UC Berkeley

UC Berkeley Previously Published Works

Title

Mode specific THz spectra of solvated amino acids using the AMOEBA polarizable force field.

Permalink

<https://escholarship.org/uc/item/4dm041qd>

Journal

Physical chemistry chemical physics : PCCP, 19(7)

ISSN

1463-9076

Authors

Esser, Alexander
Belsare, Saurabh
Marx, Dominik
et al.

Publication Date

2017-02-01

DOI

10.1039/c6cp07388c

Peer reviewed

Mode Specific THz Spectra of Solvated Amino Acids using the AMOEBA Polarizable Force Field

Alexander Esser,^{1, a)} Saurabh Belsare,^{2, a)} Dominik Marx,^{1, b)} and Teresa Head-Gordon^{3, c)}

¹⁾*Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany*

²⁾*The UCB-UCSF Graduate Program in Bioengineering, University of California Berkeley, CA 94720, United States*

³⁾*Kenneth S. Pitzer Center for Theoretical Chemistry, Department of Chemistry, Department of Bioengineering, Department of Chemical and Biomolecular Engineering, University of California Berkeley, CA 94720, United States*

(Dated: January 26, 2017)

We have used the AMOEBA model to simulate the THz spectra of two zwitterionic amino acids in aqueous solution, which is compared to the results on these same systems using *ab initio* molecular dynamics (AIMD) simulations. Overall we find that the polarizable force field shows promising agreement with AIMD data for both glycine and valine in water. This includes the THz spectral assignments and the mode-specific spectral decomposition into intramolecular solute motions as well as distinct solute-water cross-correlation modes some of which cannot be captured by non-polarizable force fields that rely on fixed partial charges. This bodes well for future studies for simulating and decomposing the THz spectra for larger solutes such as proteins or polymers for which AIMD studies are presently intractable. Furthermore, we believe that the current study on rather simple aqueous solutions offers a way to systematically investigate the importance of charge transfer, nuclear quantum effects, and the validity of computationally practical density functionals, all of which are needed to fully quantitatively capture complex dynamical motions in the condensed phase.

I. INTRODUCTION

Understanding the molecular motions that arise from solute-solvent interactions is one of the key problems in Solvation Science. Terahertz (THz) and related spectroscopies have proven to be a sensitive tool in order to probe solvation shell dynamics around a variety of solutes, from simple single monovalent ions to complex biological systems like proteins and enzymes^{1–7}. But in order to understand the experimentally observed spectra in molecular detail, theoretical methods are required. For small molecular species and simple ions, *ab initio* molecular dynamics (AIMD) simulations⁸ have proven to give a reasonably faithful description of the THz experimental observable, and hence can be relied upon to decompose the motions of solvation shell dynamics^{9–13}; see Ref. 14 for a review of the techniques underlying the AIMD approach to theoretical infrared (and thus THz) spectroscopy. In particular for zwitterionic glycine in aqueous solution, AIMD interpreted the THz observable to have three major modes of motion, including rigid body translational motions of the whole molecule at low frequencies ($\sim 80\text{ cm}^{-1}$), intermolecular cross correlation modes due to the interaction of the zwitterion with the solvent at $\sim 200\text{ cm}^{-1}$, after which purely intramolecular angle bending modes are present¹².

However, the computational cost is a limiting factor for extending AIMD to larger systems or to faithfully probe

solvation dynamics beyond the second solvation shell. In principle, force field simulations should easily allow extension to larger systems due to their more tractable cost even when using polarizable versions^{15–28}, however there is a need for validation against AIMD to ensure that the THz spectra and mode decomposition are consistent using the simpler model to describe the interatomic interactions. In the THz regime, electronic polarization and/or charge transfer effects, which are included in AIMD simulations since they rely on solving self-consistently the electronic structure problem on-the-fly⁸, are of particular importance¹⁰. In turn the more tractable force fields can probe any potential problems with finite system size effects, as well as cross-validate the AIMD protocols for simulating the THz spectra and assumptions for interpreting the low frequency modes.

In this study the AMOEBA polarizable model^{24,29} is tested for its ability to reproduce the results given by AIMD on the solvent induced intramolecular and intermolecular motions of the zwitterionic form of single glycine and valine molecules in water. We have chosen AMOEBA since validation studies on bulk water have demonstrated that the THz observable is qualitatively reproduced (Fig.1). It is noteworthy that the signature of the intermolecular vibrations of the water network in the $\approx 200\text{ cm}^{-1}$ (or $\approx 6\text{ THz}$) region is captured by the direct polarization iAMOEBA³⁰ and full mutual polarization AMOEBA models³¹, whereas if we turn-off the many-body polarization component, this feature is lost from the simulated THz spectrum (Fig.1). This suggests that more standard fixed partial charge models would be insufficient for representing intermolecular interactions probed by the THz experiment^{10,32}, hence

^{a)}These two authors contributed equally

^{b)}Electronic mail: dominik.marx@rub.de

^{c)}Electronic mail: thg@berkeley.edu

Figure 1. THz spectra, $\alpha(\omega)n(\omega)$, of pure bulk water for AMOEBA water with (solid red line) and without (dashed red line) the induced dipole contributions (see text) compared to the corresponding experimental data (adapted from the SI of Ref. 30). Note: AMOEBA 14 is only a water model, and that water model had been consistently used throughout all the simulations in this paper, for both the AMOEBA original and AMOEBA new simulations outlined in the methods and results sections.

we require at least many-bodied polarization²⁷ as a minimum level of physics for the force field that might replace the AIMD simulations for larger systems. However there are other important quantum mechanical features that are not currently accounted for in AMOEBA, but are clearly present in the AIMD simulations, including charge penetration and charge transfer. Although active work for incorporating these important short-ranged interactions are under active development within the force field community^{23,33–39}, and are starting to be introduced in more standard molecular mechanics models^{40–48}, they are not present in the current version of AMOEBA.

As will be detailed based on comprehensive analyses for glycine and valine in water, we find that the AMOEBA model performs well in comparison to AIMD in terms of capturing the intramolecular modes and the hindered translation (cage rattling) and hindered rotation (libration) modes of the zwitterions, as well as the intermolecular cross correlation modes of the zwitterion with water. It is noted in passing that the AMOEBA parameters for the two single zwitterionic amino acids in AMOEBA14 water had to be developed based on a systematic protocol. What is remarkable is the level of agreement between the polarizable force field and electronic structure based treatments given the differences in how the molecular dipole moments are calculated and the assumptions that go into the mode decomposition that uses a charge weighted velocity cross-correlation matrix. An additional benefit to the AMOEBA investigation here is to examine the potential influence of finite size effects on the calculated THz observables in the AIMD study, for which we find no issues, except for the simple loss of information

for outer water shell dynamics beyond the first solvation shell in the present case.

The remainder of this paper is outlined as follows. In section II we describe the theoretical models and methods. The resulting data and analysis for AMOEBA simulations of the THz spectra and mode decompositions of the two amino acids are compared against the AIMD results and discussed in Sec. IV. The insights gained from these benchmark calculations of AMOEBA are discussed in Sec. V and plans for future studies are discussed.

II. THEORETICAL MODELS AND SIMULATION METHODS

A. AMOEBA Model

The AMOEBA potential energy is formulated as

$$U = U_b + U_\theta + U_{tors} + U_{b\theta} + U_{oop} + U_{vdW} + U_{ele}^{perm} + U_{ele}^{pol} \quad (1)$$

where U_b and U_θ correspond to harmonic bond and angle deformations, U_{tors} is a truncated Fourier series to describe rotations around bonds, $U_{b\theta}$ is a Urey-Bradley coupling term, and U_{oop} comprises the out-of-plane bending energy, while the last three terms embody the non-bonded interactions. Given that non-bonded terms are the most important aspect of solute-solvent interactions, we describe them in more detail.

The first non-bonded term is the permanent electrostatics (U_{ele}^{perm}) based on an atom-centered point multipole on each atomic site i , comprising monopole (q_i), dipole (μ_i), and quadrupole (Q_i) moments:

$$\mathbf{M}_i = [q_i, \mu_{ix}, \mu_{iy}, \mu_{iz}, Q_{ixx}, Q_{ixy}, Q_{ixz}, Q_{iyy}, Q_{iyz}] \quad (2)$$

The total permanent electrostatics contribution is then evaluated as the pairwise sum of interactions between different atomic sites:

$$E_{ele}^{perm} = \sum_{i < j} \mathbf{M}_i \mathbf{T}_{ij} \mathbf{M}_j \quad (3)$$

where $\mathbf{T}_{ij}$ is the “composite” multipole interaction tensor between sites i and j , whose exact form can be found in Ref. 49.

The polarization effect in AMOEBA is modeled by induced dipoles placed on each atomic site, whose magnitude is determined by the site-specific isotropic polarizability and the total external electric field exerted:

$$\mu_i^{ind} = \alpha_i (\mathbf{E}_i + \mathbf{E}'_i) \quad (4)$$

where $\mathbf{E}_i$ is the electric field owing to the permanent multipoles on other fragments, and $\mathbf{E}'_i$ is the field generated by the induced dipoles on all the other atomic sites:

$$\mathbf{E}_i = \sum_j \mathbf{T}_{ij}^d \mathbf{M}_j^{perm} \quad (5)$$

$$\mathbf{E}'_i = \sum_{j \neq i} \mathbf{T}_{ij}^{d-d} \mu_j^{ind} \quad (6)$$

Since the RHS of Eq. 6 relies on the induced dipoles, μ_i^{ind} 's are solved self-consistently to capture many-body polarization effects. With converged $\{\mu_i^{ind}\}$, the associated energy lowering (the contribution of induced electrostatics) is determined by

$$E_{ele}^{ind} = -\frac{1}{2} \sum_i \mu_i^{ind} \cdot \mathbf{E}_i \quad (7)$$

One artifact of a distributed interactive induced electrostatics model is the so-called ‘‘polarization catastrophe’’⁵⁰, i.e., the electric field generated by point multipoles can severely overpolarize at short range and even lead to divergence. To ensure the finite nature of intermolecular induction effect, a Thole-style damping scheme is employed by AMOEBA, which is equivalent to replacing a point multipole with a smeared charge distribution. The damping function forms for all multipoles are reported in Ref. 19. In practice, the damping functions are built in the formation of multipole interaction tensors in Eq. 5 and 6. Atomic polarizabilities are obtained as derived by Thole⁵⁰ and used with a modified damping factor (0.39) from Thole’s original (0.567) as outlined in Ref. 29.

For the van der Waals interaction, AMOEBA adopts a pairwise additive buffered 14-7 potential that was originally proposed by Halgren⁵¹:

$$E_{vdw} = \sum_{i < j} \epsilon_{ij} \left(\frac{1 + \delta}{\rho_{ij} + \delta} \right)^7 \left(\frac{1 + \gamma}{\rho_{ij}^{\gamma} + \gamma} - 2 \right) \quad (8)$$

where ϵ_{ij} is the depth of the potential well, ρ_{ij} is the dimensionless distance between sites i and j : $\rho_{ij} = R_{ij}/R_{ij}^0$, where R_{ij}^0 is the equilibrium distance. γ and δ are two constants whose values are set to 0.12 and 0.07, respectively. The combination rules for heterogeneous atom pairs that determine ϵ_{ij} and R_{ij}^0 are:

$$R_{ij}^0 = \frac{(R_{ii}^0)^3 + (R_{jj}^0)^3}{(R_{ii}^0)^2 + (R_{jj}^0)^2}, \quad \epsilon_{ij} = \frac{4\epsilon_{ii}\epsilon_{jj}}{(\epsilon_{ii}^{1/2} + \epsilon_{jj}^{1/2})^2} \quad (9)$$

B. Parameterization of Zwitterionic Amino Acids

The parameters in the AMOEBA force field have been developed with applications to large proteins in mind, and thus no parameters exist for single amino acid side chains in their zwitterionic form⁵². Furthermore given the centrality of water to the solvation study and the need for accuracy, we opted to work with the new AMOEBA14 water model³¹ which provides a robust description of bulk water properties, but which requires reparameterization to work with other solutes. Thus a new set of non-bonded parameters of the glycine and valine solutes were required for compatibility with the AMOEBA14 water model. The standard AMOEBA parameterization

protocol²⁹ was followed, with the exception for deriving the multipoles, since the first step of performing geometry optimization in vacuum converts the zwitterions into neutral molecules as expected. Instead, 100 structures spaced 2 ps apart from a 200 ps AIMD trajectory served as input structures for the following parameterization calculations.

While most of the valence parameters in Eq. 1 are defined and thus transferable from the existing AMOEBA13 parameter set, the parameters for the $N - C_{\alpha} - C$ bond angle parameter specific to a zwitterion do not exist, and the vdW parameters on the carboxyl oxygens and amino hydrogens required optimization to account for modified interactions with the AMOEBA14 water model. The ForceBalance software⁵³ was used to derive the van der Waals parameters and bond angle force constant using static quantum chemical *ab initio* calculations as reference data. We generated the quantum mechanical energy and force calculation fitting data based on the MP2 method together with the 6-311G(1d,1p) basis, to maintain consistency with the charge and multipole parameterization protocol below, using *Q-chem*⁵⁴. All of the newly derived zwitterionic non-electrostatic parameters (vdW and bond angle) parameters obtained for glycine were transferred to valine without re-optimization, demonstrating transferability.

Using the five lowest energy of the 100 available structures, we obtained the permanent atomic multipoles from the distributed multipole analysis *via* Stone’s *DMA* program⁵⁵ based on single point MP2/6-311G(1d,1p) calculations using *Gaussian*⁵⁶. The *TINKER poledit* utility was used to rotate the atomic multipoles obtained from *DMA* to *TINKER* defined local frames. This also defines Thole intramolecular polarization, for which polarization groups are defined based on Ref. 52: methyl, carbonyl and amine groups. This gives us an initial estimate of the multipole values. These values are further refined by performing a single point MP2/aug-cc-pVTZ calculation in *Gaussian*⁵⁶, which is used to derive the electron density and subsequently construct the electrostatic potential on a grid of points outside the vdW envelope using *Cubegen*⁵⁶. The *TINKER potential* program then refines the atomic multipoles based on the quantum mechanical electrostatic potential. The *DMA* monopole values are not modified from the initial values in the refinement step.

C. Validation: AMOEBA versus AIMD

Figure 2 shows a comparison of the glycine-water radial distribution functions (RDFs) computed using data obtained from simulations from the original and modified AMOEBA parameters, and compared with the same RDFs from the AIMD calculations and from experiment⁵⁷. It is observed that the first solvation shells of the two charged groups of the zwitterion as obtained from the AIMD simulations agree convincingly

Figure 2. Radial distribution functions for a single zwitterionic glycine molecule in water: (a) amide-H and water-H, (b) carboxyl-O and water-H, (c) amide-H and water-O, (d) carboxyl-O and water-O sites according to the legend; see text for the corresponding methods and references.

with the experimental data within rather small differences in peak positions, whereas the second shells feature increased deviations. The original AMOEBA model, in stark contrast, does not reliably capture even the first solvation shell structure, both in terms of peak positions and peak heights, and thus does not accurately represent the hydrogen bonding pattern of these important hydrophilic functional groups (note, however, that AMOEBA was never parameterized to study individual zwitterionic amino acids in water!). After reparameterization without any reference or fitting to our AIMD data, the RDFs of the resulting modified AMOEBA model shift much closer to the AIMD results and thus to experiment. Similar agreement between AIMD and AMOEBA is found for valine as shown in the supplementary information (SI). Further structural comparisons in terms of intramolecular angle and dihedral distributions are also shown in the SI, in which the modified AMOEBA and AIMD distributions match similarly well. Hence, it is clear that the principal structural information of the amino acids is retrieved, and comparable to the results obtained from the AIMD calculations.

Within the AIMD approach to theoretical infrared spectroscopy¹⁴, the corresponding linear absorption cross sections, $\alpha(\omega)$, are obtained from auto-correlation functions of the dipole moments that are obtained from concurrent electronic structure calculations. Thus, in addition to the solvation shell structure around the solute molecule, also its dipole moment in solution is of key importance. Figure 3 compares the dipole distributions of glycine and valine of the original and reparameterized AMOEBA force field to AIMD. As expected, the

Figure 3. Dipole moment distribution functions for a single zwitterionic glycine and valine molecule in water split into solute and solvent contributions according to the legend; see text for the corresponding methods.

water dipole distribution shows good agreement when compared with the AIMD distribution whereas for the amino acids, the original parameters does not agree with the AIMD distribution well enough for the present purpose. However, after reparameterization the dipole distributions align with the AIMD to a reasonable extent.

D. Simulation Protocols: AIMD and AMOEBA

For the AIMD calculations⁸, the glycine and valine aqueous solutions were simulated using the PBE functional with pseudo potentials and a plane wave cut-off energy of 400 Ry and a TZV2P basis set using the *CP2k* program package⁵⁸; see the SI of Ref. 12 for comprehensive computational details. A cubic simulation cell is used with a side length of 9.85 Å for glycine and 12.49 Å for valine. Each cell contains one amino acid molecule and 30 water molecules in the glycine case and 60 water molecules in the valine case. After equilibration long AIMD simulations have been carried out in the NVT ensemble using Nosé-Hoover chain thermostats⁸ at a rescaled temperature of 400 K to approximately counterbalance its systematic underestimation by about 20–30 % thus following the approach introduced some time back for pure water and aqueous solutions^{9,10,12,59}. This *ad hoc* method not only provides agreement of radial distribution functions and the diffusion coefficient of water, but also accounts for the proper THz intensities compared to experiment⁹. From the NVT trajectory, 80 independent starting structures (and corresponding velocities) for glycine and 60 for valine are sampled at equidistant points as starting structures for NVE trajectories. Each NVE trajectory is simulated for 20 ps with an integration time step of 0.5 fs, and the maximally localized Wannier functions (MLWFs)⁸ are computed every 2 fs.

For the AMOEBA force field simulations, glycine is solvated with 30 water molecules and valine with 60 wa-

ter molecules, resulting in a cubic boxes of side lengths of 10.20 Å for glycine and 13.01 Å for valine. After initial equilibration in the NPT ensemble, we generated a 200 ps long NVT trajectory at 300 K using a Bussi thermostat, and sampled conformations and velocities every 2 ps. These served as starting structures and velocities for 100 independent NVE trajectories that were also simulated for 20 ps. The time step of integration is 0.5 fs and the configurations and induced dipoles are written out every 2 fs. Due to much faster computations using the AMOEBA model, the systems can be easily increased in the number of water molecules in order to probe finite size effects. Towards that goal, larger systems of both glycine and valine with 253 and 256 water molecules, respectively, were also studied. These systems had box lengths of 19.73 Å for glycine and 20.17 Å for valine. All the AMOEBA simulations were performed in the *TINKER* molecular dynamics package.

III. THEORETICAL THZ SPECTROSCOPY

A. Computing THz Spectra

In the limit of classical nuclear motion⁶⁰, the total linear infrared absorption cross section is given by¹⁴

$$\alpha(\omega) = \frac{1}{n(\omega)} \frac{1}{6\epsilon_0 V c} \frac{1}{k_B T} I(\omega) , \quad (10)$$

where V is the volume of the solution that is simulated at temperature T , ω is the frequency, $n(\omega)$ is the refractive index, and ϵ_0 and c are the vacuum permittivity and speed of light, respectively. It is noted in passing that the prefactor in that expression includes the “harmonic quantum correction factor”⁶⁰ if using $I(\omega)$ expressed in terms of charge current time-correlations as specified next (thus taking into account the resulting extra ω^{-2} factor in front of $I(\omega)$ in Eq. 10). Here, the line-shape function $I(\omega)$ is given *via* the Fourier transform of the dipole velocity auto-correlation as follows

$$I(\omega) = \int_{-\infty}^{+\infty} dt \langle \dot{\mathbf{M}}(0) \dot{\mathbf{M}}(t) \rangle e^{-i\omega t} , \quad (11)$$

where the total dipole moment $\mathbf{M}(t)$ can be defined as the vector sum of the (effective) molecular dipole moments $\boldsymbol{\mu}_J(t)$ in solution,

$$\mathbf{M}(t) = \sum_{J=1}^{N_M} \boldsymbol{\mu}_J , \quad (12)$$

where N_M is the total number of molecules in the entire system. The time-derivative of the total dipole moment vector (being the total charge current) of the simulation box, $\dot{\mathbf{M}}(t)$, is computed as a finite difference quantity from consecutive configuration frames. Restricted

summation within Eq. 12 allows one to compute spectral contributions stemming from specified subsystems, for instance “solute-only” spectra if only the (effective) dipole moment (velocity) for instance of glycine, $\boldsymbol{\mu}_{J=\text{Gly}}$, is considered. Since $I(\omega)$ is easily accessible to molecular dynamics trajectories, it is the product of the absorption cross section and the refractive index, $\alpha(\omega)n(\omega)$, that is straightforwardly obtained and thus mostly reported in the literature. However, $n(\omega)$ can be computed by applying the Kramers-Kronig relation as described in the SI of Ref. 12 so that the absorption coefficient $\alpha(\omega)$ itself is obtained which is indeed the experimental observable.

Computation of the molecular dipole moments $\boldsymbol{\mu}_J$ is fundamentally different between AIMD and AMOEBA simulations. For AIMD the molecular dipole moment is simply the sum of the product of charges and their cartesian positions of all charge centers $i = 1, \dots, N_J$ in the molecule J that is considered⁸,

$$\boldsymbol{\mu}_{J,\text{aimd}} = \sum_{i=1}^{N_J} q_i \mathbf{r}_i \quad (13)$$

where i labels the charge centers irrespective of their nature. In AIMD, each nucleus is a charge center that contributes its positive nuclear core charge $q_i = +Z^{\text{core}}|e|$ (thus taking into account the reduction of the bare nuclear charge Z whenever pseudo potentials are used to replace core electrons), whereas each Wannier charge center position carries a negative charge of $q_i = -2|e|$ in case of the mostly used doubly-occupied closed-shell representation of the valence electronic structure in terms of (maximally localized Wannier valence) molecular orbitals; note that this charge would be $-|e|$ in open-shell spin-polarized calculations where singly-occupied spin orbitals are used.

AMOEBA, being a polarizable point multipole based force field, uses both permanent and induced dipoles centered at each atom I , in addition to monopoles, all of which contribute to the dipole moment of molecule J ,

$$\boldsymbol{\mu}_{J,\text{amoeba}} = \sum_{I=1}^{N_J} (\boldsymbol{\mu}_I^{\text{perm}} + \boldsymbol{\mu}_I^{\text{ind}}) , \quad (14)$$

where $\boldsymbol{\mu}_I^{\text{perm}}$ is the contribution to the dipole from the permanent electrostatics; thus no additional pseudo interaction sites carrying only charges and/or multipoles are introduced in AMOEBA. In this fashion, the total dipole moment for molecule J is calculated by summing over all multipolar contributions of all atoms $I = 1, \dots, N_J$ in that molecule. The effective molecular and thus also the total dipole moments obtained this way from AIMD and AMOEBA for solvent and solute species are comparable as already demonstrated in the validation section (cf. Fig. 3).

B. Decomposing THz Spectra

In order to understand the signals in the total THz spectra $\alpha(\omega)$ at the molecular level, decomposition of the THz observable in terms of atomic motions is necessary. Our mode decomposition as developed in Refs. 14,61–63 to analyze infrared spectra of floppy molecules in the gas phase and extended to dissect THz spectra of aqueous solutions in Ref. 12 leads to mode specific lineshape functions and thus absorption cross sections $\alpha_k(\omega)$ that allow for an understanding of the spectrum in terms of explicit molecular displacements. In the following, we provide only a concise exposition of the key ideas and refer the interested reader to a review¹⁴ and to the SI of Ref. 12 for comprehensive theoretical background including the treatment of nuclear quantum effects in the realm of theoretical infrared spectroscopy and technical details of the present approach, respectively.

Our particular computational approach¹² to decompose total infrared spectra $\alpha(\omega)$ into dynamical modes k and associated lineshape functions $\alpha_k(\omega)$ has been formulated specifically for analyzing AIMD trajectories including the electronic structure based on the charge current cross-correlation matrix that involves all charge centers i in the system,

$$C_{\zeta,\xi}(\omega) = \int_{-\infty}^{+\infty} dt \langle \dot{\mu}_{i,\zeta}(0) \cdot \dot{\mu}_{j,\xi}(t) \rangle e^{-i\omega t} \quad (15)$$

$$= \int_{-\infty}^{+\infty} dt \langle q_i v_{i,\zeta}(0) \cdot q_j v_{j,\xi}(t) \rangle e^{-i\omega t}, \quad (16)$$

as expressed in terms of charge-weighted velocities, where charges q_i weight the velocity of charge center i at position $\mathbf{r}_i(t)$; here cartesian velocity components ζ and ξ are used that have been rotated into a molecular frame of reference. Summation over all charge centers in the system thus not only includes the contributions due to the motion of the (positive) nuclei, but also the electron dynamics as represented by the motion of the Wannier centers being the (negative) electronic charge centers that are obtained from the maximally localized Wannier valence molecular orbitals. This charge-weighting obviously provides the required dipolar (cross-) correlations in terms of the time-derivative of the dipole moment vectors of all charge centers, $\{q_i \dot{\mathbf{r}}_i(t)\}$, thus including also the full electronic contribution to the charge current. Importantly, the corresponding sum over all nuclear and electronic charge centers within a specific molecule J at time t leads to its total charge current $\dot{\boldsymbol{\mu}}_{J,\text{aimd}}(t)$ corresponding to Eq. 13, which is finally required according to Eq. 12 in order to compute the total infrared spectrum $\alpha(\omega)$ from Eq. 10 *via* Eq. 11 where the resulting total charge current $\dot{\mathbf{M}}(t)$ enters.

We note in passing that a decomposition approach which neglects the explicit electronic contributions altogether and cross-correlates the $\sqrt{\text{mass}}$ -weighted atom velocities instead of the dipole velocities introduces a generalization of the vibrational density of states (VDOS)⁹

and thus provides access to its decomposition in terms of modes (something that is accessible experimentally via inelastic neutron scattering). This procedure obviously does not provide infrared intensities, and thus no access to THz spectra¹⁰, yet the same mode decomposition as performed here for the dipole correlations yields very similar mode displacement patterns of the atoms in real space as explicitly demonstrated for the present example in the SI.

It is key to observe that the cross-correlation matrix as defined *via* Eq. 15 not only includes the particle velocities with the associated core charges (and thus the contribution of the molecular skeleton like in non-polarizable force field simulations), but in particular also the velocities of the Wannier centers which represent the electron dynamics in AIMD simulations within the Born-Oppenheimer approximation⁸. By taking into account the Wannier orbital dynamics in that sense, the purely electronic contributions to infrared absorption spectra, such as polarization and charge transfer effects, are included in the computation of $\alpha(\omega)$ based on AIMD trajectories.

At this stage, the mode-specific absorption cross sections (or mode spectra) $\alpha_k(\omega)$ are obtained after diagonalization of $\mathbf{C}(\omega)$, where the off-diagonal rest term α_{cross} is a measure of the remaining cross-correlations. The dipole displacement vectors corresponding to the k th mode can be determined from the transformation matrix that approximately diagonalizes the cross-correlation matrix (as explained in the SI of Ref. 12), which is close in spirit to the atomic displacement vectors that are obtained in traditional normal mode analysis. Finally, the total absorption cross section,

$$\alpha(\omega) = \sum_k \alpha_k(\omega) + \alpha_{\text{cross}}(\omega), \quad (17)$$

can be recovered by summing over all decoupled modes k after adding the remaining cross terms. Most importantly, this provides one with a systematic tool to probe, mode by mode, how the lineshape of the total THz spectrum is generated by considering selected subsets of modes.

On the other hand, AMOEBA uses a completely different approach for calculating the molecular dipoles according to Eq. 14. As a result, the aforementioned computational approach and in particular Eq. 15 cannot be applied directly to the AMOEBA data, since the polarization contributions to the modes is located at the atom centers and thus are coupled directly into the molecular motion itself, whereas they are represented explicitly by the Wannier center dynamics in AIMD. Thus, in order to decouple the polarization modes (arising mostly from solute-water interactions) from the intramolecular atomic displacements solely for the purpose of spectra calculations, we need to introduce charged pseudo-sites in order to capture the polarization contributions separately as explained in the following.

For a water molecule J in aqueous solution, we can do this in a straightforward way by computing one effective

charge center or pseudo-site, $\tilde{\mathbf{r}}_J$, which exclusively carries a charge $\tilde{q}_J$ in order to approximately capture the polarization contributions *via*

$$\tilde{q}_J \tilde{\mathbf{r}}_J(t) = \boldsymbol{\mu}_{J,\text{amoeba}}(t) - \sum_{I=1}^{N_J} q_I \mathbf{r}_I(t) , \quad (18)$$

where N_J is the number of atoms in molecule J , q_I is the charge of atom I at position $\mathbf{r}_I(t)$ at time t and $\boldsymbol{\mu}_{J,\text{amoeba}}(t)$ is the total dipole moment of water molecule J in solution as given by the full AMOEBA force field including the instantaneous polarization effects at time t according to Eq. 14. In order to conform as closely as possible with the AIMD approach to spectral decomposition, the atom charges q_I have been chosen to be identical to the nuclear core charges which underly the pseudo potential representation of the electronic structure in AIMD (i.e. $q_O = +6|e|$ and $q_H = +1|e|$ in case of oxygen and hydrogen atoms). Thus, the effective charge $\tilde{q}_J$ attached to the pseudo-site at position $\tilde{\mathbf{r}}_J$ is identical to the full valence charge of the molecule as given by the sum of all its Wannier charges, which is $-8|e|$ in case of a neutral water molecule.

Following this procedure, the position of the pseudo charge site, $\tilde{\mathbf{r}}_J(t)$, can be uniquely computed from Eq. 18 once the total dipole moment of water molecule J at time t is determined in the AMOEBA simulation of the solution for the corresponding molecular configuration $\{\mathbf{r}_I(t)\}$ of that (polarized) water molecule. Based on this approach, the total dipole moment of water molecule J in solution is represented by the following expression in AMOEBA,

$$\boldsymbol{\mu}_{J,\text{amoeba}}(t) = \sum_{I=1}^{N_J} q_I \mathbf{r}_I(t) + \tilde{q}_J \tilde{\mathbf{r}}_J(t) , \quad (19)$$

which is solely used in that form when evaluating the dipolar cross-correlations according to Eq. 15 with the aim to dissect the total infrared spectrum of the solution, $\alpha(\omega)$, in terms of modes k and the corresponding mode-specific absorption cross sections $\alpha_k(\omega)$. Thus, a single water molecule in AIMD is composed of three nuclear and four electronic (Wannier) charge centers, the latter representing the eight paired valence electrons, whereas it is approximated in AMOEBA by the same three nuclear charge centers at the atom positions together with one effective electronic pseudo charge center that carries the full valence charge.

Since a single pseudo charge center would be a rather crude approximation for molecules much larger than water, we adopted a fully additive “divide and conquer” approach by introducing one such pseudo-site $\tilde{q}_J$ for each functional group. In case of the two amino acids, the following such fragments have been defined: the protonated amino NH_3^+ and deprotonated carboxyl COO^- groups, the side chain groups for glycine (H) and valine ($\text{CH}(\text{CH}_3)_2$), as well as the $\text{C}_\alpha\text{H}_\alpha$ group. Each effective center is computed exactly in the way described above

for a single water molecule while only taking into account all those atoms (including the hydrogens) that belong to the respective functional group. Upon representing a molecule additively by a sum of atoms requires to cut covalent bonds that connect these functional groups, which is done here in the crudest way by dividing up the respective electron pairs democratically between the two functional groups. This results also in the correct net charge in case of charged groups, for instance NH_3^+ consists of one N and three H atom sites (thus providing a total nuclear charge of $+8|e|$ in view of the nuclear core charges $q_N = +5|e|$ and $q_H = +1|e|$) and seven electrons (three electron pairs $-2|e|$ from the N-H bonds and one electron $-1|e|$ from the cut C-N bond) and thus a pseudo charge of $\tilde{q}_{\text{NH}_3^+} = -7|e|$, which leads to a net charge of $+1|e|$ as required. At this stage, the same spectral analysis machinery as developed for AIMD can be carried over to analyze the AMOEBA trajectories, which also carries over to much larger systems such a peptides by virtue of the additive fragmentation approach in terms of well-defined functional groups.

Finally, the THz modes, in particular those that couple solute and hydrogen-bonded solvent molecules, are obtained by employing the supermolecular solvation complex (SSC) approach as introduced in Ref. 12. The SSC is composed of the solute molecule, in this case glycine or valine, and either three water molecules at the amino group (which is denoted as SSC(+)) or one water molecule at the carboxylate group (SSC(-)) as assessed in the SI of Ref. 12. Employing the SSC(+/-) analysis enables us to compute modes that take into account the coupled motions of the solute with the solvation water molecules at the hydrophilic sites.

IV. RESULTS

A. THz Spectra: AMOEBA versus AIMD

Overall the new set of AMOEBA parameters provide properties of aqueous solutions of amino acids that are in reasonable agreement with AIMD, and undoubtedly are an improvement over the original set of parameters. Therefore we can now proceed to compare the theoretical THz spectra resulting from AIMD and AMOEBA simulations for glycine and valine, and the cross-correlation spectra of each amino acid with their respective water environment. We will then compare the mode-specific absorptions computed *via* the mode decomposition scheme described above. Since the zwitterionic amino acid modes are found to be very similar in most cases, glycine will be discussed in detail and any differences with valine will be shown when applicable.

The total spectrum of glycine computed from the AMOEBA simulations shows qualitative agreement with the one obtained from AIMD simulations as seen in Fig. 4. An absorption peak is seen at 80 cm^{-1} where the low frequency rigid body motions are located, as

Figure 4. Total THz spectrum $\alpha(\omega)$ of glycine and valine based on AMOEBA (blue) and AIMD (red) simulations obtained within the SSC(+) approach. The three signals can be assigned to intramolecular bending modes (at ~ 300 cm^{-1}), rigid body motions (~ 80 cm^{-1}) and intermolecular solute-solvent coupling modes (~ 200 cm^{-1}).

Figure 5. Cross-correlation spectrum of glycine with surrounding water molecules based on AMOEBA (blue) and AIMD (red) simulations.

well as the most characteristic absorption, the NCCO open/close mode at 300 cm^{-1} , which is much sharper due to the more harmonic nature of the corresponding intramolecular motion according to the AMOEBA force field. The intermolecular absorption signal, originating from the interaction with the solvating water molecules due to hydrogen-bonded stretching is present at 200 cm^{-1} as shown in Fig. 5. It originates from polarization effects since fixed charge force fields do not exhibit this cross-correlation signal. From this comparison, however, it is clear that not all of the cross-correlation is necessarily captured in view of some missing intensity. This could arise due to the need to conform to using charged-weighted velocities to represent the dipoles *via* Eq. 15 and the sum rule that recovers the total spectrum from Eq. 17. This computational approach has been shown to work well for AIMD but is a less natural definition for the AMOEBA force field where it requires the introduc-

tion of charged pseudo-sites *via* Eq. 18 to approximately capture polarization contributions. Another possibility is that the solute-solvent mode is comprised of more than just pure polarization, such that the remaining missing intensity might be attributed to lack of charge transfer in the AMOEBA model while it is captured by AIMD, since simulated infrared spectra are known to be sensitive to this molecular interaction. Valine shows very similar behavior to that observed for glycine (see SI material). Nonetheless, while the AMOEBA intensities are smaller and frequencies are slightly shifted, the principal lineshapes follow the trends observed in AIMD.

B. THz Modes and Spectral Decomposition

From our previous AIMD analysis¹² we have found intramolecular motions of the amino acid itself (e.g. opening and closing of NCCO, twist around the CC bond), quasi rigid body motions that describe the hindered translations of the molecule within the water environment (rattling) as well as hindered water rotations (librations) and water stretching and bending motions that describe intermolecular interactions of water with the amino acid directly. In Fig. 6 representative examples of the glycine modes are visualized in terms of the displacement vectors, with a similar set found for valine which also includes additional rotameric motions of the aliphatic side chain. Therefore, in order to further compare the AMOEBA and AIMD calculations, we decompose the AMOEBA spectrum by assigning each band a molecular displacement using the SSC(+/-) approach.

Comparing the resulting modes obtained from the AMOEBA simulation and from AIMD, we see very good agreement for the glycine modes shown in Fig. 7 and for the valine modes in Fig. 8. From visual inspection it is evident that the intramolecular modes at the high frequency end of the THz spectrum, Fig. 6(a-c), are very similar and show basically identical displacement patterns between AMOEBA and AIMD. Glycine and valine both show the characteristic NCCO open/close mode above 300 cm^{-1} (305 cm^{-1} for glycine and 352 cm^{-1} for valine); while this mode is broad in the AIMD case, it is a sharper mode in glycine due to the harmonic nature of the corresponding force field term in AMOEBA. Valine shows additional intramolecular modes (at 335 cm^{-1} and 321 cm^{-1}) involving the side chain rotomers, although they are slightly red-shifted compared to AIMD (317 cm^{-1} and 281 cm^{-1} , respectively).

The AIMD study revealed that the CC-twisting mode shows a strong coupling to the water hydrogen-bonded network in the first solvation shell, as shown by the coupled motion of the twisting atoms together with the hydrogen-bond stretching of the water molecules. Furthermore, this mode dominates the 200 cm^{-1} signal that is associated with the solute-solvent coupling. Both of these key observations are also true for the same mode

Figure 6. Glycine mode displacement vectors for AMOEBA(above) and AIMD(below) obtained with the SSC(+) approach: NCCO open/close, C-C twist coupled to hydrogen-bond stretch, C_α out-of-plane, quasi rigid body cage rattling I, and quasi rigid body cage libration I. The corresponding mode-specific THz spectra are shown in Fig. 7.

Figure 7. Mode-specific THz absorption spectra $\alpha_k(\omega)$ of glycine based on AIMD and AMOEBA simulation data obtained within the SSC(+) approach. Only the THz modes with intensity greater than one wavenumber are shown.

obtained from the AMOEBA simulation. In direct comparison to the AIMD mode (Fig. 7) the intensity of the mode is reduced by roughly half, while the displacement is very similar (Fig. 6b). This is in agreement with the overall lower cross-correlation signal of glycine with water (Fig. 5) with reduced intensity that could stem from the pseudo charges that are adopted from the AIMD result. The modes derived by AMOEBA for valine show very similar trends, but additional intramolecular modes are obtained due to the side chain, which are in good accord between the AIMD results and the AMOEBA force field as well.

The rattling modes due to hindered translations as observed in the AIMD simulation show a concerted uni-

Figure 8. Mode-specific THz absorption spectra $\alpha_k(\omega)$ of valine based on AIMD and AMOEBA simulation data obtained within the SSC(+) approach. Only the THz modes with intensity greater than one wavenumber are shown

directional motion of all amino acid atoms and water molecules (within the SSC(+/-) approaches), whereas the AMOEBA modes show a more disorganized motion for both glycine and valine. It is a systematic problem that could arise from either the method of introducing the additional centers to separate out the polarization modes, the charges taken from AIMD to define the pseudo charge center, or the result of the lack of charge transfer in the AMOEBA force field itself that manifests mostly in the solute-solvent interactions.

C. Assessing Finite Size Effects

Simulations with the polarizable AMOEBA model allow for much larger system sizes than the AIMD simulations. Since the spectrum and the modes obtained from AMOEBA agree well in general to AIMD data, increased system sizes were investigated with AMOEBA in order to probe finite size effects. Tables I and II show the peak frequencies of the modes obtained from the small and large simulation boxes. This analysis is based on the SSC(+) approach thus including the first solvation shell of the protonated amino group. Only minor frequency shifts are observed which are attributed to the broad overall peak shape since only the maximum is reported. After inspection of the mode displacement assignments it is clear that all modes of the small and large simulation boxes agree with each other, see SI for visual inspection. We conclude that although the simulation box sizes used in AIMD appear to be small, the THz spectra and their interpretation in terms of the mode displacement vectors according to the SSC(+/-) approaches do not suffer from finite size effects at the required level of accuracy.

Table I. Peak positions of mode-specific absorption spectra of glycine in water obtained within the SSC(+) approach with significant contribution to the THz spectrum (see Fig. 7) depending on system size.

Mode	AMOEBA 30 Wat	AMOEBA 253 Wat	AIMD 30 Wat
NCCO open/close	305	306	304
C-C twist + HB stretch	236	237	247
HB stretch I	214	177	210
HB stretch II	146	149	218
C $_{\alpha}$ out-of-plane	135	137	125
HB bend I	92	96	102
cage libration III	84	87	89
cage libration II	71	74	90
HB bend II	68	73	90
cage libration I	68	66	82
cage rattling II	63	64	73
cage rattling III	61	29	64
cage rattling I	50	25	62

Table II. Peak positions of mode-specific absorption spectra of valine in water obtained within the SSC(+) approach with significant contribution to the THz spectrum (see Fig. 8) depending on system size.

Mode	AMOEBA 60 Wat	AMOEBA 256 Wat	AIMD 60 Wat
NCCO open/close	352	353	349
NCCC I	335	335	317
NCCC II	321	320	280
COO-CCC	236	239	197
C-C twist + HB stretch	196	201	189
HB stretch I	131	140	112
HB bend I	86	95	101
R libration I	75	79	85
HB bend-X3	63	76	83
cage libration I	52	57	68
Cage rattling	51	52	56
R libration II	38	40	57

V. DISCUSSION AND CONCLUSIONS

Overall we conclude that our reparameterized AMOEBA polarizable model provides qualitative, and in some instances quantitative agreement with the AIMD reference results for the THz spectra that report on the solvation dynamics of small zwitterionic amino acids in aqueous solution. Still, this polarizable force field leads to overstructured total THz absorption spectra of glycine and valine in water as judged by comparing to both AIMD and experimental results. At the level of the computed spectra, the cross-correlations between solute and solvent molecules seem to be less pronounced than in the AIMD simulations. This effect is particularly evident in the intensity of those mode-specific THz spectra that are dominated by strong couplings of the solute to the water hydrogen-bond network such as in case of the C-C twisting modes of both glycine and valine that are located in the 200 cm^{-1} region. The hindered translational motion, giving rise to cage rattling modes in THz spectra, is less

clearly pronounced according to the analysis of the polarizable force field data when gauged with the AIMD mode displacement patterns. Despite such caveats at a detailed level of assessment, the overall AMOEBA performance bodes well for future THz studies on larger systems such as enzymes in water or extended aqueous interfaces with a focus on qualitative insights and trends given that the computational cost is much lower in comparison to AIMD simulations.

To arrive at these encouraging results, we have presented a straightforward approach to THz spectral analysis using the AMOEBA model by engineering additional charged pseudo-sites in a manner that allows the decomposition of the spectrum into mode-specific absorption coefficients as was done earlier in the AIMD simulations. Importantly, this scheme relies on an additive “divide and conquer” idea based on functional groups or molecular fragments that can be readily transferred to much more complex molecular systems such as proteins or lipid membranes. Our approach to the calculation of THz spectra from AMOEBA approximately includes electronic polarization effects which are known to play a key role in determining the correct intensity modulations as a function of frequency and thus the overall lineshape function. Nonetheless further improvement in the methodology for decomposing the THz spectra is warranted for the polarizable force field since the computational approach of decomposing the total absorption spectrum into mode-specific cross sections based on the full charge current auto-correlation function as devised for AIMD simulations (where direct access to localized molecular orbitals via the Wannier centers and thus the full charge current are readily available) is ill-suited for the AMOEBA model (where no such purely electronic information is straightforwardly accessible). Since AMOEBA uses a point multipole representation of the permanent electrostatics and polarization contributions that are atom centered, this introduces both distortions of the modes and/or higher off-diagonal terms in the spectral decomposition than observed in AIMD. One future modification of the approach would replace the weighting of the modes by formal charges from AIMD for one which fully takes advantage of the AMOEBA point monopole, dipole, and induced dipoles in the weighting scheme also at the level of analysis.

Another source of future investigation is to better understand the limitations introduced by the lack of charge transfer in the AMOEBA model. Since charge transfer is certainly present in the AIMD calculations, and contributes to the intensity of the cross-correlations between the water molecules and the amino acid solutes, it is expected to play a significant role in the intensities of the modes⁶⁴. Therefore introducing charge transfer into the AMOEBA force field will further improve the capabilities of describing the THz spectrum in solutions, and its decomposition into assigned modes of the dynamics, and will be the subject of future efforts.

Finally, the AIMD results were computed using the

PBE density functional, which has both well established strengths and weaknesses, but has been well-validated against experimental THz spectra of glycine and valine aqueous solutions and neutron diffraction data of the solvation shell structure of glycine in water. Nonetheless there has been an explosion of new density functionals that show demonstrable improvements in properties ranging from binding and isomerization energies, barrier heights, through to thermochemistry that should be considered in future validation studies. Unfortunately, reliable calculations of THz spectra are computationally much more demanding than computing radial distribution functions or alike since on the order of 100 independent microcanonical AIMD runs based on uncorrelated initial conditions sampled from a long canonical AIMD simulation are typically required in order to converge lineshapes at such low frequencies and thus their subtle modulations which encode the desired molecular information after mode-specific decompositions. Furthermore, while raising the temperature or using quantum thermostats can be an *ad hoc* way to roughly emulate missing nuclear quantum effects, quantum delocalization contributions are difficult to predict *a priori* since they can also give rise to strengthening hydrogen-bonding that counteracts the effect of raising temperature.

ACKNOWLEDGMENTS

This work results from Ph.D. internships of AE and SB at UCB and RUB, respectively, that have been carried out and funded in the framework of the RESOLV Graduate School “Solvation Science” (GSS) and CalSolv, respectively. THG and SB were supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The work of AE and DM was partially supported by the Deutsche Forschungsgemeinschaft *via* MA 1547/11 and is also part of the Cluster of Excellence RESOLV (EXC 1069). Computational resources were provided by SuperMUC@LRZ, HPC-RESOLV, HPC@ZEMOS, BOVILAB@RUB as well as RV-NRW. We would like to thank Professor Fabio Bruni, (University of Roma Tre) for kindly providing experimental partial RDFs of glycine in water and Dr. Harald Forbert (Ruhr-Universität Bochum) for useful discussions on the spectral decomposition technique.

REFERENCES

- ¹U. Heugen, G. Schwaab, E. Bründermann, M. Heyden, X. Yu, D. M. Leitner, and M. Havenith, *Proc. Nat. Acad. Sci.* **103**, 12301 (2006).
- ²C. Zhang and S. M. Durbin, *J. Phys. Chem. B* **110**, 23607 (2006).
- ³I. A. Heisler and S. R. Meech, *Science* **327**, 857 (2010).
- ⁴M. Heyden and M. Havenith, *Methods* **52**, 74 (2010).
- ⁵K. J. Tielrooij, N. Garcia-Araez, M. Bonn, and H. J. Bakker, *Science* **328**, 1006 (2010).
- ⁶K. J. Tielrooij, S. T. van der Post, J. Hunger, M. Bonn, and H. J. Bakker, *J. Phys. Chem. B* **115**, 12638 (2011).
- ⁷M. Grossman, B. Born, M. Heyden, D. Tworowski, G. B. Fields, I. Sagi, and M. Havenith, *Nat. Struct. Mol. Biol.* **18**, 1102 (2011).
- ⁸D. Marx and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods* (Cambridge University Press, Cambridge, 2009).
- ⁹M. Heyden, J. Sun, S. Funkner, G. Mathias, H. Forbert, M. Havenith, and D. Marx, *Proc. Nat. Acad. Sci.* **107**, 12068 (2010).
- ¹⁰M. Heyden, J. Sun, H. Forbert, G. Mathias, M. Havenith, and D. Marx, *J. Phys. Chem. Lett.* **3**, 2135 (2012).
- ¹¹M. Śmiechowski, H. Forbert, and D. Marx, *J. Chem. Phys.* **139**, 014506 (2013).
- ¹²J. Sun, G. Niehues, H. Forbert, D. Decka, G. Schwaab, D. Marx, and M. Havenith, *J. Am. Chem. Soc.* **136**, 5031 (2014).
- ¹³M. Śmiechowski, J. Sun, H. Forbert, and D. Marx, *Phys. Chem. Chem. Phys.* **17**, 8323 (2015).
- ¹⁴S. D. Ivanov, A. Witt, and D. Marx, *Phys. Chem. Chem. Phys.* **15**, 10270 (2013).
- ¹⁵M. Sprik and M. L. Klein, *J. Chem. Phys.* **89**, 7556 (1988).
- ¹⁶D. Van Belle, M. Froeyen, G. Lippens, and S. J. Wodak, *Mol. Phys.* **77**, 239 (1992).
- ¹⁷J. W. Ponder and D. A. Case, *Adv. Prot. Chem.* **66**, 27 (2003).
- ¹⁸G. Lamoureux, A. D. MaxKerell Jr., and B. Roux, *J. Chem. Phys.* **119**, 5185 (2003).
- ¹⁹P. Ren and J. P. Ponder, *J. Phys. Chem. B* **107**, 5933 (2003).
- ²⁰G. A. Kaminski, H. A. Stern, B. J. Berne, and R. A. Friesner, *J. Phys. Chem. A* **108**, 621 (2004).
- ²¹S. Patel and C. L. I. Brookes, *J. Comput. Chem.* **25**, 1 (2004).
- ²²E. Harder, V. M. Anisimov, I. V. Vorobyov, P. E. M. Lopes, S. Y. Noskov, A. D. MacKerell Jr., and B. Roux, *J. Chem. Theory Comput.* **2**, 1587 (2006).
- ²³J.-P. Piquemal, H. Chevreau, and N. Gresh, *J. Chem. Theory Comput.* **3**, 824 (2007).
- ²⁴J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio Jr., M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon, *J. Phys. Chem. B* **114**, 2549 (2010).
- ²⁵R. Kumar, F.-F. Wang, G. R. Jenness, and K. D. Jordan, *J. Chem. Phys.* **132**, 014309 (2010).
- ²⁶P. E. M. Lopes, J. Huang, J. Shim, Y. Luo, H. Li, B. Roux, and A. D. MaxKerell Jr., *J. Chem. Theo. Comput.* **9**, 5430 (2013).
- ²⁷O. N. Demerdash, E. H. Yap, and T. Head-Gordon, *Annu. Rev. Phys. Chem.* **65**, 149 (2014).
- ²⁸A. Albaugh, H. A. Boateng, R. T. Bradshaw, O. N. Demerdash, J. Dziedzic, Y. Mao, D. T. Margul, J. Swails, Q. Zeng, D. A. Case, P. Eastman, J. W. Essex, M. Head-Gordon, V. S. Pande, J. W. Ponder, Y. Shao, C.-K. Skylaris, I. T. Todorov, M. E. Tuckerman, and T. Head-Gordon, *J. Phys. Chem. B* **in press** (2016).
- ²⁹P. Ren, C. Wu, and J. W. Ponder, *J. Chem. Theo. Comp.* **7**, 3143 (2011).
- ³⁰L.-P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande, *J. Phys. Chem. B* **117**, 9956 (2013).
- ³¹M. L. Laury, W. Lee-Ping, V. Pande, T. Head-Gordon, and J. W. Ponder, *J. Phys. Chem. B* **119**, 9423 (2015).
- ³²M. Śmiechowski, C. Schran, H. Forbert, and D. Marx, *Phys. Rev. Lett.* **116**, 027801 (2016).
- ³³P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, H. Basch, and D. Cohen, *J. Chem. Phys.* **105**, 1968 (1996).
- ³⁴C. Millot, J.-C. Soetens, M. T. C. Martins Costa, M. P. Hodges, and A. J. Stone, *J. Phys. Chem. A* **102**, 754 (1998).
- ³⁵M. A. Freitag, M. S. Gordon, J. H. Jensen, and W. J. Stevens, *J. Chem. Phys.* **112**, 7300 (2000).
- ³⁶J.-P. Piquemal, N. Gresh, and C. Giessner-Prettre, *J. Phys. Chem. A* **107**, 10353 (2003).
- ³⁷R. Chelli, M. Pagliai, P. Procacci, and G. Cardini, *J. Chem.*

- Phys. **122**, 074504 (2005).
- ³⁸L. Slipchenko and M. S. Gordon, J. Comput. Chem. **28**, 276 (2007).
- ³⁹A. J. Stone, J. Phys. Chem. A **115**, 7017 (2011).
- ⁴⁰B. Wang and D. G. Truhlar, J. Chem. Theory Comput. **10**, 4480 (2014).
- ⁴¹Q. Wang, J. A. Rackers, C. He, R. Qi, C. Narth, L. Lagardere, N. Gresh, J. W. Ponder, J.-P. Piquemal, and P. Ren, J. Chem. Theo. Comp. **11**, 2609 (2015).
- ⁴²C. Narth, L. Lagardere, E. Polack, N. Gresh, Q. Wang, D. R. Bell, J. A. Rackers, J. W. Ponder, P. Y. Ren, and J.-P. Piquemal, J. Comp. Chem. **37**, 494 (2016).
- ⁴³A. J. Lee and S. W. Rick, J. Chem. Phys. **134**, 184507 (2011).
- ⁴⁴M. Soniat and S. W. Rick, J. Chem. Phys. **137**, 044511 (2012).
- ⁴⁵M. Soniat and S. W. Rick, J. Chem. Phys. **140**, 184703 (2014).
- ⁴⁶M. Soniat and S. W. Rick, J. Chem. Theo. Comp. **11**, 1658 (2015).
- ⁴⁷M. Soniat, R. Kumar, and S. W. Rick, J. Chem. Phys. **143**, 044702 (2015).
- ⁴⁸M. Soniat, G. Pool, L. Franklin, and S. W. Rick, Fluid Phase Equilib. **407**, 31 (2016).
- ⁴⁹A. J. Stone, *The Theory of Intermolecular Forces* (Oxford University Press, 1996).
- ⁵⁰B. T. Thole, Chem. Phys. **59**, 341 (1981).
- ⁵¹T. A. Halgren, J. Am. Chem. Soc. **114**, 7827 (1992).
- ⁵²Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, J. Chem. Theo. Comp. **9**, 4046 (2013).
- ⁵³L.-P. Wang, J. Chen, and T. Van Voorhis, J. Chem. Theo. Comp. **9**, 452 (2013).
- ⁵⁴Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuš, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. Hanson-Heine, P. H. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyayev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. Gill, and M. Head-Gordon, Mol. Phys. **113**, 184 (2015).
- ⁵⁵A. J. Stone, J. Chem. Theo. Comp. **1**, 1128 (2005).
- ⁵⁶M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian Inc. Wallingford CT 2009.
- ⁵⁷E. Scoppola, A. Sodo, S. E. McLain, M. A. Ricci, and F. Bruni, Biophys. J. **106**, 1701 (2014).
- ⁵⁸J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, Comp. Phys. Comm. **167**, 103 (2005).
- ⁵⁹J. Sun, D. Bousquet, H. Forbert, and D. Marx, J. Chem. Phys. **133**, 114508 (2012).
- ⁶⁰R. Ramírez, T. López-Ciudad, P. Kumar P, and D. Marx, J. Chem. Phys. **121**, 3973 (2004).
- ⁶¹S. D. Ivanov, O. Asvany, A. Witt, E. Hugo, G. Mathias, B. Redlich, D. Marx, and S. Schlemmer, Nat. Chem. **2**, 298 (2010).
- ⁶²G. Mathias and M. D. Baer, J. Chem. Theo. Comp. **7**, 2028 (2011).
- ⁶³G. Mathias, S. D. Ivanov, A. Witt, M. D. Baer, and D. Marx, J. Chem. Theo. Comp. **8**, 224 (2012).
- ⁶⁴E. Ramos-Cordoba, D. S. Lambrecht, and M. Head-Gordon, Farad. Discuss. **150**, 345 (2011).