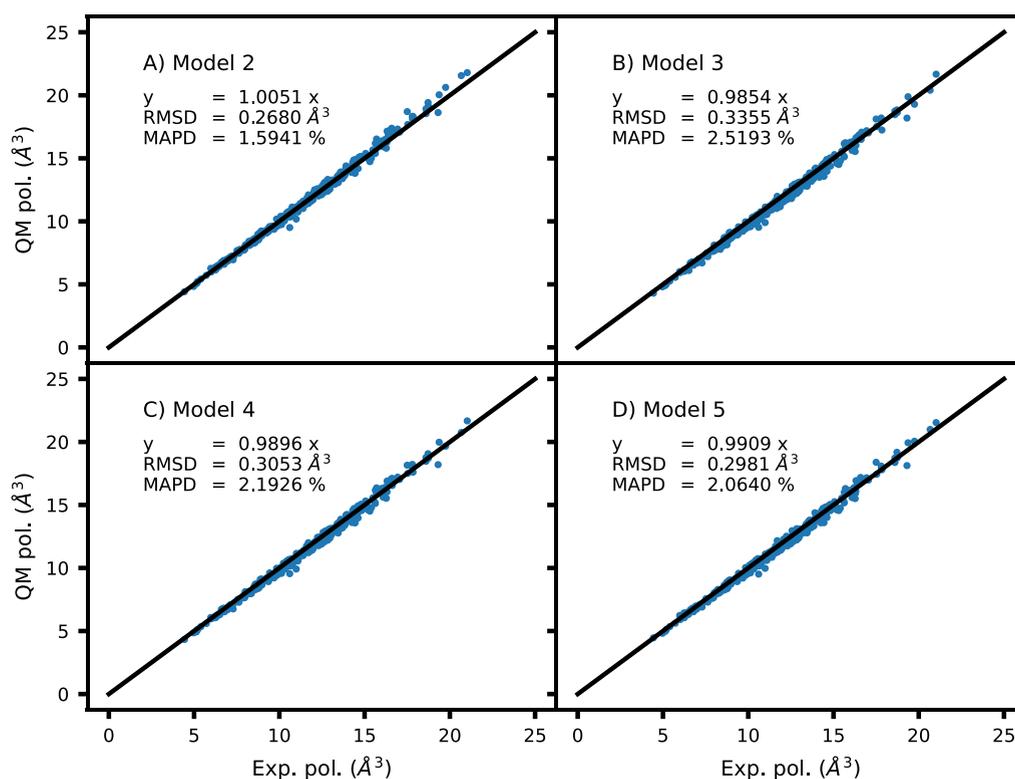


Figure 1. Comparison between experimental and calculated isotropic molecular polarizabilities. Experimental values were obtained from Bosque and Sales.²¹ Models 2–4 were obtained from our earlier work.¹⁴ For consistency, only those molecules that were calculated by all methods are shown in this comparison. Model labels are explained in Table 1 caption.

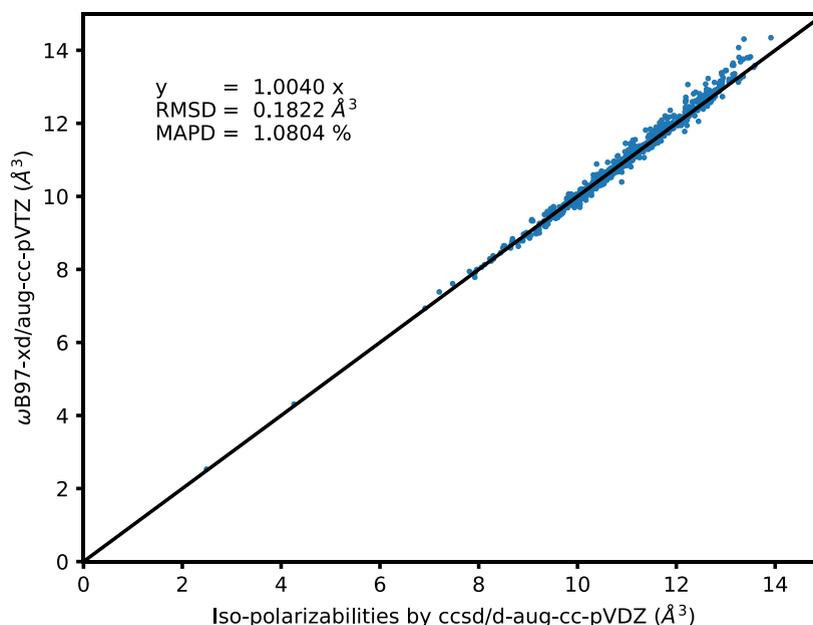


Figure 2. Comparison between the isotropic polarizabilities calculated by ω B97X-D/aug-cc-pVTZ and LR-CCSD/d-aug-cc-pVDZ of the randomly selected 682 molecules from the QM7b database.

As shown in Table 1 and Figure 1, among the five methods, results of “model 2” by B3LYP/aug-cc-pVTZ were closest to the experimental measurements, in terms of both root-mean-square-error (RMSE = 0.268 Å³) and mean-absolute-percent-difference (MAPD = 1.59%), whereas the results of “model 1” by B3LYP/aug-cc-pVDZ had largest RMSE = 0.344 Å³ and those of “model 3” by MP2/aug-cc-pVDZ had the largest MAPD = 2.52%. In

both cases, the difference between smaller basis set and large counter parts suggests the basis set convergence issue existing in the calculations of polarizabilities.

It is also interesting to note that both B3LYP/aug-cc-pVTZ and ω B97X-D/aug-cc-pVTZ are closer to the experimental results than MP2/aug-cc-pVTZ. We speculate that basis set convergence might have played a role, given the notably smaller

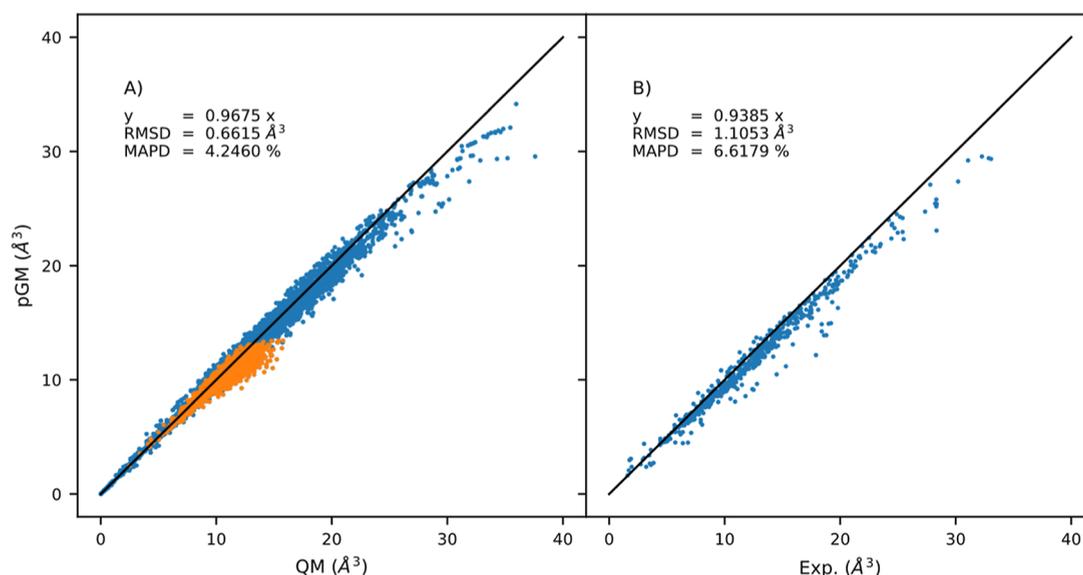


Figure 3. Comparison between the isotropic molecular polarizabilities calculated by pGM against (A) by ω B97X-D/aug-cc-pVTZ (blue) and LR-CCSD/d-aug-cc-pVDZ (light brown) and (B) experimental values.

improvements from MP2/aug-cc-pVDZ to MP2/aug-cc-pVTZ than from B3LYP/aug-cc-pVDZ to B3LYP/aug-cc-pVTZ. Another interesting observation is that B3LYP/aug-cc-pVTZ was notably closer to experimental results than ω B97X-D/aug-cc-pVTZ. This was somewhat surprising given the considerable improvement of ω B97X-D over B3LYP for their ability in reproducing the MP2 and CCSD ESPs as we have shown earlier.¹⁸ Nevertheless, we opted to use ω B97X-D/aug-cc-pVTZ polarizability tensors in the development of the pGM force field, not the B3LYP/aug-cc-pVTZ data. Our choice was informed by the observations that (1) ω B97X-D was one of the two DFT methods that can accurately reproduce MP2 and CCSD ESPs,¹⁸ notably better than the B3LYP method, and (2) MP2 and CCSD polarizability calculations can become prohibitively expensive for large molecules.

Another interesting comparison is the consistency among the polarizability calculation methods. As a direct measurement of similarities, we calculated the root-mean-square-differences (RMSD). In Table 2, interestingly, the smallest RMSD = 0.080 \AA^3 was the one between models 4 and 5, namely, MP2/aug-cc-pVTZ and ω B97X-D/aug-cc-pVTZ. This was even smaller than the ones between MP2/aug-cc-pVTZ and MP2/aug-cc-pVDZ (RMSD = 0.095 \AA^3) and between B3LYP/aug-cc-pVTZ and B3LYP/aug-cc-pVDZ (RMSD = 0.154 \AA^3).

In this iteration, we also included the QM7b data set. The QM7b set comprises diverse small molecules with up to seven heavy atoms. In this database, polarizabilities were calculated by B3LYP/d-aug-cc-pVDZ, SCAN0/d-aug-cc-pVDZ, and LR-CCSD/d-aug-cc-pVDZ methods. Because of the diversity, QM7b has been used widely in the development of models using machine learning techniques.

To compare the LR-CCSD/d-aug-cc-pVDZ method against the ω B97X-D/aug-cc-pVTZ method, we randomly selected a set of 682 molecules from the total of 7211 molecules in the QM7b set. The polarizabilities of these 682 molecules were calculated by the ω B97X-D/aug-cc-pVTZ method using the geometry presented in the QM7b database.

Shown in Figure 2 is the scatter plot of the isotropic static molecular polarizabilities by both the LR-CCSD/d-aug-cc-pVDZ and ω B97X-D/aug-cc-pVTZ methods. Overall, the two

polarizability sets are in excellent agreement with a slope of 1.004 and $R^2 = 0.9997$, with RMSD = 0.1822 \AA^3 and MAPD = 1.08%. It is also interesting to note that the polarizabilities covered in the QM7b set were mostly within the range between 7 and 14 \AA^3 , whereas the ranges covered in other sets used in this study were within 40 \AA^3 , significantly broader than the QM7b set.

Fitting Quality Assessment: pGM Model Compared to QM and Experimental Polarizabilities. Shown in Figures 3, S1, and S2 are scatter plots comparing the polarizabilities calculated by pGM and those calculated by the ω B97X-D/aug-cc-pVTZ and LR-CCSD/d-aug-cc-pVDZ methods. Here, the slope, RMSD, and MAPD in Figure 3 were obtained for the entire data set, whereas those in Figures S1 and S2 were obtained for their respective data sets. Relative to QM and experimental values, the pGM model tends to underestimate the polarizabilities. The regression slopes were 0.9675 and 0.9385, respectively, for QM and experimental data, the RMSDs were 0.6615 and $1.1053 \text{ (\AA}^3)$, respectively, and the MAPDs were 4.25% and 6.62%, respectively.

Remarkably, these results are comparable to those of Litman et al.²² In Litman et al., the atomic polarizabilities were averages of two sets: one was from training with a set of 773 molecular polarizabilities, and the other was from training a set of 316 probing ESPs of 37 small molecules. Apparently, their MAPDs of 4.7% (training) and 4.4% (validating) and the present work 4.25% MAPDs were similar, and both were notably worse than the average 2.31% error of Bosque and Sales²¹ who utilized a rather simple linear additive fitting without consideration of intramolecular polarization. We attribute these observations to the inclusion of ESPs in Litman et al. and our work and the lack of consideration of ESPs in Bosque and Sales. Furthermore, these observations suggest that molecular polarizabilities alone may not be sufficient to determine atomic polarizabilities for an accurate representation of molecular ESPs.

It is important to note that in this work, we combine both molecular polarizability tensors and the ESPs in the development. Therefore, we anticipate that our fitting quality, as measured by the polarizability difference from both QM and experimental data, would be somewhat large compared to those

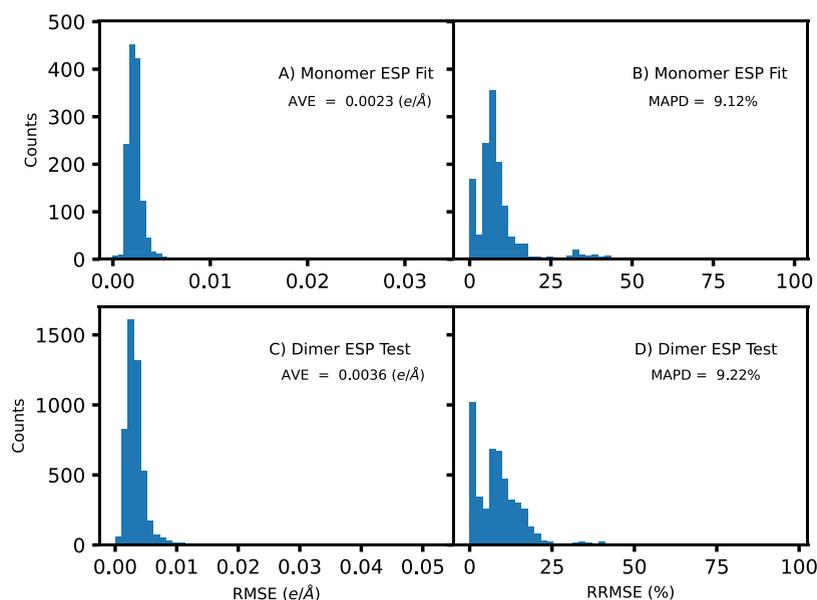


Figure 4. Histograms of RMSE and RRMSE values of the ESP of monomers (A,B) and dimers (C,D) by pGM. The dimer ESPs were calculated using parameters derived from monomer ESPs.

that were fit solely to the polarizability tensors. In addition, in this work, we limit our independent fitting variables to atomic polarizabilities only and derive the atomic radii by applying a formula developed by Elking et al.¹² This significantly reduces the number of independent variables and degrees of freedom of fitting that naturally lead to numerically less accurate fit.

Since polarization plays the pivotal role in facilitating transferability across multiple oligomeric states, multiple conformations, and multiple dielectric environments, we will focus on the transferability of parameters in this work. Specifically, we will measure how well the parameters developed in the monomeric state and one dielectric environment can be transferred to the dimeric state and other dielectric environments.

Fitting Quality Assessment: Monomer ESPs in Multiple Solvents. Figure 4A,B illustrates the fitting quality of the combined multisolvant ESPs as measured by the histograms of the RMSE (in e/Å) and relative RMSE (in percent) of the calculated monomer ESPs. The data set in this fitting includes the monomers from the DES data set, the polarizability data set we use earlier, as well as amino acid side chain and nucleotide analogues, for a total of 1345 compounds. The ESPs in four solvents (ETH, DCE, ACT, and WAT) were calculated for all molecules, for a combined total of 1345×4 ESP sets. A two-stage fitting was performed by combining the four ESPs for each molecule. Among the 1345 compounds, 506 (37%) had RMSE less than 0.002 e/Å and only 20 (or 1.5%) had RMSE greater than 0.005 e/Å. In terms of relative RMSE, 322 of the 1345 compounds (24%) were smaller than 5%, 1024 (76%) were smaller than 10%, and 1251 (93%) were smaller than 20%. Only 9 out of 1345 (or 0.7%) were greater than 50%. Overall, the average RMSE was 0.0023 e/Å, and the average RRMSE was 9.12%. Thus, judging by the small RMSE and RRMSE overall, fitting of the ESPs was considered to be good quality.

Detailed information about the 9 outliers is presented in Table S4. Among them, 3 are cycloalkanes. These molecules have a weak electrostatic field. Their root-mean-square potentials were about 0.004 e/Å, compared to the average of 0.060 e/Å of all 1345 monomers. Their RMSE's were about

0.0019 e/Å, which is slightly smaller than the average RMSE of 1345 monomers (0.0023 e/Å). Thus, the elevated RRMSEs of these 3 cycloalkanes were due to their weak electrostatic field. Among the remaining 6 monomers, 4 are diatomic molecules (including 3 homonuclear) and 4 are halogen molecules. The observation suggests that these molecules have unique chemistry that may require separate atom-type assignments or more elaborate models.

ESP Transferability from Monomer to Dimer in Four Solvents. In this test, we apply the parameters developed from fitting the monomer ESPs to calculate the ESPs of the dimers in four solvents. There were 4743 dimers used in this assessment. Among them, 3827 were from the DES data set, 888 were amino acid analogue dimers (including water and noble gas to amino acid analogue dimers), and 28 were nucleotide base pairs. All ESPs were calculated in four solvents (ETH, DCE, ACT, and WAT) for a total of 4743×4 ESPs in this test set.

The histograms of RMSE and RRMSE values of the calculated dimer ESPs (Figure 4C,D) show that the parameters developed using multisolvant monomer ESPs can be readily transferred to calculate the dimer ESPs in multiple solvents. The dimer histograms (Figure 4C,D) were similar to those of the monomer histograms (Figure 4A,B). Among the 4743 dimer RMSEs, 821 (17.2%) were smaller than 0.002 e/Å, and only 92 (1.9%) were greater than 0.01 e/Å. Among the RRMSEs, 2996 (63.2%) were smaller than 10%, 4461 (94.1%) were smaller than 20%, and 20 (0.4%) were greater than 50%. Overall, the average RMSE was 0.0036 e/Å, and the average RRMSE was 9.22%. The maximum RMSE was 0.0515 e/Å. Compared to those from the monomer fitting, the average RMSE of the dimers was notably larger than the average RMSE 0.0023 e/Å of the monomers, and the average RRMSE was similar to the average of the monomers (9.12%). Thus, the increased average RMSE of the dimer ESPs was mostly due to the fact that dimer ESPs are overall larger than monomer ESPs. Given the high degree of similarities between the average RRMSEs, we conclude that the ESP parameters of the monomers are highly transferable to dimers, and the polarizability parameters are applicable to both monomers and dimers.

ESP Parameter Transferability across Multiple Solvents. A key impetus of polarizable force field development is to enable simulations across diverse dielectric environments. Therefore, the ability to model ESPs in multiple solvents is considered an important test. Here, we assess the transferability using two data sets that were used in our earlier work.²⁰ The first set of ESPs were those of monomers from the DES data set, and the second set of ESPs were those of tetrapeptides. Both sets were calculated in four solvents and gas phase. Table 3 illustrates

Table 3. Average Relative Root-Mean-Square-Fitting Errors of the pGM Model in Four Solvents and Gas-Phase Conditions Using DES Monomer and Tetrapeptide Data Sets

	GAS	ETH	DCE	ACT	WAT	AVE
	DES/RRMSE (%)					
RESP	24.70	21.52	20.60	20.33	19.52	21.33
pGM-ind	18.68	16.73	16.12	15.93	15.35	16.56
pGM-perm	12.37	11.03	10.60	10.48	10.04	10.90
	TET-peptide/RRMSE (%)					
RESP	10.24	8.45	7.78	7.70	6.73	8.18
pGM-ind	6.50	6.00	5.55	5.36	4.71	5.62
pGM-perm	4.45	3.94	3.68	3.60	3.21	3.78

the performance of the pGM models with the polarizability set. Here, pGM-ind and pGM-perm are two variants of the pGM models; both pGM-ind and pGM-perm share identical formalism, except pGM-ind does not contain permanent dipoles, whereas pGM-perm does. In our earlier work,²⁰ in which the atomic radii were treated as independent adjustable variables, the average fitting RRMSEs of the pGM-ind were 19.5% and

7.6%, respectively, for DES monomer and TET-pep in four solvents plus gas-phase ESPs (Figure 1 of ref 20). In this work, even though all atomic radii were assigned using the formula developed by Elking et al.,¹² the average fitting RRMSEs of pGM-ind were reduced to 16.6% and 5.6% for DES and TET-pep sets, respectively. We conclude that the updated parameter set improves the fitting quality of the pGM-ind model. On the other hand, the average RRMSEs of the pGM-perm model were somewhat mixed. Improvement, albeit small, was observed in the DES set (from previous 11.7% to current 10.9%), whereas for the TET-pep set, the fitting quality was comparable to earlier work (3.7%, previous vs 3.8%, current).

Figures 5 and 6 show the heat maps of the cross-media transferability of DES and TET-pep sets, respectively, as measured by the RRMSE of the calculated ESPs using parameters developed under another dielectric condition. Once again, for the DES set, the updated polarizability set exhibits improvements over the set we reported earlier.²⁰ About 2% improvement was observed for the pGM-ind method, whereas pGM-perm was also improved by less than 1%.

A similar trend was observed for the TET-pep set. In this case, the pGM-ind model exhibits an average of 1.6% improvement (reduced from 7.9% of our earlier work²⁰ to 6.3% of present result), and pGM-perm had a comparable RRMSE (from 4.5% in an earlier work²⁰ to 4.7% in this work). However, if we exclude all entries that involve gas-phase data, only focus on the other four media, as seen in the col under the title “AVE-gas” (Tables S1 and S2), pGM-ind improved from 7.5% to 5.6%, and pGM-perm remained essentially unchanged (4.1%–4.0%). It is noteworthy that the polarizabilities were obtained using the pGM-perm method. Here, we see notable improvements with

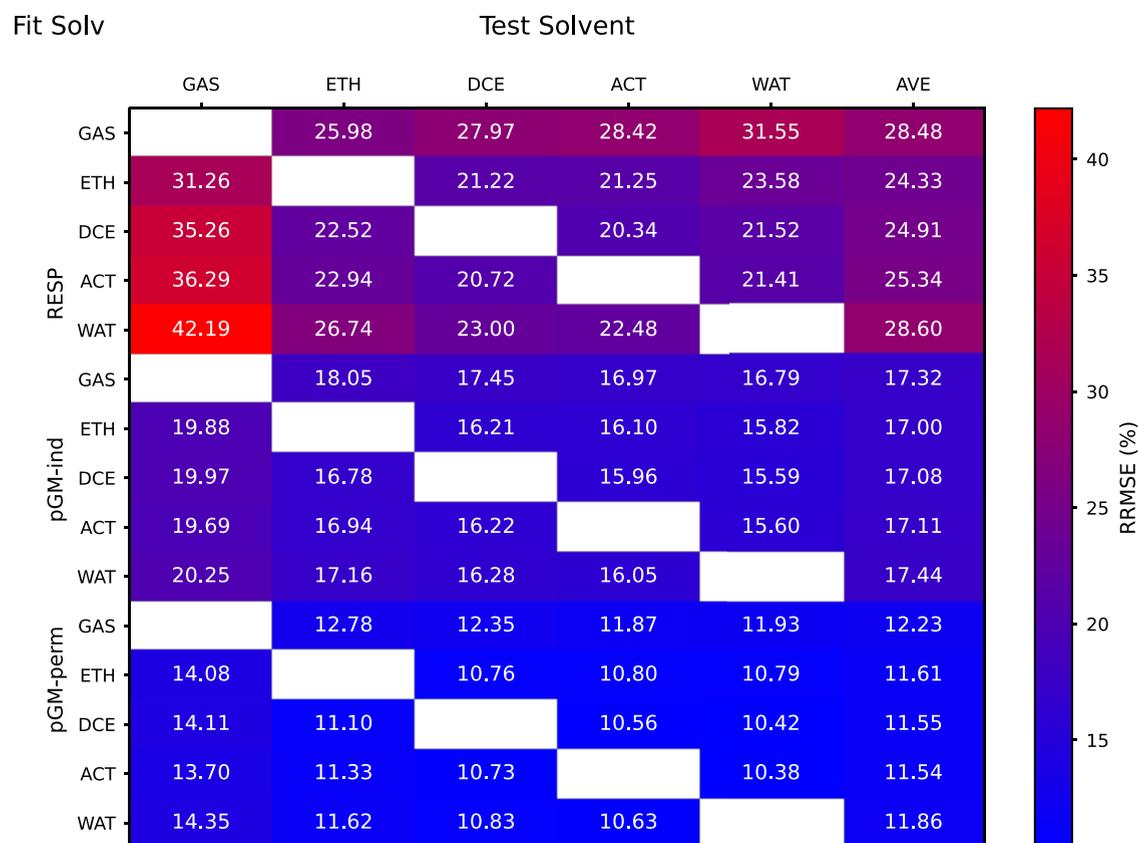


Figure 5. Transferability for the DES monomer data set. Column titles indicate the testing solvents. Fitting methods and solvents are labeled on the left.

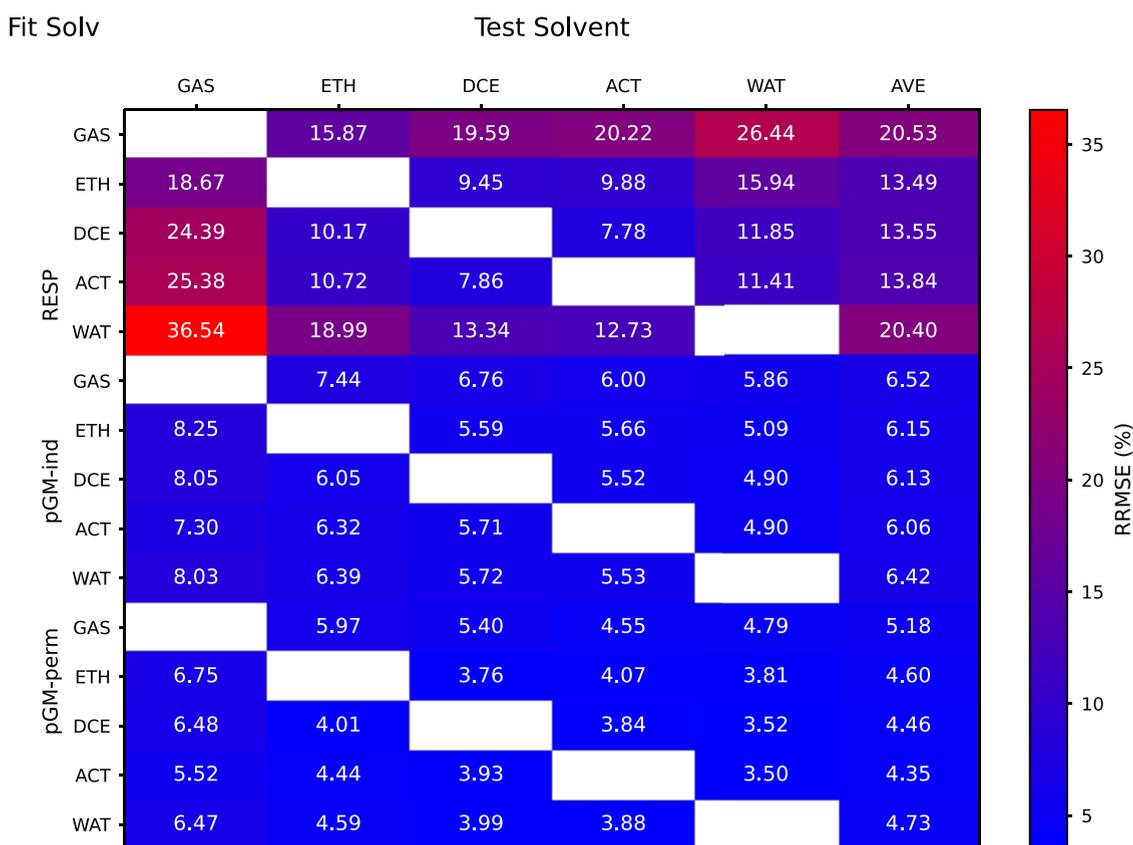


Figure 6. Transferability test for Tetrapeptide set. Column titles indicate the testing solvents. Fitting methods and solvents are labeled on the left.

pGM-ind in both DES and TET-pep data sets, even though the numbers of degrees of freedom in ESP fitting were identical to those used in the RESP method. Therefore, the improvements observed in pGM-ind suggest that the pGM-ind model is significantly more transferable than the nonpolarizable point charge models. Furthermore, because the difference between RESP and pGM-ind is the inclusion of the induced dipoles, the significant improvements over the RESP method strongly suggests that induced dipole polarization significantly enhances transferability.

On the other hand, the pGM-perm method performs consistently better than both RESP and pGM-ind. Since pGM-perm differs from pGM-ind by the presence of the permanent atomic dipoles, it is not surprising that fittings using pGM-perm were systematically better than those using pGM-ind due to additional parameters. However, we also see systematic improvements in transferability when the parameters were used to calculate the ESP in another solvent. This indicates that the addition of permanent dipoles enhances the model quality and reliability. This suggests strongly that, in addition to the significantly improved transferability due to induced dipoles, pGM-perm with monopoles, induced and permanent dipoles, is highly transferable and reliable.

Remarkably, after excluding the gas-phase data, the average RRMSEs were 16.03% and 10.54% (Table 3), for fittings using the pGM-ind and pGM-perm methods, respectively. These were just slightly better than those average RRMSEs, 16.25% and 10.83% (Table S1), of the transfer ESPs, for pGM-ind and pGM-perm, respectively. Similar observations can be made with respect to the TET-pep set. With exclusion of the gas-phase data, the average RRMSEs were 5.41% and 3.61% (Table 3), for

fittings using the pGM-ind and pGM-perm methods, respectively. These were just 0.21% and 0.34% better than the average RRMSEs, 5.62% and 3.95% (Table S2), of the transfer ESPs, for pGM-ind and pGM-perm, respectively. Thus, we conclude that both pGM-ind and pGM-perm with the refined polarizabilities are highly transferable across multiple solvents for both the small molecule DES monomer and the TET-pep sets.

Atomic Radii in pGM Can Be Derived Directly from Polarizabilities. In the work of Elking et al.,¹² the Gaussian exponent, β , is given by

$$\beta = a / \left(\alpha \frac{2}{3\sqrt{2\pi}} \right)^{1/3} \quad (13)$$

where $a < 1$ is a scaling constant, and α is the isotropic polarizability. In pGM

$$\beta = 1 / (\sqrt{2} r) \quad (14)$$

For simplicity, the monopole and all multipoles of each atom share identical r and β . The Gaussian radii can be calculated as

$$r = \left(\alpha \frac{2}{3\sqrt{2\pi}} \right)^{1/3} / (\sqrt{2} a) \quad (15)$$

Importantly, constant a affects the rate of convergence¹² of the induced dipole calculations which are typically calculated iteratively. When a is too large $a > 1$, the induced dipole iteration may not converge,¹² leading to the so-called “polarization catastrophe”.¹⁰ For this reason, $a < 1$ and r must satisfy¹²

$$r > \left(\alpha \frac{2}{3\sqrt{2\pi}} \right)^{1/3} / \sqrt{2} \quad (16)$$

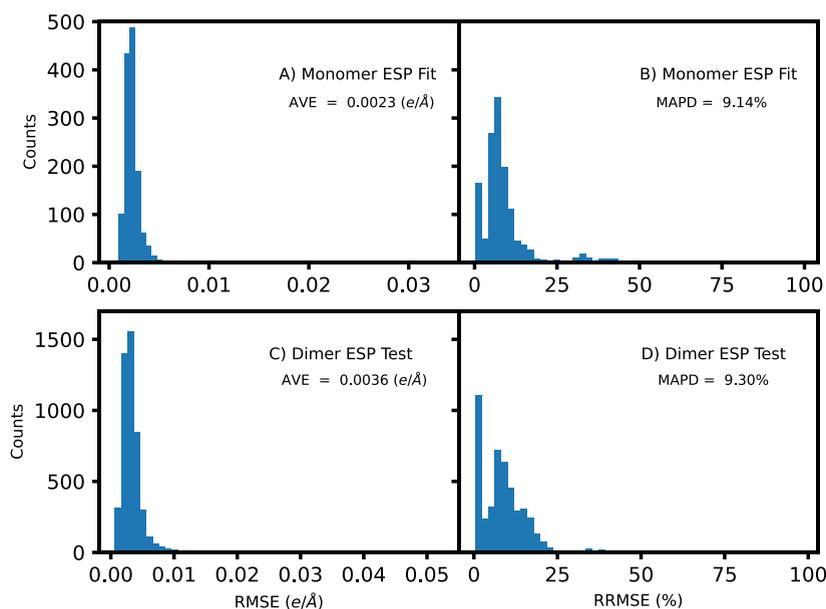


Figure 7. Histograms of RMSE and RRMSE values of the (A,B) monomer and (C,D) dimer ESPs in pGM with variable radii.

Furthermore, when the same constant a is used for all atoms, we call such a scheme the “universal screening factor” (USF). Alternatively, a can vary for different atoms, leading to independent Gaussian radii, and the scheme is called the “variable screening factor” (VSF). Although the two types of polarizability sets may share identical polarization schemes, they differ significantly by the number of independent fitting variables. In the USF type, the atomic radii are not independent fitting parameters, whereas they are in the VSF. Therefore, the number of independent fitting parameters in the VSF scheme is essentially twice as many as that in the USF scheme. Naturally, due to the significant increase of independent fitting parameters, the VSF types are expected to have smaller numerical fitting errors. However, such decreased errors may or may not be a true reflection of the underlying physical property and in fact can be a result of overfitting. In our case, to assess whether or not such increase of independent fitting parameters constitutes overfitting, we compared the performances of the two parameter sets.

The reported results in this work so far are those from the USF scheme in which $a = 0.45069523$ is a constant for all atoms (Table S3). The atomic radii in Table S3 were calculated directly using eq 15. It is noteworthy that the value of a is about half of that reported by Elking et al.¹² Because a affects the rate of convergence of induced dipole calculations, we anticipate that our parameter set may have better convergence rate than Elking et al.¹²

We also evaluated the VSF scheme. We started from the USF atomic polarizabilities and radii in Table S3 and further optimized both the polarizabilities and radii by minimizing the WSSE in eq 12. Shown in Figure S2 is the scatter plot of QM (x -axis) versus pGM molecular isotropic polarizabilities (y -axis). The data set in Figure S2 was identical to those used in Figure 3A which includes both ω B97X-D/aug-cc-pVTZ and LR-CCSD/d-aug-cc-pVDZ polarizabilities. Overall, the isotropic polarizabilities of the VSF scheme had an RMSD = 0.651 (\AA^3) and MAPD = 4.17%. Compared to RMSD = 0.6615 (\AA^3) and MAPD = 4.25% from the USF set, the VSF set shows slight reduction of the fitting error by less than 2% of the RMSD. This trend holds true also for the ESPs. For monomer ESPs in multiple solvents, the average relative RMSE was 9.14%, which was only slightly

better than the 9.12% observed in the USF fitting. Figure 7A,B shows the histograms of RMSE and RRMSE of VSF from the monomer fitting, and Figure 7C,D shows the dimer testing results, respectively. Here, the high degree of similarity to the USF results in Figure 4 is apparent. For the dimer tests, the average relative RMSE of VSF was 9.30% slightly inferior to 9.22% of USF. Because of these observed small changes and the notably increased number of independent fitting parameters in VSF, we conclude that the USF set is sufficiently accurate, and that addition of atomic radii as independent variables did not significantly improve the fitting quality or transferability, that atomic radii in pGM can be derived directly from the polarizabilities, and that the large increase of number of independent fitting variables in the VSF scheme is unnecessary.

CONCLUSIONS

Atomic polarizabilities are considered fundamental parameters in polarizable force fields that affect the accuracy and transferability. In this work, we refined the atomic polarizabilities using both a large set of molecular polarizability tensors and multisolvant ESPs. The resulting polarizabilities were examined by comparing the multisolvant dimer ESPs that were calculated using parameters obtained from monomer ESPs. Further examinations were made for cross-solvent transferability of small molecules and tetrapeptides. In all these examinations, both the pGM-ind and pGM-perm methods with the updated polarizabilities (Table S3) were found to be highly transferable from monomers to dimers and from one solvent to another. Remarkably, the average transfer RRMSEs in all cases were comparable to the average fitting RRMSEs after excluding the gas-phase data.

Electron polarization can be broadly categorized into two types. In addition to local induction, which is characterized by localized small scale movement of electrons, electrons can also undergo large scale and delocalized movement. The induced dipole polarizable models, such as pGM, are more suitable to treat local induction, which is the focus of the present work. To treat the large-scale delocalized induction, however, additional terms such as charge transfer can be introduced.

■ ASSOCIATED CONTENT

Data Availability Statement

All polarizability tensors, electrostatic potentials, and surface charges have been deposited to zenodo.org (DOI: 10.5281/zenodo.13906160).

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.4c02175>.

Comparison of the isotropic molecular polarizabilities calculated by wB97X-D/aug-cc-pVTZ and CCSD/d-aug-cc-pVDZ; transferability data of DES monomers and tetrapeptides; and atomic polarizabilities and radii for the pGM force field (PDF)

■ AUTHOR INFORMATION

Corresponding Author

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

Authors

Junmei Wang – Department of Pharmaceutical Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States; orcid.org/0000-0002-9607-8229

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

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

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jcim.4c02175>

Author Contributions

YD initiated, formulated, and performed most of the computation, fitting, and analysis, and wrote and edited the manuscript. JMW constructed most of the initial structures and performed some computation. PC and RL performed some computation and edited the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge the research support from NIH (GM79383 to Y.D., GM130367 to R.L., and GM147673 to J.W.). The authors acknowledge no conflicts of interest.

■ REFERENCES

(1) (a) Cieplak, P.; Caldwell, J.; Kollman, P. A. 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* (10), 1048–1057. (b) Cieplak, P.; Dupradeau, F. Y.; Duan, Y.; Wang, J. M. Polarization effects in molecular mechanical force fields. *J. Phys.: Condens. Matter* **2009**, *21* (33), 333102. (c) 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. (d) Wang, J.; Cieplak, P.; Cai, Q.; Hsieh, M. J.; Wang, J. M.; 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* (28), 7999–8008. (e) Wang, J. M.; 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* (24), 7088–7101.

(2) 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.

(3) (a) Ren, P. Y.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003**, *107* (24), 5933–5947. (b) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., Jr.; et al. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114* (8), 2549–2564. (c) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046–4063.

(4) 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.

(5) (a) Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D.; 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* (2), 87–92. (b) 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* (1–3), 245–249. (c) 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* (11), 2873–2885.

(6) (a) Mortier, W. J.; Vangenechten, K.; Gasteiger, J. Electronegativity Equalization - Application and Parametrization. *J. Am. Chem. Soc.* **1985**, *107* (4), 829–835. (b) Mortier, W. J.; Ghosh, S. K.; Shankar, S. Electronegativity Equalization Method for the Calculation of Atomic Charges in Molecules. *J. Am. Chem. Soc.* **1986**, *108* (15), 4315–4320. (c) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; 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* (16), 1515–1531. (d) Friesner, R. A. Modeling Polarization in Proteins and Protein-ligand Complexes: Methods and Preliminary Results. *Adv. Protein Chem.* **2005**, *72*, 79–104. (e) Patel, S.; Brooks, C. L. CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004**, *25* (1), 1–15. (f) Patel, S.; Mackerell, A. D., Jr.; Brooks, C. L., III 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* (12), 1504–1514.

(7) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. ACKS2: Atom-condensed Kohn-Sham DFT approximated to second order. *J. Chem. Phys.* **2013**, *138* (7), 074108.

(8) (a) Islam, M. M.; Kolesov, G.; Verstraelen, T.; Kaxiras, E.; van Duin, A. C. T. eReaxFF: A Pseudoclassical Treatment of Explicit Electrons within Reactive Force Field Simulations. *J. Chem. Theory Comput.* **2016**, *12* (8), 3463–3472. (b) Cools-Ceuppens, M.; Dambre, J.; Verstraelen, T. Modeling Electronic Response Properties with an Explicit-Electron Machine Learning Potential. *J. Chem. Theory Comput.* **2022**, *18* (3), 1672–1691. (c) Leven, I.; Head-Gordon, T. C-GeM: Coarse-Grained Electron Model for Predicting the Electrostatic Potential in Molecules. *J. Phys. Chem. Lett.* **2019**, *10* (21), 6820–6826. (d) Guan, X. Y.; Leven, I.; Heidar-Zadeh, F.; Head-Gordon, T.

Protein C-GeM: A Coarse-Grained Electron Model for Fast and Accurate Protein Electrostatics Prediction. *J. Chem. Inf. Model.* **2021**, *61* (9), 4357–4369.

(9) Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y. X.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M.; et al. The ReaxFF reactive force-field: development, applications and future directions. *npj Comput. Mater.* **2016**, *2*, 15011.

(10) (a) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59* (3), 341–350. (b) van Duijnen, P.; Swart, M. Molecular and atomic polarizabilities: Thole's model revisited. *J. Phys. Chem. A* **1998**, *102*, 2399–2407.

(11) 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* (9), 2952.

(12) Elking, D.; Darden, T.; Woods, R. J. Gaussian induced dipole polarization model. *J. Comput. Chem.* **2007**, *28* (7), 1261–1274.

(13) (a) Elking, D. M.; Cisneros, G. A.; Piquemal, J. P.; Darden, T. A.; Pedersen, L. G. Gaussian Multipole Model (GMM). *J. Chem. Theory Comput.* **2010**, *6* (1), 190–202. (b) Elking, D. M.; Perera, L.; Duke, R.; Darden, T.; Pedersen, L. G. Atomic Forces for Geometry-Dependent Point Multipole and Gaussian Multipole Models. *J. Comput. Chem.* **2010**, *31* (15), 2702–2713.

(14) Wang, J. M.; 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* (2), 1146–1158.

(15) (a) Wei, H. X.; Qi, R. X.; Wang, J. M.; Cieplak, P.; Duan, Y.; Luo, R. Efficient formulation of polarizable Gaussian multipole electrostatics for biomolecular simulations. *J. Chem. Phys.* **2020**, *153* (11), 114116. (b) Wei, H. X.; Cieplak, P.; Duan, Y.; Luo, R. Y. Stress tensor and constant pressure simulation for polarizable Gaussian multipole model. *J. Chem. Phys.* **2022**, *156* (11), 114114. (c) Huang, Z.; Zhao, S. J.; Cieplak, P.; Duan, Y.; Luo, R.; Wei, H. X. Optimal Scheme to Achieve Energy Conservation in Induced Dipole Models. *J. Chem. Theory Comput.* **2023**, *19* (15), 5047–5057.

(16) Zhao, S. J.; Wei, H. X.; Cieplak, P.; Duan, Y.; Luo, R. PyRESP: A Program for Electrostatic Parameterizations of Additive and Induced Dipole Polarizable Force Fields. *J. Chem. Theory Comput.* **2022**, *18* (6), 3654–3670.

(17) Zhao, S. J.; Wei, H. X.; Cieplak, P.; Duan, Y.; Luo, R. Accurate Reproduction of Quantum Mechanical Many-Body Interactions in Peptide Main-Chain Hydrogen-Bonding Oligomers by the Polarizable Gaussian Multipole Model. *J. Chem. Theory Comput.* **2022**, *18* (10), 6172–6188.

(18) Zhu, Q.; Wu, Y. X.; Zhao, S. J.; Cieplak, P.; Duan, Y.; Luo, R. Streamlining and Optimizing Strategies of Electrostatic Parameterization. *J. Chem. Theory Comput.* **2023**, *19* (18), 6353–6365.

(19) (a) Zhao, S. J.; Cieplak, P.; Duan, Y.; Luo, R. Y. Transferability of the Electrostatic Parameters of the Polarizable Gaussian Multipole Model. *J. Chem. Theory Comput.* **2023**, *19*, 924. (b) Zhao, S.; Cieplak, P.; Duan, Y.; Luo, R. Assessment of Amino Acid Electrostatic Parameterizations of the Polarizable Gaussian Multipole Model. *J. Chem. Theory Comput.* **2024**, *20*, 2098–2110.

(20) Duan, Y.; Niu, T.; Wang, J.; Cieplak, P.; Luo, R. PCMRESP: A Method for Polarizable Force Field Parameter Development and Transferability of the Polarizable Gaussian Multipole Models Across Multiple Solvents. *J. Chem. Theory Comput.* **2024**, *20*, 2820–2829.

(21) Bosque, R.; Sales, J. Polarizabilities of solvents from the chemical composition. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (5), 1154–1163.

(22) Litman, J.; Liu, C.; Ren, P. Atomic Polarizabilities for Interactive Dipole Induction Models. *J. Chem. Inf. Model.* **2022**, *62*, 79–87.

(23) (a) Cancès, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.* **1997**, *107* (8), 3032–3041. (b) Scalmani, G.; Frisch, M. J. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.* **2010**, *132* (11), 15.

(24) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105* (8), 2999–3093.

(25) Montilla, M.; Luis, J. M.; Salvador, P. Origin-Independent Decomposition of the Static Polarizability. *J. Chem. Theory Comput.* **2021**, *17* (2), 1098–1105.

(26) (a) McGibbon, R. T.; Taube, A. G.; Donchev, A. G.; Siva, K.; Hernández, F.; Hargus, C.; Law, K. H.; Klepeis, J. L.; Shaw, D. E. Improving the accuracy of Møller-Plesset perturbation theory with neural networks. *J. Chem. Phys.* **2017**, *147* (16), 161725. (b) Donchev, A. G.; Taube, A. G.; Decolvenaere, E.; Hargus, C.; McGibbon, R. T.; Law, K. H.; Gregersen, B. A.; Li, J. L.; Palmo, K.; Siva, K.; et al. Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Sci. Data* **2021**, *8* (1), 55.

(27) Kesharwani, M.; Karton, A.; Sylvetsky, N.; Martin, J. The S66 Non-Covalent Interactions Benchmark Reconsidered Using Explicitly Correlated Methods Near the Basis Set Limit. *Aust. J. Chem.* **2018**, *71*, 238–248.

(28) (a) Chai, J. D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10* (44), 6615–6620. (b) Dunning, T. H. Gaussian-Basis Sets for Use in Correlated Molecular Calculations 0.1. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007–1023.

(29) (a) Wilkins, D.; Grisafi, A.; Yang, Y.; Lao, K.; DiStasio, R.; Ceriotti, M. Accurate molecular polarizabilities with coupled cluster theory and machine learning. In *Proceedings of the national academy of sciences of the united states of america*, 2019, *116*, 3401–3406. (b) Yang, Y.; Lao, K.; Wilkins, D.; Grisafi, A.; Ceriotti, M.; DiStasio, R. Quantum mechanical static dipole polarizabilities in the QM7b and AlphaML showcase databases. *Sci. Data* **2019**, *6*, 152.

(30) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113* (18), 6378–6396.

(31) York, D. M.; Karplus, M. A smooth solvation potential based on the conductor-like screening model. *J. Phys. Chem. A* **1999**, *103* (50), 11060–11079.

(32) Jiang, J. L.; Wu, Y. B.; Wang, Z. X.; Wu, C. Assessing the Performance of Popular Quantum Mechanics and Molecular Mechanics Methods and Revealing the Sequence-Dependent Energetic Features Using 100 Tetrapeptide Models. *J. Chem. Theory Comput.* **2010**, *6* (4), 1199–1209.

(33) (a) Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16* (5), 548–558. (b) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5* (2), 129–145.

(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian 16*, revision C.01, 2016.