Methodology of Parameterization of Molecular Mechanics Force Field From Quantum Chemistry Calculations using Guided Genetic Algorithm: A case study of methanol

Ying Li,\textsuperscript{a} Hui Li,\textsuperscript{b} Maria K. Y. Chan,\textsuperscript{c,d} Subramanian Sankaranarayanan,\textsuperscript{c} Benoît Roux\textsuperscript{b,d}

\textsuperscript{a}Leadership Computing Facility, Argonne National Laboratory, IL 60439, USA
\textsuperscript{b}Department of Biochemistry and Molecular Biophysics, University of Chicago, IL 60637, USA
\textsuperscript{c}Computational Institute, University of Chicago, IL 60637, USA
\textsuperscript{d}Center for Nanoscale Materials, Argonne National Laboratory, IL 60439, USA

\textbf{ABSTRACT}

In molecular dynamics (MD) simulation, force field determines the capability of an individual model in capturing physical and chemistry properties. The method for generating proper parameters of the force field form is the key component for computational research in chemistry, biochemistry, and condensed-phase physics. Our study showed that the feasibility to predict experimental condensed phase properties (i.e., density and heat of vaporization) of methanol through problem specific force field from only quantum chemistry information. To acquire the satisfying parameter sets of the force field, the genetic algorithm (GA) is the main optimization method. For electrostatic potential energy ($E_{ESP}$), we optimized both the electrostatic parameters of methanol using the GA method, which leads to low deviations of $E_{ESP}$ between the quantum mechanics (QM) calculations and the GA optimized parameters. We optimized the van der Waals (vdW) parameters both using GA and guided GA methods by calibrating interaction energy ($\Delta E$) of various methanol homoclusters, such as nonamers, undecamers, or tridecamers. Excellent agreement between the training dataset from QM calculations (i.e., MP2) and GA optimized parameters can be achieved. However, only the guided GA method, which eliminates the overestimation of interaction energy from MP2 calculations in the optimization process, provides proper vdw parameters for MD simulation to get the condensed phase properties (i.e., density and heat of vaporization) of methanol. Throughout the whole optimization process, the experimental value were not involved in the objective functions, but were only used for the purpose of justifying models (i.e., nonamers, undecamers, or
tridecamers) and validating methods (i.e., GA or guided GA). Our method is general and can be extended to any type of small organic molecule systems, as well as other descriptive polarizable force field.

SECTION: Molecular Mechanics, Quantum Chemistry, Force Field and General Theory

INTRODUCTION

Molecular mechanics (MM) modeling have been extensively employed in the research of chemistry, biochemistry, and condensed-phase physics, due to its capability of providing atomistic resolution data within relatively small to moderate computational cost. The accuracies of the MM calculation results relies on several important factors, such as sampling techniques, the underlying function forms and the parameters employed in the potential, with the latter two being the indispensable components of MM force field. Force field determines the capability of a certain model in capturing physical and chemistry properties in molecular dynamics (MD) simulations. The construction of a force field and its parameterization are non-trivial processes that require thoroughly understanding of the underlined physics and chemistry, carefully designing of parameterization protocols and appropriately applying of the optimization techniques. While several force field parameterization strategies have been widely employed and well-documented elsewhere, these approaches rely more or less on both experimental and quantum mechanics (QM) calculated properties. The present study intends to extend the existing scopes of force field developing strategies, to explore the possibility to use only QM calculation approaches for force field parameterization.

In this study, we probed the procedure for parameterization of the polarizable force field in the form of Atomic Multipole Optimized Energetics for Biomolecular Applications (AMOEBA) based on QM calculations. Comparing with non-polarizable force fields, which describes the system dipole moments in an averaged and thus fixed fashion, polarizable models were designed in a way, so that they can effectively describe the capability of chemistry species to change electronic distribution under the influence of its electrostatics environment. The polarizable models represented a significant improvement in the field of theoretical chemistry since last decade and had led to important scientific insights such as in the study of ionic liquids and ligand binding to biological macromolecules. Knowledge on efficient and accurate parameters development of polarizable models will lead to more discoveries in biophysics, biochemistry and material science. There are several polarizable force fields available for simulating various chemical and biomolecular systems, such as the Drude polarizable force field that utilizes charges on springs, the CHARMM charge equilibration (CHEQ) model that employs fluctuating charges, the AMBER ff02 using partial charges with inducible dipoles, and the AMOEBA force field that incorporates contributions from monopoles (charges), dipoles and quadrupoles, that are of higher order multipoles. The AMOEBA polarizable force field has illustrated its capability of providing reasonable predictions of interaction energy and structural properties of small molecules and clusters in gas phase, as well as the physicochemical properties of bulk phase, such as density, self-diffusivity and static dielectric constants.
In order to improve the ability to describe the physical models, AMOEBA polarizable force field invokes the utilization of a vast number of variables at the atomic site. The increments of the dimension of variables lead to a series of challenges in parameter searching and thus have been a focus of the present study. To address this issue, we implement the genetic algorithm (GA) to approach the global optimization regarding parameterizations. GA is an evolutionary algorithm that mimics the process of natural selection. For traditional parameters optimization, gradient-based algorithms, such as the steepest descent method, conjugate gradient method, etc. are the standard approach. However, for the gradient-based method, the optimal result depends strongly on the existing knowledge of the objective function. For example, the convergence of the optimal parameters is not guaranteed when the surface of the objective function is considerably rugged or non-differentiable, or when the initial value deviates significantly from the global minimum. In parameters searching of force field, the value range of parameters exponentially expands as the complexity of the force field functional form increases. To tackle the optimization algorithm issue, statistical search heuristic (such as GA, simulated annealing, etc.) is much needed to drive the parameter space towards the optimal region for the objective function. One successful example is using the GA to determine the parameters for the reactive force field (ReaxFF) for materials system.

Another issue to address in this study is the selection of reference data. Force fields that have been extensively employed and validated are often based on reference data from both experiments and theoretical calculations. For example, the development of almost all the widely employed water models are based on achieving a balance between the high level \textit{ab initio} QM calculated dataset and condensed phase properties of bulk water measured by experimentalists. Calibration of force field parameters can be achieved by minimizing the difference of the desired properties between the MM calculations and QM calculations/experimental data. For small molecules and clusters in gas phase, QM calculations are often straightforward and within reasonable computational cost. While the evaluation of the bulk phase properties from QM calculations is often computationally unfeasible, thus referring to experimental data (such as densities, dielectric constants, etc.) is necessary. However, for many molecules and especially mixtures system, experimental data are either unavailable or have significant discrepancies between different studies, due to difficulties in achieving precise measurements. Thus it is necessary to explore the possibility to select reference data of small molecules or clusters solely from theoretical calculations, \textit{i.e.} \textit{ab initio} QM calculations, as the training dataset for force field development to make the predictive calculation in computational modeling.

METHODS

The principal goal of this work is to establish guidance for constructing a framework of force field parameters development using primarily QM calculation data. In this initial attempt, we selected liquid methanol as the model system. Methanol represents a molecule of significant importance to organic liquids and polymers that contain bulky alkyl moieties. The hydroxyl group leads the molecule to the polar nature. The formation
of the methanol clusters primarily comes from the contribution of hydrogen bond networks.\textsuperscript{13} These ubiquitous aspects make methanol as an excellent candidate for designing and examination of force field development.

Ren\textit{ et al.} suggested a general protocol of parameter development for the AMOEBA polarizable force field.\textsuperscript{7,14} In the AMOEBA model, the electrostatic potential explicitly considers the effect of atomic multipoles (charge, dipoles, and quadrupoles) and polarization effect on induced dipole. The van der Waals (vdW) term is modeled as a buffered 14-7 potential proposed by Halgren.\textsuperscript{15} Compared to the 12-6 potential, the buffered 14-7 potential has the advantage of optimizing structures with initial crude geometries. For hydrogen atoms, an additional parameter (the so-called reduction factor) is used to scale the position of the hydrogen atom interacting site along the corresponding covalent bond. It is meant to reflect the degree of which hydrogenic electron density displaced toward the heavy atom when covalent bonding takes place.\textsuperscript{3e} The full potential energy function of the AMOEBA force field has been described in full detail elsewhere.\textsuperscript{2c,14} The general energy expression of the AMOEBA force field is written in equation (1), where parameterizations of electrostatic potential energy $E_{\text{ESP}}$ and vdW interaction energy $E_{\text{vdW}}$ terms are studied in this work. Bonded interaction, including bond, angle, dihedral and torsional parameters, were kept consistent with that in the original AMOEBA force field (amoeba09.prm).\textsuperscript{7}

$$E_{\text{AMOEBA}} = E_{\text{bonded}} + E_{\text{no-bonded}} = E_{\text{bonded}} + E_{\text{ESP}} + E_{\text{vdW}}$$  \(1\)

In Ren\textit{ et al.}’s protocol, the parameters of multipoles and torsional bonded terms are optimized using QM results. However, besides of taking QM calculations as the reference, the protocol adapts data from existing database of experimental values, which are not entirely available for every desired system. We followed their approaches with the context of parameterizing standard AMOEBA force field for atomic multipoles, polarizabilities and vdW interaction by GA methods using solely results from QM calculations, which are available for the methanol system within given computational resources.

All the QM calculations were performed with Gaussian 09\textsuperscript{16} package of electronic structure programs. The second order Møller–Plesset perturbation theory\textsuperscript{17} (MP2) was employed, with the basis set superposition error (BSSE) correction.\textsuperscript{18} This level of theory and basis set to predict energetics and structural properties has been widely verified by previous studies.\textsuperscript{19} The electronic density calculation of methanol monomer is initially used for calibrating the magnitude of the multipole moments at the atomic positions. Combining with the Distributed Multipole Analysis (GDMA 2.2) tool developed by Anthony Stone,\textsuperscript{20} the Tinker\textsuperscript{2c,21} package is used to optimize the permanent atomic multipoles by fitting the electrostatic potential measured on the Connolly surface of the methanol monomer. Table S1 in the Supporting Information is showing the electrostatic potential results for methanol monomer using different basis sets.
In the parameterization process, the training dataset for the optimization should contain a good sampling of the possible structural configurations and their respective energies. For instance, to predict the polarizable effect between molecules, methanol dimers with a continuous range of electrostatic potential energies \( E_{\text{ESP}} \) should be included. In AMOEBA, to compute clusters electrostatic potential energy \( E_{\text{ESP}} \) including the polarizable effect between methanol molecules, 44 independent parameters are needed. Those parameters are monopole \( (q) \), dipole \( (\mu_x, \mu_y, \mu_z) \), quadrupole-a traceless and systemic matrix- \( (Q_{xx}, Q_{yx}, Q_{yy}, Q_{zx}, Q_{zy}) \), atomic polarizability \( (\alpha) \) and Thole’s description of damping factor \( (\lambda) \) for four types of atoms (O, H(-O), C and H(-C)) in methanol (The detailed description of the parameter form can be found in somewhere else\(^{2c, 14}\)). We optimized all 44 independent parameters listed above, using an extensive training dataset taking electronic density calculations for 4943 methanol dimers. The methanol dimer configurations were optimized via Guassian09 program at the MP2 level using 6-311G(d, p) basis set. The electrostatic potential energy \( E_{\text{ESP}} \) of 4943 dimers was computed using the same level of theory and basis set. The detailed description of optimization for electrostatic parameters is in the Supporting Information Section S2.

To ensure an adequate representation of the various possible coordination environments and cluster sizes as well as the energy landscape, we sampled three training datasets containing different methanol homoclusters, which consisting of 9, 11, and 13 molecules referred as nonamer, undecamer and tridecamer in the descriptions below, respectively, to calibrate the vDW interaction energy \( E_{\text{vdW}} \). For methanol, there are vDW potential depth \( (\epsilon) \) and minimum energy distance \( (R^*) \) for four types of atoms, and additional reduction factor \( (\lambda) \) for two types of hydrogen atoms, which constitutes 10 independent parameters to optimize. We optimized these 10 independent parameters using interaction energy \( (\Delta E) \) results of the sampled 502 nonamers, 157 undecamers and 94 tridecamers at the MP2/6-31G (d, p), respectively. For instance, the nonamer was sampled through randomly placing 9 C atoms in a 10 Å x 10 Å x 10 Å computational supercell, where one C atom placed in the center and surrounded by other 8 C atoms with relative distance at least of 3.25 Å, the minimum distance of carbon bond networks in methanol,\(^{22}\) to each other. Then, methanol with random orientations will be assigned to the position where the C atoms sit. The methanol nonamer configuration is optimized in Gaussian09 using the eigenvalue-following algorithm\(^{23}\) with constraints of fixed C atoms positions. The relaxed nonamer configurations and their interaction energies \( (\Delta E) \) between the center methanol and surrounding methanols are then employed in the training dataset for optimizing the vDW parameters. This procedure was applied to undecamers and tridecamers as well for getting the relaxed configuration and the interaction energy, respectively. Figure 1 (a), (b) and (c) are showing example configurations of nonamer, undecamer and tridecamer.
In the AMOEBA force field, except for the electrostatic potential energy ($E_{ESP}$), vdW interaction energy ($E_{vdw}$) is the only term consisting of interaction energy between molecules. The vdW interaction energy ($E_{vdw}$) can be calibrated from the interaction energy ($\Delta E$) by separating apart molecules, with the assumption that the deformation energy of each molecule upon binding is reasonably small. Through sampling configurations widely and calculating the coordinated interaction energies, we consider the vdW interaction by taking the effect of electrostatic interaction inclusively. The interaction energy ($\Delta E$) is calculated as the total energy of the cluster subtracts the total energy of the center methanol molecule and that of the surrounding methanol clusters, as shown in equation (2). The BSSE correction is applied for getting the $\Delta E$ of every configuration. The MP2/6-31G (d, p) computed interaction energies of the 502 nonamer, 157 undecamer and 94 tridecamer configurations exhibit a Gaussian distribution as shown in figure 1 (d), (e) and (f), which signify the sampling of those configurations are adequate and in the equilibrated states.

$$\Delta E = E_{\text{cluster}} - E_{\text{center}} - E_{\text{sourounding}}$$ (2)
Using the training dataset described above, we optimized the non-bonded parameters in AMOEBA force field by employing genetic algorithm; the procedure is outlined in Figure 2.

Taking the optimization of vdW parameters as an illustration, we begin the optimization process by generating a population of $N_p = 120$ parameter sets randomly, such that their values lie within physically allowable limits (parameter search ranges are listed in Supporting Information Table S2). Each set of parameters in this population is called a member. For each member $i$, we compute the interaction energies for all structures in the training dataset using MD simulation package Tinker$^{2c,21}$ and evaluate the objective function $\Delta_i$ given by equation (3),

$$\Delta_i = \sum_j \left( |\Delta E_j^{MP2} - \Delta E_j^{AMOEBA} | \right)^2$$

where $\Delta E_j^{MP2}$ and $\Delta E_j^{AMOEBA}$ are the MP2 calculated and the AMOEBA calculated interaction energies for the structure $j$ in the training dataset, respectively. Since the electrostatic potential energy ($E_j^{ESP}$) is computed for the structure $j$ to represent the atomic electrostatic potential precisely, the only uncertain term in the interaction energy is the vdW interaction energy ($E_j^{vdW}$), represented by a member $i$, a set of 10 independent parameters. Here, members are then ranked in an ascending order of $\Delta_i$. After the ranking, non-linear roulette wheel selection$^{24}$ will be performed to select the top 60% members, i.e., the ones with lowest values of $\Delta_i$, which are then subjected to genetic operations: crossover with crossover-rate 3% and mutation. These mutations introduce sufficient diversity into the population, and the non-linear selection scheme helps to avoid premature convergence of the GA run. After the genetic operations, both the old and the new members are ranked by their $\Delta$s. The best $N_p$ parameter sets (members) were then chosen to constitute the next

---

Figure 2. Flowchart describing the sequence of steps employed in this work for optimization of vdW parameters in AMOEBA force field.
generation. Such an optimization routine ensures that only satisfactory parameter sets survive after each generation; upon repeating this workflow for sufficient generations and sampling viable regions in the parameters space, we perform three separate GA runs starting with different random populations. From each of the converged GA runs, we choose the final parameter set corresponding the lowest $\Delta$.

To overcome one of the serious shortcomings of MP2 theory, a noticeable overestimation of the interaction energy, which plays a major role in stability and formation of condensed phase molecular structure, we improved the form of objective function $\Delta$ in the GA program. We introduced a penalty parameter $\delta$ in the GA optimization procedure as the unexplained discrepancy between the AMOEBA results and MP2 calculations, as shown in equation (4), which obeys the design concept of guided genetic algorithm (guided GA). This penalty parameter $\delta$ never participates as the form of AMOEBA force field in MD simulations, adds up to 11 independent parameters as a subset of $i^{th}$ member in the GA optimization process for vdW parameters.

$$
\Delta_i = \sum_j \left( (\Delta E_j^{MP2} - \delta) - \Delta E_j^{AMOEBA} \right)^2 = \sum_j \left( (\Delta E_j^{MP2} - (\Delta E_j^{AMOEBA} + \delta)) \right)^2
$$

**RESULTS**

The vdW parameters are critical in the force field for getting the correct condensed phase structures, such as right intermolecular distance, through molecular dynamic simulations. In this section, we present the optimization results of the vdW parameters from both GA and guided GA program by calibrating interaction energy between molecules. We follow with showing binding curves for methanol dimers using both GA and guided GA optimized parameters. Finally, we show the MD simulation results (i.e., density and heat of vaporization) of a condensed phase methanol system from both GA and guided GA optimized parameters using AMOEBA force field.

**3.1. VDW parameters of nonamers, undecamers, and tridecamers from GA and Guided GA**

The GA and guided GA programs were applied to optimize the vdW parameters using the interaction energies ($\Delta E$) calculated at MP2/6-31G (d, p) with BSSE correction for 502, 157 and 94 methanol nonamers, undecamers and tridecamers, respectively. The excellent correlations ($R = 0.970$, 0.976, and 0.968) of the $\Delta E$ between MP2 calculations and AMOEBA with GA optimized parameters for nonamers, undecamers, and tridecamers, respectively, are shown in figure 3 (a), (b) and (c). Meanwhile with apparent overestimation of the MP2 calculation, figure 3 (d), (e) and (f) are also showing excellent correlations ($R = 0.956$, 0.965 and 0.946) of the $\Delta E$ between MP2 calculations and AMOEBA with guided GA optimized parameters for nonamers, undecamers, and tridecamers, respectively. There are in total six sets of vdW parameters from nonamers, undecamers, and tridecamers all using GA and guided GA optimization method, respectively.
3.2. VDW parameters from Dimers

To validate the transferability of the GA and guided GA optimized AMOEBA parameters from nonamers, undecamers, and tridecamers for gas-phased methanol molecules, we use the six sets of parameters to calibrate the binding energy curves of methanol dimers. We fully optimized 141 methanol dimer configurations at MP2/6-31G (d, p) with BSSE correction with the constraint of pre-defined distance between carbon atoms in the two methanol molecules. Figure 4 is showing the binding energy of methanol dimers as a function of the distance between two carbon atoms.
optimization method, respectively. Figure 5 (a), (b) and (c) are showing the good agreement of the binding energy between the MP2 calculation and the GA optimized parameters from nonamers, undecamers, and tridecamers, respectively. We also obtained the binding energy curves for those dimers using the other three sets of vdw parameters using guided GA optimization method from nonamers, undecamers, and tridecamers. Figure 5 (d), (e) and (f) are showing the binding energy curves of methanol dimers from the MP2 calculations and from the guided GA optimized vdw parameters by nonamers, undecamers, and tridecamers, respectively. From the lower row of figure 5, we can see discrepancies of binding energy between the gas-phased MP2 calculations and the guided GA optimized parameters from clusters, where the overestimation of MP2 calculations is conspicuous near the lowest energy curve.

![Figure 5](image)

The shifting down of the binding energy curves from the guided GA optimized parameters implies the interaction energy of the same level of QM theory (i.e., MP2/6-31G(d, p) + BSSE) calculations from gas-phased molecules leads to the overestimation of real potential energy surface (PES). This discrepancy of interaction energy from MP2 calculations is commonly seen as the incompleteness of the basis sets for the particular level of MP2 theory and the insufficiency sampling of configurations in the parameterization of force field.19, 25

3.3 VDW parameters for Condensed Phase Methanol

To verify the ability of our parameters of predicting the condensed phase properties, sets of MD simulations using the developed parameters of AMOEBA force field were performed. Here, we applied the six sets of optimized parameters to perform MD simulations to the condensed phase system consisting of 344 methanol molecules. For getting the density ($\rho$) and heat of vaporization ($H_{\text{vap}}$), we used Tinker package
(omm_dynamics program -- GPU version of Tinker dynamics program\textsuperscript{27}) to run MD simulation. NPT ensemble simulations were performed at $T = 298.15$ K and $P = 40$ atm with an integration time step of 1 fs. With three different random seeds for initiating velocity on the system, the averaged value of density and heat of vaporization were obtained. Table 1 shows the result of density and heat of vaporization from Tinker MD simulations with the GA optimized AMOeba force field parameters. Among all the results, the simulated density from tridecamers is the best, while the winner of the heat of vaporization is from undecamers. However, from Table 1, it is clear that the value of density and heat of vaporization are nowhere close to the experimental value.

Table 1. The density and heat of vaporization calculated from Tinker MD simulation with the GA optimized AMOeba force field from 502 nonamers, 157 undecamers, and 94 tridecamers, respectively.

| Number of methanol in cluster | Number of configurations | $\rho$ (g/mL) | $H_{\text{vap}}$ (kcal/mol) |
|-----------------------------|-------------------------|---------------|-----------------------------|
| 9                           | 502                     | 0.383         | 6.202                       |
| 11                          | 157                     | 0.415         | 6.839                       |
| 13                          | 94                      | 0.501         | 5.995                       |
| Experimental\textsuperscript{28} |                         | 0.786         | 8.950                       |

Table 2 shows the results of density and heat of vaporization from Tinker MD simulations with the guided GA optimized AMOeba force field parameters. The simulated density and heat of vaporization from tridecamers are much closer to the experimental values. We can see that using the guided GA, which introduces a penalty parameter $\delta$, eliminates the overestimation of interaction energy from MP2 calculation. The guided GA optimized parameters describe better condensed phase properties (i.e., density and heat of vaporization) than the GA optimized parameters in MD simulation. Among all the results, the simulated density and heat of vaporizations from more number of methanol molecule clusters are better. MP2 calculations of interaction energy of tridecamers and undecamers are more descriptive than nonamers for optimizing vdW parameters, due to more detailed information can be calibrated as more molecules in the solvation shell of methanol involved.

Table 2. The density and heat of vaporization calculated from Tinker MD simulation with the guided GA optimized AMOeba force field from 502 nonamers, 157 undecamers, and 94 tridecamers, respectively.

| Number of methanol in cluster | Number of configurations | $\rho$ (g/mL) | $H_{\text{vap}}$ (kcal/mol) |
|-----------------------------|-------------------------|---------------|-----------------------------|
| 9                           | 502                     | 0.726         | 6.807                       |
| 11                          | 157                     | 0.722         | 9.014                       |
| 13                          | 94                      | 0.753         | 8.825                       |
| Experimental\textsuperscript{28} |                         | 0.786         | 8.950                       |

CONCLUSIONS

Our study showed that the feasibility to predict experimental condensed phase properties (i.e., density and heat of vaporization) of methanol through problem specific (here, we use AMOeba) force field from only quantum chemistry information. To acquire the satisfying parameter sets, the genetic algorithm (GA) is the main optimization method. For electrostatic potential energy ($E_{\text{ESP}}$), we optimized the both the multipoles and
polarizability, Thole damping factors of methanol using the GA method, which leads to lower deviations of $E_{\text{ESP}}$ between QM calculations and the GA optimized parameters than that from amoeba09.prm. We optimized the vdW parameters both using GA and guided GA methods by calibrating interaction energy ($\Delta E$) of various methanol homo-clusters, such as nonamers, undecamers, or tridecamers. Excellent agreement between the training dataset from MP2 calculations and GA optimized parameters can be achieved. However, only the guided GA method, which eliminates the overestimation of interaction energy from MP2 calculations in the optimization process, provides proper vdW parameters for MD simulation to get condensed phase properties (i.e., density and heat of vaporization) of methanol. Throughout the whole optimization process, the experimental value were not involved in the objective functions, but were only used for the purpose of justifying models (i.e., nonamers, undecamers, or tridecamers) and validating methods (GA or guided GA). We conclude that the main difficulty of parameterizing the force field for liquid phase methanol from solely QM calculations could be coming from the reliability of the training dataset (i.e., MP2 calculations), as it is well known that different basis sets of MP2 could lead to different accuracy of QM calculations, the conspicuous overestimation of interaction energy from MP2 calculations. Some of the other higher level of QM calculations (e.g., MP3/4, CCSD(T), etc.) can only deal with up to certain amount of atoms/electrons. It is ambiguous to study the property of condensed phase properties from a few molecules. In fact, we tried to use the interaction energy ($\Delta E$) calculated for methanol tridecamers from another type of QM calculation, Symmetry-Adapted Perturbation Theory (SAPT), as the training dataset using GA optimization, which yields a lousy performance of the condensed phase properties. However, our method was able to utilize the QM calculations (i.e., MP2 calculations) from the limited calculation resource to tune the force field parameters for reasonable agreement between the MD simulated values and experimental data. The method is general and can be extended to any type of small organic molecule systems, as well as other descriptive polarizable force field.

ACKNOWLEDGMENTS
This work was supported by XXX Grant No. YYY. Simulations were performed on Blues computers at the Laboratory Computing Resource Center and Cooley computers at the Leadership Computing Facility at the Argonne National Laboratory, and on the Midway computers of the Research Computing Center at the University of Chicago. We appreciate the Margaret Butler Postdoctoral Fellowship at Argonne Leadership Computing Facility for supporting the work. We thanks to Dr. Frank C. Pickard IV and Dr. Bernard R. Brooks from the Computational Biophysics Section of the Laboratory of Computational Biology at the National Institutes of Health for the very useful discussion on calculation on SAPT method.

REFERENCES
1. (a) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorikiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B
1998, 102 (18), 3586-3616; (b) Foloppe, N.; MacKerell, A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. J Comput Chem 2000, 21 (2), 86-104; (c) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 2004, 25 (11), 1400-1415; (d) Klauda, J. B.; Venable, R. M.; MacKerell, A. D.; Pastor, R. W. Considerations for lipid force field development. Curr Top Membr 2008, 60, 1-48; (e) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. J Phys Chem B 2010, 114 (23), 7830-7843; (f) Lim, J. B.; Rogaski, B.; Klauda, J. B. Update of the Cholesterol Force Field Parameters in CHARMM. J Phys Chem B 2012, 116 (1), 203-210.

2. (a) Lamoureux, G.; MacKerell, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. J Chem Phys 2003, 119 (10), 5185-5197; (b) Lamoureux, G.; Roux, B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. J Chem Phys 2003, 119 (6), 3025-3039; (c) Ren, P. Y.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. J Phys Chem B 2003, 107 (24), 5933-5947; (d) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator. J Chem Theory Comput 2005, 1 (1), 153-168; (e) Zhang, J.; Tuguldur, B.; van der Spoel, D. Force Field Benchmark of Organic Liquids. 2. Gibbs Energy of Solvation (vol 55, pg 1192, 2015).

3. (a) Hensen, C.; Hermann, J. C.; Nam, K. H.; Ma, S. H.; Gao, J. L.; Holttje, H. D. A combined QM/MM approach to protein-ligand interactions: Polarization effects of the HIV-1 protease on selected high affinity inhibitors. J Med Chem 2004, 47 (27), 6673-6680; (b) Yan, T. Y.; Li, S.; Jiang, W.; Gao, X. P.; Xiang, B.; Voth, G. A. Structure of the liquid-vacuum interface of room-temperature ionic liquids: A molecular dynamics study. J Phys Chem B 2006, 110 (4), 1800-1806; (c) Khoruzhii, O.; Donchev, A. G.; Galkin, N.; Illarionov, A.; Olevanov, M.; Ozrin, V.; Queen, C.; Tarasov, V. Application of a polarizable force field to calculations of relative protein-ligand binding affinities. P Natl Acad Sci USA 2008, 105 (30), 10378-10383; (d) Borodin, O. Polarizable Force Field Development and Molecular Dynamics Simulations of Ionic Liquids. J Phys Chem B 2009, 113 (33), 11463-11478; (e) Ponder, J. W.; Wu, C. J.; Ren, P. Y.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. J Phys Chem B 2010, 114 (8), 2549-2564; (f) Chaban, V. Polarizability versus mobility: atomistic force field for ionic liquids. Phys Chem Chem Phys 2011, 13 (35), 16055-16062; (g) Soderhjelm, P. Polarization effects in protein-ligand calculations extend farther than the actual induction energy. Theor Chem Acc 2012, 131 (3).

4. Harder, E.; Anisimov, V. M.; Whitfield, T. W.; MacKerell, A. D.; Roux, B. Understanding the dielectric properties of liquid amides from a polarizable force field. J Phys Chem B 2008, 112 (11), 3509-3521.

5. Bauer, B. A.; Patel, S. Recent applications and developments of charge equilibration force fields for modeling dynamical charges in classical molecular dynamics simulations. Theor Chem Acc 2012, 131 (3).

6. Wang, Z. X.; Zhang, W.; Wu, C.; Lei, H. X.; 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 (6), 781-790.

7. Ren, P. Y.; Wu, C. J.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. J Chem Theory Comput 2011, 7 (10), 3143-3161.

8. McCall, J. Genetic algorithms for modelling and optimisation. J Comput Appl Math 2005, 184 (1), 205-222.

9. (a) Powell, M. J. D. Nonconvex Minimization Calculations and the Conjugate-Gradient Method. Lect Notes Math 1984, 1066, 122-141; (b) Fliege, J.; Svaiter, B. F. Steepest descent methods for multicriteria optimization. Math Method Oper Res 2000, 51 (3), 479-494.

10. (a) van Duin, A. C. T.; Dасguptа, S.; Lorant, F.; Goddard, W. A. ReaxFF: A reactive force field for hydrocarbons. J Phys Chem A 2001, 105 (41), 9396-9409; (b) Pahari, P.; Chaturvedi, S. Determination of best-fit potential parameters for a reactive force field using a genetic algorithm. J Mol Model 2012,
18 (3), 1049-1061; (c) Larsson, H. R.; van Duin, A. C. T.; Hartke, B. Global optimization of parameters in the reactive force field ReaxFF for SiOH. J Comput Chem 2013, 34 (25), 2178-2189.

11. (a) Chen, B.; Xing, J. H.; Siepmann, J. I. Development of polarizable water force fields for phase equilibrium calculations. J Phys Chem B 2000, 104 (10), 2391-2401; (b) Ren, P. Y.; Ponder, J. W. Temperature and pressure dependence of the AMOEBA water model. J Phys Chem B 2004, 108 (35), 13427-13437; (c) Paesani, F.; Iuchi, S.; Voth, G. A. Quantum effects in liquid water from an ab initio-based polarizable force field. J Chem Phys 2007, 127 (7); (d) Paesani, F.; Xantheas, S. S.; Voth, G. A. Infrared Spectroscopy and Hydrogen-Bond Dynamics of Liquid Water from Centroid Molecular Dynamics with an Ab Initio-Based Force Field. J Phys Chem B 2009, 113 (39), 13118-13130; (e) Laury, M. L.; Wang, L. P.; Pande, V. S.; Head-Gordon, T.; Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. J Phys Chem B 2015, 119 (29), 9423-9437.

12. Pinnick, E. R.; Erramilli, S.; Wang, F. Predicting the melting temperature of ice-Ih with only electronic structure information as input. J Chem Phys 2012, 137 (1).

13. Kobayashi, T.; Shishido, R.; Mizuse, K.; Fuji, A.; Kuo, J. L. Structures of hydrogen bond networks formed by a few tens of methanol molecules in the gas phase: size-selective infrared spectroscopy of neutral and protonated methanol clusters. Phys Chem Chem Phys 2013, 15 (24), 9523-9530.

14. Wu, J. C.; Chattree, G.; Ren, P. Y. Automation of AMOEBA polarizable force field parameterization for small molecules. Theor Chem Acc 2012, 131 (3).

15. Halgren, T. A. Representation of Vanderwaals (Vdw) Interactions in Molecular Mechanics Force-Fields - Potential Form, Combination Rules, and Vdw Parameters. J Am Chem Soc 1992, 114 (20), 7827-7843.

16. 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 Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; 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.; Foreman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Gaussian, Inc.: Wallingford, CT, USA, 2009.

17. Headgordon, M.; Pople, J. A.; Frisch, M. J. Mp2 Energy Evaluation by Direct Methods. Chem Phys Lett 1988, 153 (6), 503-506.

18. Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors (Reprinted from Molecular Physics, vol 19, pg 553-566, 1970). Mol Phys 2002, 100 (1), 65-73.

19. Li, H.; Ngo, V.; Da Siva, M. C.; Salahub, D. R.; Callahan, K.; Roux, B.; Noskov, S. Y. Representation of Ion-Protein Interactions Using the Drude Polarizable Force-Field. J Phys Chem B 2015, 119 (29), 9401-9416.

20. (a) Stone, A. J. Distributed Multipole Analysis, or How to Describe a Molecular Charge-Distribution. Chem Phys Lett 1981, 83 (2), 233-239; (b) Stone, A. J.; Alderton, M. Distributed multipole analysis - Methods and applications (Reprinted from Molecular Physics, vol 56, pg 1047-1064, 1985). Mol Phys 2002, 100 (1), 221-233.

21. (a) Ponder, J. W.; Richards, F. M. An Efficient Newton-Like Method for Molecular Mechanics Energy Minimization of Large Molecules. J Comput Chem 1987, 8 (7), 1016-1024; (b) Kundrot, C. E.; Ponder, J. W.; Richards, F. M. Algorithms for Calculating Excluded Volume and Its Derivatives as a Function of Molecular-Conformation and Their Use in Energy Minimization. J Comput Chem 1991, 12 (3), 402-409; (c) Dudek, M. J.; Ponder, J. W. Accurate Modeling of the Intramolecular Electrostatic Energy of Proteins. J Comput Chem 1995, 16 (7), 791-816; (d) Kong, Y.; Ponder, J. W. Calculation of the reaction field due to off-center point multipoles. J Chem Phys 1997, 107 (2), 481-492; (e) Pappu, R. V.; Hart, R. K.; Ponder, J. W. Analysis and application of potential energy smoothing and search methods for global optimization. J Phys Chem B 1998, 102 (48), 9725-9742.

22. Dougan, L.; Bates, S. P.; Hargreaves, R.; Fox, J. P.; Crain, J.; Finney, J. L.; Reat, V.; Soper, A. K. Methanol-water solutions: A bi-percolating liquid mixture. J Chem Phys 2004, 121 (13), 6456-6462.
23. (a) Cerjan, C. J.; Miller, W. H. On Finding Transition-States. *J Chem Phys* **1981**, *75*(6), 2800-2806; (b) Simons, J.; Jorgensen, P.; Taylor, H.; Ozment, J. Walking on Potential-Energy Surfaces. *J Phys Chem Us* **1983**, *87*(15), 2745-2753; (c) Khait, Y. G.; Puzanov, Y. V. Search for stationary points on multidimensional surfaces. *J Mol Struc-Theochem* **1997**, *398*, 101-109.

24. (a) Lipowski, A.; Lipowska, D. Roulette-wheel selection via stochastic acceptance. *Physica A* **2012**, *391* (6), 2193-2196; (b) Razali, N. M.; Geraghty, J. In *Genetic Algorithm Performance with Different Selection Strategies in Solving TSP*, Proceedings of the World Congress on Engineering, 2011.

25. (a) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys Chem Chem Phys* **2006**, *8*(17), 1985-1993; (b) Cybulski, S. M.; Lytle, M. L. The origin of deficiency of the supermolecule second-order Moller-Plesset approach for evaluating interaction energies. *J Chem Phys* **2007**, *127*(14); (c) Tkatchenko, A.; DiStasio, R. A.; Head-Gordon, M.; Scheffler, M. Dispersion-corrected Moller-Plesset second-order perturbation theory (vol 131, 094106, 2009). *J Chem Phys* **2009**, *131*(12).

26. (a) Leng, L. T. Guided genetic algorithm. *University of Essex, A thesis submitted for the degree of Ph. D in Computer Science, Department of Computer Science. 1999*; (b) El-Mihoub, T. A.; Hopgood, A. A.; Nolle, L.; Battersby, A. Hybrid Genetic Algorithms: A Review. *Engineering Letters* **2006**, *13*(2), 124-137.

27. TINKER-OPENMM-Contributors. Main Page -- TINKER-OPENMM. [http://biomol.bme.utexas.edu/tinker-openmm/index.php?title=Main_Page&oldid=18](http://biomol.bme.utexas.edu/tinker-openmm/index.php?title=Main_Page&oldid=18).

28. Riddick, J. A. B., W. B.; Sakano, T.; Weissberger, A. Organic Solvents: Physical Properties and Methods of Purification. *Wiley: New York 1986*, (4th ed.).

29. (a) Woon, D. E. Accurate Modeling of Intermolecular Forces - a Systematic Moller-Plesset Study of the Argon Dimer Using Correlation Consistent Basis-Sets. *Chem Phys Lett* **1993**, *204*(1-2), 29-35; (b) Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. Scaled MP3 Non-Covalent Interaction Energies Agree Closely with Accurate CCSD(T) Benchmark Data. *ChemPhysChem* **2009**, *10*(1), 282-289; (c) Řezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the “Gold Standard,” CCSD(T) at the Complete Basis Set Limit? *J Chem Theory Comput* **2013**, *9*(5), 2151-2155.