JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Anharmonic Infrared Spectroscopy through the Fourier Transform of Time Correlation Function Formalism in ONETEP

Valerio Vitale,[†,‡] Jacek Dziedzic,[†,¶] Simon M.-M. Dubois,[§] Hans Fangohr,[∥,⊥] and Chris-Kriton Skylaris*[,†]

[†]School of Chemistry, University of Southampton, Southampton, United Kingdom

[‡]Institute of Complex Systems Simulation, University of Southampton, Southampton, United Kingdom

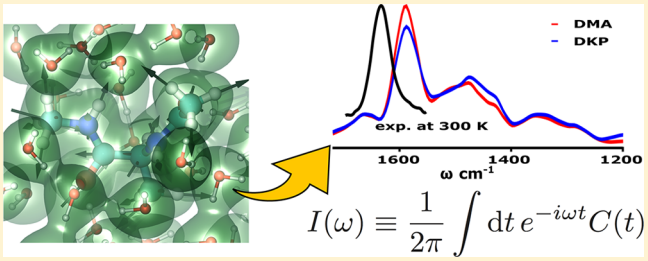
[¶]Faculty of Applied Physics and Mathematics, Gdańsk University of Technology, Gdańsk, Poland

[§]Institut de la Matière Condenseée et des Nanosciences, Universitè Catholique de Louvain B-1348 Louvain-la-Neuve, Belgium

[∥]Engineering and the Environment, University of Southampton, Southampton, United Kingdom

[⊥]Institute of Complex Systems Simulation, University of Southampton, Southampton, United Kingdom

ABSTRACT: Density functional theory molecular dynamics (DFT-MD) provides an efficient framework for accurately computing several types of spectra. The major benefit of DFT-MD approaches lies in the ability to naturally take into account the effects of temperature and anharmonicity, without having to introduce any *ad hoc* or *a posteriori* corrections. Consequently, computational spectroscopy based on DFT-MD approaches plays a pivotal role in the understanding and assignment of experimental peaks and bands at finite temperature, particularly in the case of floppy molecules. Linear-scaling DFT methods can be used to study large and complex systems, such as peptides, DNA strands, amorphous solids, and molecules in solution. Here, we present the implementation of DFT-MD IR spectroscopy in the ONETEP linear-scaling code. In addition, two methods for partitioning the dipole moment within the ONETEP framework are presented. Dipole moment partitioning allows us to compute spectra of molecules in solution, which fully include the effects of the solvent, while at the same time removing the solvent contribution from the spectra.



1. INTRODUCTION

Time-correlation functions (TCFs) have been extensively used to unveil the dynamical and structural properties of molecular systems.[1,2] Unlike in the past, where TCFs were derived from experimental spectra, currently, we are able to generate TCFs from molecular dynamics simulations.

Molecular mechanics (MM) simulations[3−6]—and, more recently, *ab initio* molecular dynamics (AIMD) simulations[7−10]—have been proven to be a powerful tool in achieving this goal. However, in MM simulations, the interatomic potential is defined in terms of a parametrized force field, where parameters are obtained from experimental works or from high-level quantum-mechanical calculations. Conversely, AIMD simulations do not rely on any external parameter, and forces are computed "on-the-fly" directly from first-principles electronic structure calculations. This endows AIMD simulations with a higher predictive power over MM simulations, especially when dealing with solvent interactions, hydrogen bonding and polarizability, where the MM description is poor or, in some force fields, missing altogether. Consequently, AIMD simulations have been successfully used to study either systems that are inherently dynamical, such as in conformational analysis of (bio)molecules at finite temperatures, chemical reaction paths in gas and condensed phase, surface (adsorption) processes, and solvation phenomena, or as an alternative method to energy minimization techniques, for the search of equilibrium structures of molecules, clusters, and surfaces under different pressure and temperature conditions. Therefore, from AIMD simulations, it is possible to construct many TCFs for the calculation of several types of spectra—in particular, vibrational, (mid)IR, and Raman spectra.

Identifying the structure of biomolecules is a very relevant matter, since it is widely accepted that the function of biological molecules is deeply related to the three-dimensional domains (central dogma of structural biology).[11] Particularly, the understanding of the IR signatures of the gas phase can be help to interpret the less-obvious IR signatures in the condensed phase, and the effect of vibrational couplings with the solvent. However, in more realistic systems with a very large number of atoms, such as in drug optimization applications or other biomolecular problems, some approximations must be made in the model. For example, when the role of the solvent is crucial in determining the properties of the system, implicit solvent models can be used. In ONETEP,[12] the implicit solvation model is based on a direct solution of the nonhomogeneous Poisson equation in real space, and the implementation is such that the linear-scaling regime is maintained.[13] This has been

Received: April 24, 2015
Published: June 10, 2015

DOI: 10.1021/acs.jctc.5b00391
J. Chem. Theory Comput. 2015, 11, 3321−3332

successfully used to study protein−ligand complexes,[14] as well as the effect of nanoparticles on the binding of peptides in a multiscale simulation.[15]

2. TIME-CORRELATION FUNCTION FORMALISM

Lineshape computation of vibrational, infrared, and Raman spectra relies on the assumption that the external field used to probe the system is weakly coupled to the system itself. The energy dissipated in the interaction is of the same magnitude of the energy of a thermal fluctuation. Under this assumption, it is legitimate to use the results of the linear response theory and time correlation functions of observables of the unperturbed system. The aim of the next two sections is to gather known results and present them in a concise but self-contained fashion. In deriving the formulas of interest, we will mainly follow the works of McQuarrie[1] and Tuckerman.[16]

Let $\hat{\mathcal{H}}_0$ be the unperturbed time-independent Hamiltonian of the ground state, for an N-body system, i.e. $\hat{\mathcal{H}}_0|j\rangle = E_j|j\rangle$. If the system now interacts weakly with an external electric field of frequency ω, transitions from the initial states $|i\rangle$ into other quantum states $|f\rangle$ (for final) will occur if the frequency of the radiation is close to $(E_f - E_i)/\hbar$. For a monochromatic field, we can write

$$\mathbf{E}(t) = \varepsilon E_0 \cos(\omega t) \tag{1}$$

where ε is the unit vector that defines the direction of the incident radiation. Assuming a uniform field or a field with a wavelength much larger than molecular dimensions, the interaction between the field and the system can be written as

$$\hat{\mathcal{H}}_1(t) = -\hat{\mu}\cdot\mathbf{E}(t) \tag{2}$$

where $\hat{\mu}$ is the dipole moment operator. Hence, the new Hamiltonian can be written as $\hat{\mathcal{H}}(t) = \hat{\mathcal{H}}_0 + \hat{\mathcal{H}}_1(t)$, where the explicit time dependence is due to the oscillatory external field, and not to the operator itself (Schrödinger picture). According to time-dependent quantum-mechanical perturbation theory, in the first order in the perturbation, the probability per unit time of a transition from the state $|i\rangle$ to the state $|f\rangle$ is given by Fermi's Golden Rule:

$$\Omega_{i\to f}(\omega) = \frac{\pi {E_0}^2}{2\hbar^2}|\langle f|\varepsilon\cdot\hat{\mu}|i\rangle|^2 \times [\delta(\omega_{fi} + \omega) + \delta(\omega_{fi} - \omega)] \tag{3}$$

where $\omega_{fi} = \omega_f - \omega_i$, and $\delta(x)$ is the Dirac delta distribution.

The rule states that the transition rate is dependent only on the square of the matrix elements of the operator $\hat{\mu}$, where the δ enforces energy conservation. We proceed by assuming, without any loss of generality, that $\omega_{fi} > 0$. If also $\omega > 0$, the argument of the first δ-function in eq 3 is always nonzero, and, consequently, $\delta(\omega_{fi} + \omega)$ can be dropped:

$$\Omega_{i\to f}(\omega) = \frac{\pi {E_0}^2}{2\hbar^2}|\langle f|\varepsilon\cdot\hat{\mu}|i\rangle|^2\delta(\omega_{fi} - \omega) \tag{4}$$

The above expression corresponds to the transition rate for a pure state system.

The transition rate for the ensemble is given by the ensemble average of $\Omega_{i\to f}(\omega)$ over initial states. In the canonical ensemble, the probability of having an initial state that is an eigenstate of $\hat{\mathcal{H}}_0$ with energy E_i, is given by the equilibrium density matrix eigenvalue ρ_i,

$$\rho_i = \frac{e^{-\beta E_i}}{Q(N, V, T)} = \frac{e^{-\beta E_i}}{\mathrm{Tr}(e^{-\beta\hat{\mathcal{H}}_0})} \tag{5}$$

Hence, the ensemble average of the transition rate is given by

$$\begin{aligned}\Omega(\omega) &= \langle\Omega_{i\to f}(\omega)\rangle \\ &= \frac{\pi E_0^2}{2\hbar^2}\sum_{i,f}\rho_i|\langle f|\varepsilon\cdot\hat{\mu}|i\rangle|^2 \times \delta(\omega_{fi} - \omega)\end{aligned} \tag{6}$$

where we must sum over all the final states that are connected to the initial states by energy conservation.

So far, we have implicitly used the Schrödinger picture. We can convert eq 6 to the interaction picture in a simple way, by introducing the Fourier transform of the Dirac delta function $\delta(x)$:

$$\begin{aligned}\Omega(\omega) = \frac{{E_0}^2}{4\hbar^2}\int \mathrm{d}t \sum_{i,f}\rho_i\langle i|\varepsilon\cdot\hat{\mu}|f\rangle\langle f|\varepsilon\cdot\hat{\mu}|i\rangle \\ \times \exp\left[\left(\frac{E_f - E_i}{\hbar} - \omega\right)it\right]\end{aligned} \tag{7}$$

where we have used the symmetry of the inner product in complex vector space, and the hermicity of the operator $\hat{\boldsymbol{\mu}}$.

The state vectors $|i\rangle$ and $|f\rangle$ are eigenstates of the unperturbed Hamiltonian $\hat{\mathcal{H}}_0$. Accordingly, in the Schrödinger picture, the time dependence of these states can be written as

$$e^{-i\hat{\mathcal{H}}_0 t/\hbar}|i(0)\rangle = e^{-iE_i t/\hbar}|i(0)\rangle \tag{8}$$

$$\langle f(0)|e^{i\hat{\mathcal{H}}_0 t/\hbar} = \langle f(0)|e^{iE_f t/\hbar} \tag{9}$$

By substituting these expressions into eq 7, we can switch from the Schrödinger picture to the interaction picture:

$$\begin{aligned}\langle i|\varepsilon\cdot\hat{\mu}|f\rangle e^{\left[\left(E_f - E_i\right)/\hbar\right]it} &= \langle i|e^{-iE_i t/\hbar}\varepsilon\cdot\hat{\mu}e^{iE_f t/\hbar}|f\rangle \\ &= \langle i|\varepsilon\cdot e^{i\hat{\mathcal{H}}_0 t/\hbar}\hat{\mu}e^{-i\hat{\mathcal{H}}_0 t/\hbar}|f\rangle \\ &= \langle i|\varepsilon\cdot\hat{\mu}(t)|f\rangle\end{aligned} \tag{10}$$

and eq 7 becomes

$$\Omega(\omega) = \frac{{E_0}^2}{4\hbar^2}\int \mathrm{d}t\ e^{-i\omega t}\sum_{i,f}\rho_i\langle i|\varepsilon\cdot\hat{\mu}|f\rangle\langle f|\varepsilon\cdot\hat{\mu}(t)|i\rangle \tag{11}$$

Since $|f\rangle$ are eigenstates of the unperturbed Hamiltonian, they form a complete basis of the Hilbert space; therefore, the closure relationship $\sum_f |f\rangle\langle f| = 1$ holds. Substituting this expression into eq 11, we find

$$\Omega(\omega) = \frac{{E_0}^2}{4\hbar^2}\int \mathrm{d}t\ e^{-i\omega t}\sum_{i}\rho_i\langle i|\varepsilon\cdot\hat{\mu}(0)\varepsilon\cdot\hat{\mu}(t)|i\rangle \tag{12}$$

Finally, multiplying by ρ_i and summing over all the initial states, we obtain an equilibrium ensemble average

$$\Omega(\omega) = \frac{1}{3}\frac{{E_0}^2}{4\hbar^2}\int \mathrm{d}t\ e^{-i\omega t}\langle\hat{\mu}(0)\hat{\mu}(t)\rangle \tag{13}$$

where the factor 1/3 is due to the assumption of an isotropic system, for which any response is independent from the direction of the incident radiation ε, and $C^{rs}(t) = \langle\hat{\mu}(0)\hat{\mu}(t)\rangle$ is

DOI: 10.1021/acs.jctc.5b00391
J. Chem. Theory Comput. 2015, 11, 3321−3332

the right-sided quantum autocorrelation function of the dipole moment operator.

Here, we note that the operators $\hat{\mu}(0)$ and $\hat{\mu}(t)$ are individually Hermitian, but since they generally do not commute, the autocorrelation function $C^{rs}(t)$ is an expectation value of a non-Hermitian operator. As a consequence, $C^{rs}(t)$ is a complex function. This fact creates a breach between the quantum and classical formulation, and, in fact, the imaginary part of the autocorrelation function can be used as a measure of quantum effects.

A connection with the classical formulation can be recovered through a complete analysis of the transition rate function in the frequency spectrum, i.e., considering the case of $\omega < 0$ as well. In this case, the argument of the second δ in eq 3 is always nonzero, which implies $\delta(\omega_{fi} - \omega)$ can be dropped. This is equivalent to substituting $-\omega$ for ω in eq 6, which corresponds to an emission process.

By interchanging the indices i and f in eq 6 and using the relation for ρ_f for a system in thermal equilibrium:

$$\rho_f = \frac{e^{-\beta E_f}}{\mathrm{Tr}(e^{-\beta\hat{\mathcal{H}}_0})} = \frac{e^{-\beta(E_i+\hbar\omega)}}{\mathrm{Tr}(e^{-\beta\hat{\mathcal{H}}_0})} = \rho_i e^{-\beta\hbar\omega} \tag{14}$$

The ensemble transition rate for the emission process becomes

$$\begin{aligned}\Omega(-\omega) &= \frac{\pi E_0^{\,2}}{2\hbar^2} e^{-\beta\hbar\omega} \sum_{i,f} \rho_i |\langle i|\varepsilon\cdot\hat{\mu}|f\rangle|^2 \times \delta(\omega_{fi} - \omega) \\ &= e^{-\beta\hbar\omega}\Omega(\omega)\end{aligned} \tag{15}$$

The expression in eq 15 represents the well-known detailed balance condition for a system in thermal equilibrium. The ensemble averaged absorption and emission rates, $\Omega(\omega)$ and $\Omega(-\omega)$, respectively, do not obey the microscopic reversibility condition.

If we carry out the same analysis that led from eq 6 to eq 13 for the transition rate $\Omega(-\omega)$, the result is

$$\Omega(-\omega) = \frac{E_0^{\,2}}{12\hbar^2}\int dt\, e^{-i\omega t}\langle\hat{\mu}(t)\hat{\mu}(0)\rangle \tag{16}$$

and since $\Omega(-\omega) \neq \Omega(\omega)$, the two correlation functions $\langle\hat{\mu}(t)\hat{\mu}(0)\rangle$ and $\langle\hat{\mu}(0)\hat{\mu}(t)\rangle$ are not equal.

The net energy absorbed per unit time at frequency ω is given by

$$\begin{aligned}\Pi(\omega) &= [\Omega(\omega) - \Omega(-\omega)]\hbar\omega \\ &= \hbar\omega\Omega(\omega)(1 - e^{-\beta\hbar\omega})\end{aligned} \tag{17}$$

However, from the detailed balance condition, we can obtain the following relation:

$$\Omega(\omega) + \Omega(-\omega) = (1 + e^{-\beta\hbar\omega})\Omega(\omega) \tag{18}$$

Since taking an ensemble average is a linear operation, we can rewrite $\Omega(\omega) + \Omega(-\omega)$ as

$$\begin{aligned}&\frac{E_0^{\,2}}{12\hbar^2}\int dt\, e^{-i\omega t}\langle\hat{\mu}(0)\hat{\mu}(t) + \hat{\mu}(t)\hat{\mu}(0)\rangle \\ &= \frac{E_0^{\,2}}{12\hbar^2}\int dt\, e^{-i\omega t}\langle\{\hat{\mu}(0), \hat{\mu}(t)\}\rangle\end{aligned} \tag{19}$$

where $\{\cdot,\cdot\}$ is the anticommutator. Finally, we can recast the formula for the net energy absorption spectrum in eq 17 as

$$\Pi(\omega) = \frac{2\omega E_0^{\,2}}{12\hbar}\tanh\left[\frac{\beta\hbar\omega}{2}\right] \times \int dt\, e^{-i\omega t}\left\langle\frac{1}{2}\{\hat{\mu}(0), \hat{\mu}(t)\}\right\rangle \tag{20}$$

where $C^{sym}(t)$ is the symmetrized autocorrelation function ($C^{sym}(t) = \langle{}^1/_2\{\hat{\mu}(0),\hat{\mu}(t)\}\rangle$). Unlike the function in eq 13, $C^{sym}(t)$ is a real function. We have found a quantum analogue candidate for the absorption spectrum, in terms of the autocorrelation function of a Hermitian operator.

We can define an absorption cross-section $\alpha(\omega)$ as the ratio $\Pi(\omega)/(V\|\mathbf{S}\|)$, where V is the volume of the system, and $\|\mathbf{S}\|$ the magnitude of the incident flux of the radiation, the Poynting vector. Different formulas for $\alpha(\omega)$ are possible, depending ultimately on which autocorrelation function one adopts. It is also useful to define an absorption line shape $I(\omega)$, as

$$I(\omega) \equiv \frac{1}{2\pi}\int dt\, e^{-i\omega t}C(t) \tag{21}$$

where $C(t)$ is the generic autocorrelation function, such that the cross-section function can be written, more generally, as

$$\alpha(\omega) = \frac{4\pi^2\omega}{3\hbar Vcn(\omega)}(1 - \exp[-\beta\hbar\omega])\zeta(\omega)I^{\zeta}(\omega) \tag{22}$$

where V is the volume of the system, $n(\omega)$ is the refractive index, c the speed of light, and $\zeta(\omega)$ is implicitly defined by defining the autocorrelation function $C(t)$. In the next section, we will show that many physically equivalent quantum autocorrelation functions can be defined. The formula in eq 22 represent the main result of our derivation, that is that the cross-section function $\alpha(\omega)$ can be expressed in terms of the Fourier transform of a given dipole moment autocorrelation function $C(t)$ of the absorbing system, in the absence of any external field.

3. DETAILED BALANCE CONDITIONS AND QUANTUM CORRECTION FACTOR

The formulas in eqs 21 and 22 are derived directly from first-principles, with the assumption of a weak coupling between the system and the external field. Under this assumption, the linear response theory[1] can be used to predict many dynamical properties of the system.

However, in Born−Oppenheimer molecular dynamics (BOMD), the trajectories of nuclei are generated by integrating the classical laws of motion, with the forces computed "on-the-fly" from electronic structure calculations. Consequently, the observables that are dependent on nuclear positions, such as the total dipole moment, must be considered as semiclassical objects. Obviously, the BOMD approach does not satisfy the assumptions made in the previous section. Nonetheless, we can formally maintain the formulas in eq 21 by replacing the quantum operators with classical vectors $\hat{\mu}\ (t) \rightarrow \mu(t)$. When dealing with classical systems, there exists only one autocorrelation function, $C^{cl}(t) = \langle\mu(0)\cdot\mu(t)\rangle$, where $\langle\cdot\rangle$ is a classical ensemble average and $\mu(t)$ is a vector in Euclidean space. Hence, by simply replacing the classical autocorrelation function in eq 21, we obtain the classical line shape function $I^{cl}(\omega)$, and the cross-section function

$$\alpha(\omega) = \frac{4\pi^2\omega}{3\hbar Vcn(\omega)}(1 - \exp[-\beta\hbar\omega])\ I^{cl}(\omega) \tag{23}$$

which is the classical analogue of eq 22.

DOI: 10.1021/acs.jctc.5b00391
J. Chem. Theory Comput. 2015, 11, 3321−3332

However, the classical and quantum lineshapes have different symmetry. In fact, $I^{cl}(\omega)$ is an even function of ω and does not satisfy the more-complicated expression $I(-\omega) = \exp[-\beta\hbar\omega]\, I(\omega)$ required by the quantum counterpart (which is, essentially, a restatement of the detailed balance condition in the frequency domain). This breaks the symmetry of the cross-section function $\alpha(\omega) = \alpha(-\omega)$. The failure of the classical approach in fulfilling the detailed balance condition is at the heart of the problem. A standard procedure to force $I^{cl}(\omega)$ to satisfy the detailed balance condition is to introduce a frequency-dependent quantum correction factor, such that the symmetry enclosed in eq 22 is restored. Alternatively, more-sophisticated approaches could be used, such as path integral molecular dynamics (PIMD).

In addition to $C^{rs}(t)$ and $C^{sym}(t)$ defined in the previous section, we can define other quantum autocorrelation functions, such as the symmetric $C^{asym}(t)$, the shifted $C^{shift}(t)$, and the Kubo autocorrelation functions $C^{Kubo}(t)$.[17] The only difference between $C^{rs}(t)$ and the other correlation functions appears in the quantum limit, $\lim_{\beta\to\infty}$, where there is a loss of dynamical information, as pointed out in ref 18.

The physical information carried by the different autocorrelation functions is the same, in the sense that if any of these functions is known, we can immediately derive the others, since their Fourier transforms are interrelated by simple expressions.[17] This also implies that $C^{sym}(t)$ and $C^{asym}(t)$ are not independent functions, which is a consequence of the Kramers–Kronig relation. Among all the possible correlation functions, the Kubo autocorrelation function is particularly important:

$$C^{kubo}(t) = \left\langle \frac{1}{\beta}\int_0^\beta d\lambda\, \hat{\mu}(0)\hat{\mu}(t + i\hbar\lambda) \right\rangle = \langle \tilde{\hat{\mu}}(0)\hat{\mu}(t)\rangle \tag{24}$$

which, in the frequency domain, is related to the right-sided lineshape through

$$I^{rs}(\omega) = \frac{\beta\hbar\omega}{1 - \exp[-\beta\hbar\omega]} I^{Kubo}(\omega) \tag{25}$$

In fact, it can be shown that $C^{Kubo}(t)$ is a real and even function of time, as is $C^{cl}(t)$, and since the Fourier transform conserves parity, it follows straightforwardly that $I^{Kubo}(\omega) = I^{Kubo}(-\omega)$.

Moreover, in the classical limit, $\beta \to 0$, the prefactor in eq 25 approaches to unity, which implies $I^{rs}(\omega) = I^{Kubo}(\omega)$, which, in turn, implies the equality of the respective autocorrelation functions. This suggests that $C^{cl}(t)$ be replaced with the Kubo autocorrelation function, and it also suggests that the corresponding prefactor be used to restore the symmetry in eq 22. The lineshape function with the correct symmetry can be written as

$$I(\omega) = Q^{QC}(\omega) I^{cl}(\omega) \tag{26}$$

In the literature, the Kubo correction factor is also known as the harmonic correction factor and is widely used to correct classical molecular dynamics results for detailed balance, since it systematically gives results in better agreement with experimental works for the mid-infrared range, compared to other correction factors.[8] By testing the different prefactors, we arrived at the same conclusions; therefore, this is the only prefactor used in our results.

4. POLARIZATION IN ONETEP

The BOMD trajectories were generated using the ONETEP linear-scaling code. In ONETEP, Kohn–Sham (KS)-like equations[19] for the single-particle density matrix P are solved self-consistently, through direct energy minimization scheme, under periodic (Born–von Kármán) boundary conditions (PBC).

The linear scaling regime can be achieved by exploiting Kohn's principle of "nearsightedness" of electronic matter.[20,21] In fact, for systems with a nonzero band gap, the density matrix decays exponentially:

$$\rho(\mathbf{r}, \mathbf{r}') = \langle \mathbf{r}|P|\mathbf{r}'\rangle \propto \exp(-\gamma|\mathbf{r} - \mathbf{r}'|) \tag{27}$$

where γ is a positive constant that can be related heuristically to the physics of the system. However, quantifying the degree of spatial localization, by relating γ to a property of materials, remains an open issue. A functional form of γ, in terms of the direct band gap, has been proposed by many authors.[22−25] On the other hand, analytical results have supported the idea that γ can be related to the ratio between $k_B T$ and Fermi velocity in metals at finite temperature.[26]

By assuming the functional form in eq 27, it becomes possible to apply a truncation, such that the resulting density matrix has a sparse band-diagonal form. This is achieved by imposing

$$\rho(\mathbf{r}, \mathbf{r}') = 0, \quad |\mathbf{r} - \mathbf{r}'| > \mathbf{R}_{cut} \tag{28}$$

Moreover, the density matrix can be expanded in separable form,[27] in terms of localized functions, and when considering large systems for which Γ-point only sampling is sufficient, this yields

$$\rho(\mathbf{r}, \mathbf{r}') = \phi_\alpha(\mathbf{r}) K^{\alpha\beta} \phi^*_\beta(\mathbf{r}') \tag{29}$$

Henceforth, the Einstein summation convention for repeated Greek indices is used. The spatially localized functions are real functions and are known as nonorthogonal generalized Wannier functions (NGWFs).[28] NGWFs are centered on atoms and are constrained to fixed localization regions $V^\alpha_{loc}(\mathbf{R}_I)$, with $\mathbf{R}_I$ denoting the center of the Ith atom. Nonorthogonality enhances localization and promotes the use of sparse algebra, at the price of a nonunit metric tensor $S_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle$. The matrix $\mathbf{K}$ is known as the kernel of the density matrix; it is a generalization of the occupation numbers to the case of nonorthogonal functions.

In ONETEP the NGWFs are optimized *in situ*, as part of the energy minimization procedure, allowing the use of a minimal set of functions and, therefore, a minimal size for the Hamiltonian matrix. For a more detailed description of the energy minimization technique, sparse algebra, and scaling in ONETEP the reader is referred to the works of Skylaris et al.[12] and Hine et al.[29]

In this framework, the electronic dipole moment for an isolated system is given by

$$\mu_{el} = 2\mathrm{Tr}(\mathbf{KR}) = 2K^{\alpha\beta}R_{\beta\alpha} \tag{30}$$

where the factor 2 comes from considering a closed shell system, with doubly occupied NGWFs, and $\mathbf{R}$ represents an array of three matrices $\mathbf{R} = (\mathbf{X}, \mathbf{Y}, \mathbf{Z})$, whose elements are given by

$$X_{\beta\alpha} = \langle\phi_\beta|x|\phi_\alpha\rangle \tag{31}$$

DOI: 10.1021/acs.jctc.5b00391
J. Chem. Theory Comput. 2015, 11, 3321−3332

for **X** and similarly for **Y** and **Z**.

A comparison between NGWFs and maximally localized Wannier functions (MLWFs)[30] has been presented in ref 31, showing the accuracy of our approach in computing electric dipole polarizabilities comparable to MLWF methods, with the advantage that NGWFs are computed "on-the-fly", obviating the need for post-processing.

For the electronic dipole moment of a nonisolated system A within a larger system A′ = A + B, a density partitioning scheme is required. Different methods for partitioning the density matrix into atomic/pointwise,[32,33] or fragment[34] contributions are known. Within MLWF approaches, partition of the density emerges naturally,[30] as a consequence of the orthogonality and localization constraints of the MLWFs: the density matrix ρ_A corresponding to system A can be built from MLWFs belonging to A only, and the number of electrons in A is correctly given by $N_A^{elec} = \mathrm{Tr}(\rho_A)$.

In the ONETEP framework, on one hand, NGWFs are more easily assigned to a given system, as they are centered on atoms, while MLWFs are not. On the other hand, nonorthogonality of NGWFs plays an important role: the density matrix ρ_A cannot be built only from NGWFs centered on atoms in A. In the following discussion, we outline two approaches to overcome this obstacle: density kernel partitioning (DKP) and distributed multipole analysis (DMA). Both methods have been implemented in ONETEP, and they are now part of its academic distribution.

4.1. Density Partitioning by Kernel Partitioning. The density kernel partitioning (DKP) method that we present computes the dipole moment of a subsystem A directly from the matrix elements of **K** and **R** of the total system A′.

To properly describe the electronic dipole of A, contributions originating from the NGWFs centered on atoms in B, whose overlap S_{BA} with NGWFs centered on atoms in A is nonzero, must be included. In our method, we consider **K** and **X**, **Y**, **Z** of the entire system A′, which are real and symmetric matrices by construction. They can be rearranged into block matrices: two square, diagonal blocks AA and BB, whose elements originate from NGWFs centered only on atoms in A or B, respectively, and two rectangular, off-diagonal blocks AB and BA, where AB = BA^T, with the obvious meaning of symbols. The double tensor contraction in eq 30 represents a matrix multiplication **M** = **K**·**R**, followed by taking the trace of the resulting matrix. The block structure is preserved under matrix multiplication. Therefore, the trace of the block AA in **M** gives the electronic dipole of the subsystem A (see Figure 1). This is equivalent to adding the trace of the matrices obtained from the multiplication of the AA blocks in **K** and **R** components and the trace of the matrices obtained from the multiplication of the BA block in **K** and AB block in the **R** components (or the transpose of the BA block since the **R** components are symmetric), which is half of the total contribution from off-diagonal blocks:

$$\begin{aligned}\mu_{el}^{A} &= 2[\mathrm{Tr}(\mathbf{KR})_{AA} + \mathrm{Tr}(\mathbf{KR})_{BA}] \\ &= 2[\sum_{\alpha,\beta\in AA} K^{\alpha\beta}R_{\beta\alpha} + \sum_{\gamma,\delta\in BA} K^{\delta\gamma}R_{\gamma\delta}]\end{aligned} \tag{32}$$

We end this section by stressing that the density kernel partitioning does not retain the exact number of electrons for the subsystems. However, in numerical tests that we performed, the electronic charge of the subsystem $\mathrm{Tr}(\mathbf{M}_{AA})$ was within 0.05% of the expected number of valence electrons.

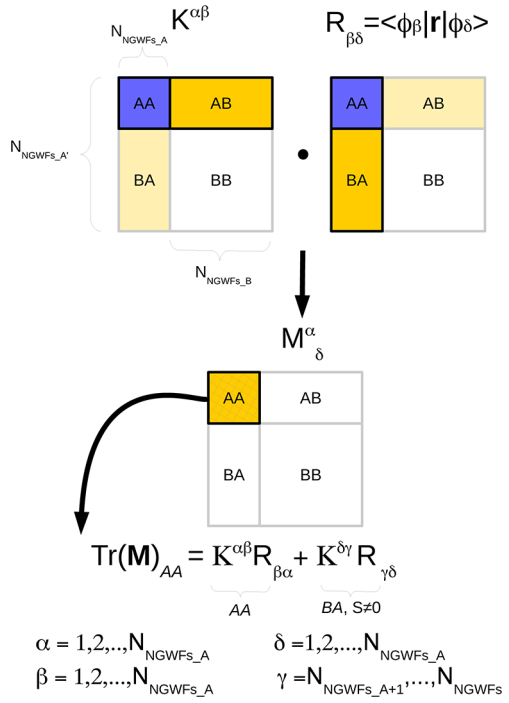


Figure 1. Schematic representation of the density kernel partitioning (DKP) method for the partitioning of the dipole moment.

4.2. Density Partitioning by Distributed Multipole Analysis. Another approach to partitioning the density between the two subsystems relies on distributed multipole analysis (DMA). Here, the charge distribution of the entire system is represented in terms of a multipole expansion. Atomic centers are usually, although not universally, used as the centers for the multipoles. The DMA approach, first proposed by Rein,[35] has been pioneered and made popular by Stone[32] and Alderton.[33] DMA is typically performed in a Gaussian basis set;[36,37] below, we outline how electronic densities represented in a localized (NGWF) basis can be similarly expanded.

We begin by decomposing the total electronic density $n(\mathbf{r})$ into on-site and off-site contributions:

$$n(\mathbf{r}) = \sum_I \sum_{i_1,i_2\in I} \phi_{i_1}(\mathbf{r})K^{i_1i_2}\phi_{i_2}(\mathbf{r}) + \sum_I \sum_{J\neq I} \sum_{i\in I} \sum_{j\in J} \phi_i(\mathbf{r})K^{ij}\phi_j(\mathbf{r}) \tag{33}$$

$$n(\mathbf{r}) = \sum_I n_{II}(\mathbf{r}) + \sum_I \sum_{\substack{J\neq I \\ S_{IJ}\neq 0}} n_{IJ}(\mathbf{r}) \tag{34}$$

In eq 34, we have explicitly separated density contributions from NGWFs on the same atomic center I and those on different atomic centers I and J with nonzero overlap S_{IJ}. By construction, NGWFs that do not overlap do not contribute to density. We denote with i or i_1, i_2, ... (and correspondingly with j, j_1, j_2, ...) the NGWFs belonging to the atom I (J).

Our next goal is to approximate each of the off-site contributions to density as a sum of on-site contributions, represented in an auxiliary basis set, i.e.,

$$\tilde{n}_{IJ}(\mathbf{r}) = \sum_s^{N_f} f_s(\mathbf{r})c_{IJ}^s \tag{35}$$

where c_{IJ}^s are the sought coefficients in the expansion, and $\{f_s(\mathbf{r})\}_{s=1}^{N_f}$ are the functions comprising our auxiliary basis set,

DOI: 10.1021/acs.jctc.5b00391
J. Chem. Theory Comput. 2015, 11, 3321−3332

$N_f/2$ of which originate on center I, and the remaining $N_f/2$ originate on center J.

We subsequently define an electrostatic metric V:

$$V_{st} = (f_s|f_t) \tag{36}$$

where, for the sake of brevity, we introduced the notation

$$(g|h) = \int\int \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}'\, g^*(\mathbf{r}) \frac{1}{|\mathbf{r}-\mathbf{r}'|} h(\mathbf{r}') \tag{37}$$

Minimization of the electrostatic self-energy of the difference between the exact and approximate (interpolated) density, i.e., of the quantity

$$\frac{1}{2}([n_{IJ} - \tilde{n}_{IJ}]|[n_{IJ} - \tilde{n}_{IJ}]) \tag{38}$$

yields the expansion coefficients as

$$c_{IJ}^s = \sum_t^{N_f} (n_{IJ}|f_t) V^{ts} \tag{39}$$

where t similarly indexes auxiliary basis functions originating on centers I and J, and V^{ts} are elements of the inverse electrostatic metric matrix. The electrostatic overlaps $(n_{IJ}|f_t)$ are equivalent to standard overlaps $\langle n_{IJ}|u_t\rangle$, where the auxiliary basis function f_t has been replaced by its Coulombic potential u_t, i.e.,

$$u_t(\mathbf{r}) = \int \mathrm{d}\mathbf{r}'\, \frac{f_t(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{40}$$

In practice, the overlaps are evaluated numerically on a Cartesian grid, and the auxiliary basis functions are chosen such that their potential can be obtained analytically. For a detailed description of the procedure, with a discussion of the practicality of using truncated spherical waves as the auxiliary basis set, the reader is referred to ref 38.

We note that the above expressions retain validity for the case of $I = J$, which lets us use the same formalism to expand on-site products for consistency. [Note that, when on-site densities are expanded, only one center is used, and the number of functions N_f in the auxiliary basis set is halved accordingly.]

We have now decomposed the electronic density of the system into a sum of atom-centered contributions, with each contribution being a linear combination of auxiliary basis functions. As each atomic center I participates in the expansion of a number of pairs (I,J), the total number $n(I)$ of coefficients associated with each center I is $n(I) = {}^1/_2 \sum_{J,S_{IJ}\neq 0} N_f$, while the number of auxiliary basis functions on each center is $N_f/2$.

For the sake of concreteness, in the text that follows, we explicitly assume the auxiliary basis functions to be truncated spherical waves (SWs),[38−40] i.e.,

$$f_s(\mathbf{r}) \equiv f_{lmq}(\mathbf{r}) = \begin{cases} j_l(qr) Z_{lm}(\hat{\mathbf{r}}) & r < a \\ 0 & r \geq a \end{cases} \tag{41}$$

where $j_l(qr)$ is a spherical Bessel function and $Z_{lm}(\hat{\mathbf{r}})$ is a real spherical harmonic. Values of q are chosen so that the truncation does not introduce a discontinuity, i.e., $j_l(qa) = 0$. The radii of localization of the SWs (a) and of the NGWFs coincide. For the purpose of calculating dipole moments, we limit the values of the angular momentum quantum number to $l \leq 1$ (although the approach is general), the magnetic quantum number $m \in [-l,l]$, and the number of distinct values of q in the auxiliary basis set (typically ∼10) is dependent on the psinc (periodic cardinal sine function) kinetic energy cutoff used in the calculations.

We now turn to the calculation of the dipole moment. The multipoles associated with an atomic center I can be calculated as

$$\mathcal{M}_{lm}(I) = \sum_{\substack{J \\ S_{IJ}\neq 0}} \sum_{s_I}^{N_f/2} c_{IJ}^{s_I} \int_{\mathcal{S}_{I,a}} \mathrm{d}\mathbf{r}\, |\mathbf{r}|^l C_{lm}(\hat{\mathbf{r}}) f_{s_I}(\mathbf{r}) \tag{42}$$

where the subscript I that is applied to s reminds the reader that only the SWs originating on atom I participate in the second sum. Because of the localization of truncated spherical waves, the integral only runs over a spherical volume $\mathcal{S}_{I,a}$ of radius a centered at $\mathbf{R}_I$. C_{lm} are spherical harmonics with Racah normalization, i.e.,

$$C_{lm}(\hat{\mathbf{r}}) = \sqrt{\frac{4\pi}{2l+1}}\, Y_{lm}(\hat{\mathbf{r}}) \tag{43}$$

where Y_{lm} denote complex spherical harmonics.

For the sake of facilitating subsequent discussion, we will now switch to explicitly labeling the SWs with the indices $l(s_I)$, $m(s_I)$, and $q(s_I)$, i.e.,

$$f_{l(s_I)m(s_I)q(s_I)}(\mathbf{r}) \equiv f_{s_I}(\mathbf{r}) \tag{44}$$

We separate the angular and radial parts in the integrals in eq 42, and employ the properties of spherical harmonics, obtaining

$$\int_{\mathcal{S}_{I,a}} \mathrm{d}\mathbf{r}\, |\mathbf{r}|^l C_{lm}(\hat{\mathbf{r}}) f_{l(s_I)m(s_I)q(s_I)}(\mathbf{r}) \tag{45}$$

$$= \delta_{ll(s_I)}\delta_{mm(s_I)} J_{l(s_I)q(s_I)} \tag{46}$$

where

$$J_{lq} = \int_0^a \mathrm{d}r\, r^{l+2} j_l(qr) \tag{47}$$

is a radial integral that can be computed analytically.

The atom-centered multipoles obtained from eq 42 are already suitably normalized; atom-centered, Cartesian dipoles are obtained as

$$\mu_{\mathrm{el}}(I) = [\mathcal{M}_{11}(I),\, \mathcal{M}_{1\bar{1}}(I),\, \mathcal{M}_{10}(I)] \tag{48}$$

The total electronic dipole moment of subsystem A, with respect to a chosen origin $\mathbf{O}$, is computed by vector addition of atom-centered multipoles, following a suitable change of frame of reference:

$$\mu_{\mathrm{el}}^{\mathrm{A}}(\mathbf{O}) = \sum_{I\in\mathrm{A}} \mu_{\mathrm{el}}(I) - (\mathbf{O} - \mathbf{R}_I)\mathcal{M}_{00}(I) \tag{49}$$

We note that the procedure described above minimizes the quantity given in eq 38, with no constraint on the total charge of the entire system $\sum_I \mathcal{M}_{00}(I)$, let alone any of its subsystems. In the numerical tests that we performed, the total monopole obtained from the expansion was within 0.05% of the expected number of valence electrons for reasonable qualities of the auxiliary basis set.

5. CALCULATION DETAILS

All the BOMD calculations and post-processing were carried out in ONETEP, whereas the classical MD trajectories were generated with the AMBER12 MD engine SANDER.[41] In the

DOI: 10.1021/acs.jctc.5b00391
J. Chem. Theory Comput. 2015, 11, 3321−3332

ONETEP BOMD engine, Newtonian equations of motion for the nuclei are integrated by adopting the well-known velocity-Verlet scheme, where forces are generated "on-the-fly" from the ground-state electronic configuration. Hybrid parallelism[42] and extrapolation schemes were exploited to minimize the walltime per MD step.

For all the *ab initio* calculations, we ran microcanonical BOMD with periodic boundary conditions. A time step of 0.5 fs was used, this value is a compromise between energy conservation, minimization of error due to time discretization, and computational cost. We employed a kinetic energy cutoff of 800 eV, with the PBE exchange-correlation functional and the D2 Grimme dispersion correction.[43,44] The radius for the NGWFs localization regions has been set to $8.0a_0$. For the valence states, we required 4 NGWFs for all the atoms but hydrogens, where only 1 NGWF was sufficient; norm-conserving pseudopotentials were adopted for dealing with core states. The total energy was minimized self-consistently through the LNV scheme.[45−47] At every time step, the starting set of NGWFs and density kernel fed to the LNV minimization scheme were obtained from a linear combination of the previous MD steps NGWFs/density kernels

$$\chi_i^{\text{init}} = 2\chi_{i-1}^{\text{scf}} - \chi_{i-2}^{\text{scf}}$$

where χ^{scf} represents a converged NGWF/density kernel. The expressions in eqs 30 and 32 were used to compute the dipole moment of single molecules in the gas phase, and the dipole moment of the solute when considering a molecule in solution, respectively. Unless stated otherwise, the parameters summarized in Table 1 were used in all BOMD simulations.

Table 1. Parameters Common to All BOMD Simulations

parameter	value
time step	0.5 fs
temperature	300 K
thermostat	none
kinetic energy cutoff	800 eV
XC functional	PBE
NGWF radius	$8.0a_0$
outer loop rms threshold	10^{-9}
inner loop rms threshold	10^{-6}
NGWF extrap. scheme	linear
density kernel extrapolation scheme	linear

In all the classical calculations, we used the general AMBER force field (GAFF). A time-step of 0.5 fs was chosen for consistency with BOMD calculations. The parameter file and the initial coordinates were generated with the LEAP graphical user interface xLEaP. We followed this protocol: we performed a geometry relaxation calculation for 2500 steps, where, in the first 1500 steps, we used steepest descents for minimizing the energy, and, in the remaining 1000 steps, we used the conjugate gradient method. The bonds containing hydrogens were kept fixed throughout, using the SHAKE algorithm. After the geometry relaxation, we performed a heating calculation, where we let the temperature increase from 0 to 300 K, using Langevin dynamics with the collision frequency coefficient of $\gamma = 2.0\ \text{ps}^{-1}$ for 10 ps with a cutoff for the electrostatic interactions of 10.0 Å. A further equilibration calculation was performed at a fixed temperature of 300 K for 10.0 ps, with the other parameters unchanged. After this equilibration phase, a production run was performed, where the SHAKE constraints and the thermostat were removed, and the total simulation length was set to 20.0 ps.

The IR spectra were then computed using eq 22, where the dipole moments are simply given by

$$\mu(t) = \sum_I \nu_I \mathbf{R}_I(t) \tag{50}$$

where ν_I are the total classical charges on each atom I, obtained from the force field parameters, and $\mathbf{R}_I$ are the atomic positions. The parameters in Table 2 were used in all the classical calculations.

Table 2. Parameters Common to All Classical Simulations

parameter	value
time step	0.5 fs
force field	GAFF
temperature	300 K
thermostat	Langevin
frequency coefficient, γ	$2.0\ \text{ps}^{-1}$
electrostatic interaction cutoff	10.0 Å
pressure relaxation time, τ	2.0 ps

Following Gaigeot,[8] we applied a Gaussian filter window in the time domain, such that each term of the correlation function $C(t_i)$ was multiplied by a Gaussian function $\exp(-0.5\sigma(t_i/\tau)^2)$, where τ is the length of the simulation, and $\sigma = 10$ for the gas phase and $\sigma = 40$ for the aqueous phase.

5.1. Gas Phase. *5.1.1. Water.* For water in the gas phase, we used a cubic cell with a length of $a = 15.0$ Å. The water molecule was originally placed at the center of the cell. The initial geometry was obtained from a geometry optimization calculation. The initial velocities were extracted from a Maxwell−Boltzmann distribution at the imposed temperature of 300 K. The total time was 8.0 ps. We let the system equilibrate for 1.0 ps, and then performed a production run of 7.0 ps. To calculate the normal modes, we used the built-in phonon module in ONETEP. In order to compute the normal modes, the forces on the nuclei and their derivatives must be evaluated with high accuracy, because the normal modes are obtained from the diagonalization of the Hessian matrix of the total energy. Thus, a geometry optimization calculation with a well-converged density is required. The procedure requires an energy of $3N$ and forces calculations and one Hessian diagonalization.

5.1.2. Carbon Dioxide. For carbon dioxide, we used a cubic cell with a length of $a = 12.0$ Å. As for the water molecule, we placed the CO_2 molecule at the center of the cell. The simulation time was 8.0 ps, where we equilibrated the system over the initial 1.0 ps. The initial velocities were extracted from a Maxwell−Boltzmann distribution at the imposed temperature of 300 K.

5.1.3. Ethanol. For the ethanol molecule in the gas phase, we adopted the same procedure described above. We used a cubic cell with a length of $a = 18.0$ Å. The temperature was fixed at 296.15 K for consistency with the experimental data. The simulation length was 5.5 ps, and a time step of 0.75 fs has been used. This slightly larger time step allowed the simulations to be accelerated while still ensuring a good conservation of total energy. As a consequence, we performed NVE-BOMD for ∼7300 steps. The systems was equilibrated for 1.5 ps, and the length of the production run was 4.0 ps.

DOI: 10.1021/acs.jctc.5b00391
J. Chem. Theory Comput. 2015, 11, 3321−3332

5.1.4. 2-Ala. In order to test the accuracy of the DKP and DMA methods of sections 4.1 and 4.2 in generating IR spectra of systems in the condensed phase, we considered a well-known system referred in the literature as alanine dipeptide (2-Ala). 2-Ala is the alanine amino-acid equipped with minimal peptide bonds and with the C-terminal and N-terminal capped with methyl groups, and constitutes one of the simplest system with a peptide motif. This system has been extensively studied by Gaigeot et al.,[8,48−50] both in the gas phase and the aqueous phase. The study of 2-Ala in the gas phase is crucial for understanding the effects of the solvent in the aqueous phase. However, addressing the issue of the preferred conformations of 2-Ala in water is not the object of this work.

For 2-Ala in the gas phase, we used a cubic box with a length of a = 20.0 Å. We ran a 2600-step NVE-BOMD simulation (1.3 ps). The initial momenta were generated from a Maxwell−Boltzmann distribution, in order to obtain an initial temperature of 300 K. Prior to the MD, the initial geometry was relaxed, by performing a geometry optimization in ONETEP. [Note that the initial geometry was taken from the xLeap coordinate file.]

5.2. Aqueous Phase. The 2-Ala-water system is not interesting *per se*, but it is a prototype of larger and more-complex peptides. Moreover, since 2-Ala represents a floppy molecule of minimal size, it is a perfect candidate for testing our methods within the BOMD approach. Studying 2-Ala conformational changes in a polar solvent such as water can help shed light on the complex dynamics of larger biomolecules *in vivo*.

5.2.1. 2-Ala in AMBER. For 2-Ala in the aqueous phase, we used periodic boundary conditions. We solvated the system with 500 water molecules through the SOLVATEBOX command in the xLEaP interface, using the TIP3P model for the water molecules. Subsequently, the system was relaxed by first relaxing the solvent for 1500 steps while keeping the solute's atomic positions fixed, and then relaxing the entire system for an additional 3500 steps. We then continued to heat the system from 0 to 300 K, followed by a constant pressure calculation at p_{ext} = 1 atm for 10.0 ps to let the entire system expand and obtain a density of 0.9847 g cm^{-3}. We used an isotropic position scaling barostat and a pressure relaxation time of τ = 2.0 ps. The final cell volume was V_{fin} = 15373.4 Å^3. Finally, we performed an NVE production run for 40.0 ps, where all the SHAKE constraints were removed.

5.2.2. 2-Ala in ONETEP. Both the DKP and DMA methods are linear scaling and they can be performed "on the fly". However, the DMA method at this stage of the implementation is not able to treat systems where the NGWFs span the boundary of the simulation cell, i.e. when periodic images of NGWFs re-enter the original cell. Moreover, the DMA method requires computation of the electrostatic metric V (from eq 36) at every MD step; although this can be very efficiently parallelized, it is still very costly for large systems. Consequently, at every MD step, we extracted, from the full system of 1522 atoms, a smaller system containing the 2-Ala molecule and the closest 27 water molecules whose NGWFs were all fully contained inside the original simulation cell. The remaining water molecules of the full system were transformed into embedding classical charges. Therefore, for every MD step, we considered 109 quantum atoms and 1413 embedding classical charges (see Figure 2). Convergence tests were performed to ensure that the size of the reduced quantum system was sufficient to obtain well-converged values for the Cartesian components of the solute

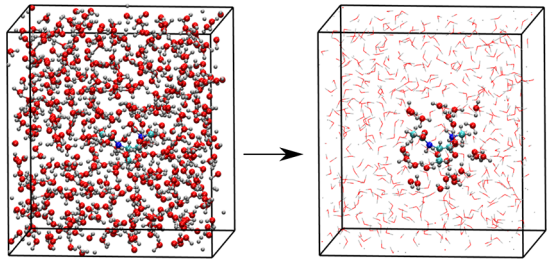


Figure 2. Reduction of the system size. Left panel shows the original cell with all atoms treated quantum-mechanically. Right panel shows the reduced-size system with 109 atoms treated quantum-mechanically and the remaining 1413 atoms represented as embedding classical charges. This transformation was applied to all the BOMD snapshots.

dipole moment. The partial charges on classical hydrogen and oxygen atoms were obtained from the TIP3P model, as explained in ref 51. Specifically, the partial charge on classical oxygens was −0.834e; for classical hydrogens, it was 0.417e. The effect of embedding classical charges is to yield a smoother profile for the 2-Ala electronic dipole moment, with respect to the system size, and therefore to reach the converged full quantum system value faster.

The initial atomic coordinates and velocities and the simulation cell for the full quantum system were obtained from the classical calculations described above, after the equilibration procedure. We ran a 2080-step NVE-BOMD simulation with 1522 atoms, using 2256 CPU cores to obtain a trajectory ∼1.0 ps long.

6. RESULTS AND DISCUSSION

The results from the BOMD and classical simulations were postprocessed to obtain the autocorrelation functions and the IR spectra.

The water molecule in the gas phase vibrates in several ways, involving combinations of the symmetric (exp. at 3657.1 cm^{-1}), and asymmetric (exp. at 3755.9 cm^{-1}) stretching, and bending of the covalent bonds (exp. at 1594.7 cm^{-1}).[53] The results for the entire simulation and the final spectra with the Kubo quantum correction factor are shown in Figure 3. As we can see from Figure 3, we found a good agreement with the experimental spectrum, where both the positions and the intensities of the different bands are very well reproduced. The breadth of the symmetric and asymmetric stretch bands is due to the rotational fine structure of these bands.

The results for the carbon dioxide are shown in Figure 4. As we can see, the intensities and positions of the asymmetric C═O stretch mode (exp. at 2249 cm^{-1}) and the two degenerate O═C═O bending modes (exp. at ∼660 cm^{-1}) are reproduced very well. As expected, the symmetric C═O stretch mode (exp. at 1338 cm^{-1}) is not IR-active, because the molecular dipole moment does not change under this vibrational mode, and, therefore, is absent in the spectrum. More interestingly, the combination bands at ∼3600 and ∼3700 cm^{-1}, which originate purely from anharmonic effects, are also recovered.

As we can see from Figure 5, also in the case of ethanol in the gas phase, we obtain good overall agreement with the experimental spectrum. The shapes, intensities, and relative positions of the different bands are reproduced well. The two strong peaks at ∼1100 cm^{-1} correspond to stretches of the C−O bond, while the double bond C═O signature

DOI: 10.1021/acs.jctc.5b00391
J. Chem. Theory Comput. 2015, 11, 3321−3332

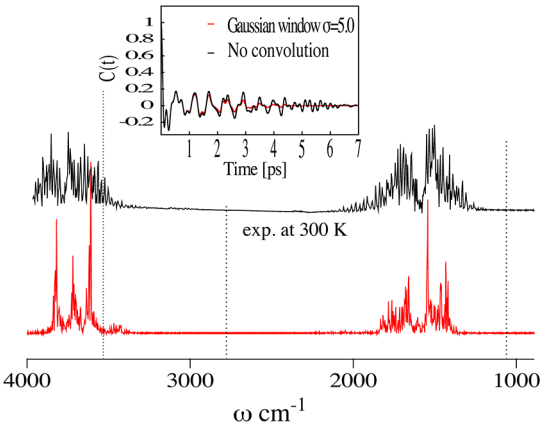


Figure 3. IR spectrum of water in the gas phase at 300 K calculated in ONETEP through NVE-BOMD simulations, using the Kubo quantum correction factor (red trace). Experimental IR spectrum of water in the gas phase at 300 K from ref 52 (black trace). Frequencies from normal-mode analysis obtained with ONETEP (denoted by dotted black vertical lines). Inset shows the autocorrelation function of the total dipole moment of water in vacuum, with and without Gaussian convolution.

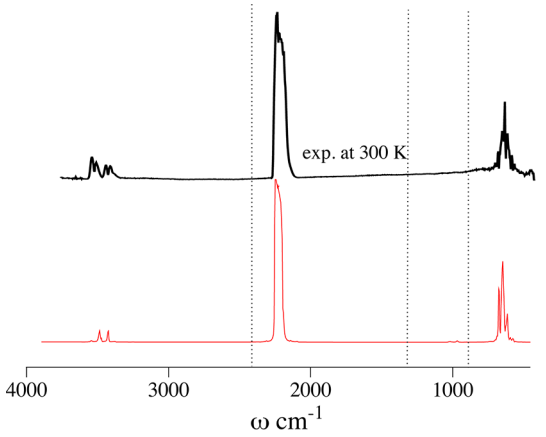


Figure 4. IR spectrum of carbon dioxide in the gas phase at 300 K computed through NVE-BOMD simulations in ONETEP, using the Kubo quantum correction factor (red trace). Experimental IR spectrum of CO_2 in the gas phase at 300 K from ref 52 (black trace). Frequencies from normal-mode analysis obtained with ONETEP are denoted by dotted black vertical lines).

corresponds to the bands in the region at ~1600 cm^{-1}. The broad band in the 2800−3100 cm^{-1} region is known to be characteristic of C−H bond stretching. Since ethanol has five C−H bonds, no more than five bands should present in the harmonic approximation results. Nonetheless, the experimental spectrum, as well as the anharmonic one, can show more than five bands, because of conformational changes (*trans-* and *gauche-* conformers) and Fermi resonance effects.[54] However, the correct assignment of fine vibrational structure is outside the scope of this work. It is worth mentioning that, in the IR spectrum obtained from the BOMD calculation, the position of the O−H stretch mode is observed at ω = 3500 cm^{-1}, whereas, in the experimental spectrum, it is observed at ~3645 cm^{-1}. Intensity and broadness are also reduced. This is possibly a consequence of using a larger time step, which affects the high-frequency modes more. Interestingly, this band is completely missing in the spectrum that was computed from the AMBER calculation with the GAFF force field with all SHAKE

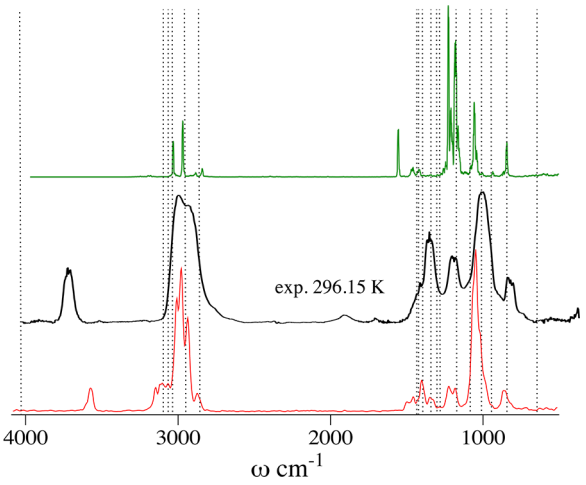


Figure 5. IR spectrum of ethanol in the gas phase at 296.15 K obtained from NVE-BOMD simulations in ONETEP, using the Kubo quantum correction factor (red trace). Experimental absorption IR spectrum of ethanol in the gas phase at 296.15 K from ref 52 (black trace). Frequencies from normal-mode analysis generated in ONETEP are denoted by dotted black vertical lines. Also shown is an AMBER-generated IR spectrum for ethanol in the gas phase at 296.15 K (green trace).

constraints removed (see Figure 5). This reinforces our statement that spectra computed at finite temperatures from an accurate molecular dynamics trajectory yield a better description of the absorption bands, compared to classical simulations or calculation with the harmonic approximation.

Because of the short length of the BOMD simulation, we are only able to sample the fast vibrational modes of 2-Ala. In fact, we are not able to properly compute a spectrum in the near-infrared range. Instead, we focused on the 1000−2000 cm^{-1} range of the infrared spectrum, as it shows the contribution of Amide I−III bands and it is the domain typically used to assess the structural properties of peptides. In Figure 6, we show the IR spectra calculated from the BOMD simulation in ONETEP

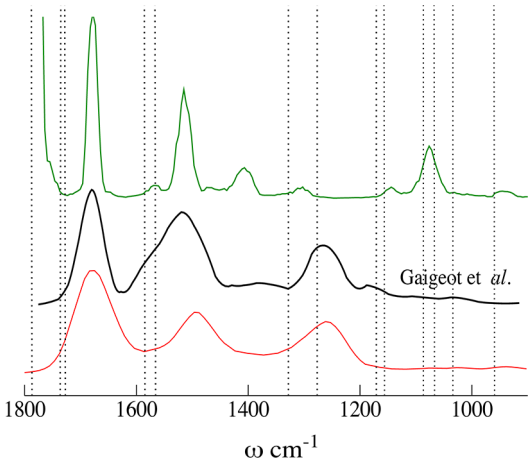


Figure 6. IR spectrum of 2-Ala in the gas phase at 300 K computed from NVE-BOMD simulations in ONETEP (red trace). Computed IR spectrum of 2-Ala in the gas phase at 300 K from classical MD simulation in AMBER (green trace). Frequencies from normal-mode analysis as obtained from ONETEP are denoted by dotted black vertical lines. Also shown is the IR spectrum of 2-Ala in the gas phase from CPMD simulations (results from ref 49) (black trace).

DOI: 10.1021/acs.jctc.5b00391
J. Chem. Theory Comput. 2015, 11, 3321−3332

and from classical MD in AMBER in vacuum. The three bands appearing in the spectrum obtained from BOMD, relative to Amide I at ∼1675 cm^{-1}, Amide II at ∼1515 cm^{-1}, and Amide III at ∼1263 cm^{-1} are in good agreement with the bands reported in refs 8 and 49, which are observed at ∼1670−1680 cm^{-1}, ∼1257 cm^{-1}, and ∼1210 cm^{-1}, respectively, even for the shorter simulation length employed here. Interestingly, the IR spectra in ref 8 were computed using Car−Parrinello MD, with the need of a fictitious electronic mass and consequent shifting of the final spectra. As explained in refs 8 and 49, of these three bands, the shape and position of the Amide I band appears to be a good candidate as the vibrational signature for 2-Ala at 300 K, when compared to larger and more-complex alanine peptide chains.

6.1. 2-Ala in the Aqueous Phase. The spectra computed with the DKP and DMA methods for 2-Ala in water are shown in Figure 7. The spectrum from classical simulation is also

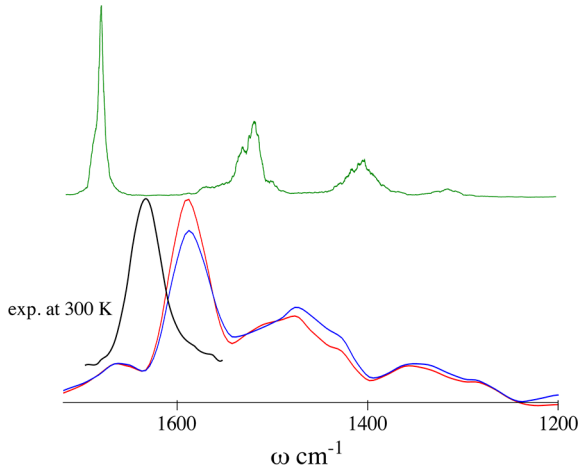


Figure 7. IR spectrum of 2-Ala in the aqueous phase at 300 K, obtained from NVE-BOMD simulations in Onetep with the DKP method (red trace) and DMA method (blue trace). IR spectrum of 2-Ala in AMBER (green trace). Experimental spectrum at 300 K of 2-Ala in deuterated water (D_2O) has been taken from ref 48.

shown for reference. DKP and DMA methods are in good agreement, especially in the 1200−1400 cm^{-1} range. In the region between 1400 cm^{-1} and 1700 cm^{-1}, the shapes and positions of the bands are the same, but not the intensities. This is likely to be due to the different way the density is partitioned in the two methods. In our simulation, the 2-Ala explores mostly the unfolded β and P_{II} conformations, since the folded α_R exhibited a slightly higher free energy (∼1 kcal mol^{-1}) at room temperature, leading to a preference for P_{II}/β population. Consequently, a single peak in the region of ∼1590−1600 cm^{-1} is found. This is in good agreement with that found by Gaigeot et al.[50] The Amide II band is blue-shifted and merged into one broader Amide I−II band. This property of folded 2-Ala in water makes the Amide I−II band featureless.[50] The experimental spectrum also shows a single peak at ∼1635 cm^{-1}. However, the solvent used in the experiment is D_2O to avoid vibrational mixing between H_2O bending modes with Amide I bands. In fact, mixing between Amide I and II modes and the bending modes of H_2O will broaden the Amide bands. The difference between the computed spectra and the experimental one can be potentially explained by the difference in solvents used. Moreover, a non-negligible role is also played by the exchange-correlation functional used. It is known that GGA functionals, such as PBE, perform quite well for water in the gas phase, whereas they tend to underestimate the binding energy in the aqueous phase, leading to an enhancement of the local structure.[55] This unsatisfactory property of GGAs could be due to a poor description of hydrogen bonds, which can be mitigated with functionals that inherently include dispersion.[56] In our case, the dispersion interactions are taken into account through the D2 Grimme correction.[43,44] The "softness" of PBEs in describing vibrational modes affects mostly the intermolecular modes, which in turn affects the 2-Ala-water coupling and consequently the positions of Amide bands. Finally, the combination of trajectories from folded and unfolded configurations will also improve the Amide I−II band position and shape.[50] The strong hydrogen bonds formed between 2-Ala and its surroundings yield to a red-shift of the position of Amide I band of ∼80.0 cm^{-1}, with respect to 2-Ala in the gas phase. On the other hand, the Amide II band is blue-shifted by ∼50 cm^{-1}. The merging of the Amide I and Amide II into one single peak is a clear effect of conformational changes and couplings with the solvent, which strongly affect the final spectrum.

7. CONCLUSIONS

We have presented two methods to compute the dipole moment of selected atoms, such as single molecules, in a larger environment, such as the solvent. Out methods are based on density kernel partitioning (DKP) and distributed multipole analysis (DMA), respectively. Both methods have been implemented in Onetep. Combining our methods with the Fourier transform time-correlation function (FTTCF) formalism and Born−Oppenheimer molecular dynamics, we are able to compute fully anharmonic infrared spectra of molecules in the gas phase and in solution at finite temperature. With this approach, conformational dynamics and anharmonic effects (anharmonicity of the potential energy surface) are naturally taken into account within the molecular dynamics simulation without the need of any *a posteriori* correction. Moreover, for systems in solution, only the spectrum of the solute is computed while retaining the effect of the vibrational coupling of the modes of the solute with the surroundings. This feature allows the rigorous study of molecules in solution, providing a formidable tool for the understanding and the assignment of bands from experimental spectra.

The theory for the FTTCF formalism is derived from first-order time-dependent quantum perturbation theory. We stress the point that the only approximation made here, from a theoretical point of view, consists of treating the nuclei classically. This approximation leads to the detailed balance condition symmetry problem that is partially recovered by introducing a theoretically derived quantum correction factor.

Several NVE-BOMD simulations at finite temperature on benchmark molecules in the gas phase have been performed. For all of the systems, very good agreement was found; when compared with the corresponding experimental spectra positions, the shapes and intensities of the bands were accurately reproduced. Note that, even for small molecules, the role played by the anharmonicity of the PESs at finite temperature is important. This is evident in the case of CO_2, where combination bands naturally emerge in our spectrum, whereas they are completely absent in the harmonic approximation. Generally, we found that spectra computed with our method are in much better agreement with the

DOI: 10.1021/acs.jctc.5b00391
J. Chem. Theory Comput. 2015, 11, 3321−3332

experiments than those computed through normal-mode analysis or a classical force field.

Finally, we performed an NVE-BOMD simulation to validate the DKP and DMA methods on the 2-Ala peptide in 500 water molecules (i.e., 1522 atoms). The spectrum for 2-Ala was compared to similar calculations from Gaigeot et al., obtained by performing CPMD with maximally localized Wannier functions for the calculation of the dipole,[8,49] and good agreement was found with these methods and with the experiment.

Schemes for accelerating BOMD in ONETEP, such as the second generation Car–Parrinello scheme,[57] are under investigation. These schemes are promising for extending the timescale of the dynamics, by reducing the number of self-consistency cycles while retaining time reversibility and energy conservation. By exploiting fast BOMD and linear scaling, computational spectroscopy based on DFT-MD can be made very competitive, even for large systems.

AUTHOR INFORMATION

Corresponding Author

*E-mail: C.Skylaris@soton.ac.uk.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank Professor Marie-Pierre Gaigeot for useful discussions at the start of this work. V.V. acknowledges the support of EPSRC the Doctoral Training Centre of the Institute for Complex System Simulations (Grant No. EP/G03690X/1). J.D. acknowledges the support of the Engineering and Physical Sciences Research Council (EPSRC Grant Nos. No EP/J015059/1 and EP/K039156/1) and of the Polish Ministry of Science and Higher Education (Grant No. IP2012 043972). The calculations in this work were carried out on the Iridis4 supercomputer of the University of Southampton, on the ARCHER national facility (access via the UKCP consortium, EPSRC Grant No. EP/K013556/1), and on the Galera Supercomputer at the TASK Computer Centre (Gdansk, Poland).

REFERENCES

(1) McQuarrie, D. *Statistical Mechanics*; University Science Books: Sausalito, CA, 2000; Chapter 21, pp 467–495.

(2) Zwanzig, R. Time-Correlation Functions and Transport Coefficients in Statistical Mechanics. *Annu. Rev. Phys. Chem.* **1965**, *16*, 67–102.

(3) Yagasaki, T.; Ono, J.; Saito, S. Ultrafast energy relaxation and anisotropy decay of the librational motion in liquid water: A molecular dynamics study. *J. Chem. Phys.* **2009**, *131*, 164511.

(4) DeVane, R.; Space, B.; Jansen, T. I. C.; Keyes, T. Time correlation function and finite field approaches to the calculation of the fifth order Raman response in liquid xenon. *J. Chem. Phys.* **2006**, *125*, 234501.

(5) Chandrasekhar, I.; Clore, G. M.; Szabo, A.; Gronenborn, A. M.; Brooks, B. R. A 500 ps molecular dynamics simulation study of interleukin-1β in water: Correlation with nuclear magnetic resonance spectroscopy and crystallography. *J. Mol. Biol.* **1992**, *226*, 239–250.

(6) Terranova, Z. L.; Corcelli, S. A. Molecular Dynamics Investigation of the Vibrational Spectroscopy of Isolated Water in an Ionic Liquid. *J. Phys. Chem. B* **2014**, *118*, 8264–8272.

(7) Thomas, M.; Brehm, M.; Fligg, R.; Vohringer, P.; Kirchner, B. Computing vibrational spectra from *ab initio* molecular dynamics. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6608–6622.

(8) Gaigeot, M.-P. Theoretical spectroscopy of floppy peptides at room temperature. A DFTMD perspective: Gas and aqueous phase. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3336–3359.

(9) Kaminski, S.; Gaus, M.; Phatak, P.; von Stetten, D.; Elstner, M.; Mroginski, M. A. Vibrational Raman Spectra from the Self-Consistent Charge Density Functional Tight Binding Method via Classical Time-Correlation Functions. *J. Chem. Theory Comput.* **2010**, *6*, 1240–1255.

(10) Schmitz, M.; Tavan, P. Vibrational spectra from atomic fluctuations in dynamics simulations. I. Theory, limitations, and a sample application. *J. Chem. Phys.* **2004**, *121*, 12233–12246.

(11) Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **1973**, *181*, 223–230.

(12) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. Introducing ONETEP: Linear-scaling density functional simulations on parallel computers. *J. Chem. Phys.* **2005**, *122*, 084119.

(13) Dziedzic, J.; Helal, H. H.; Skylaris, C.-K.; Mostofi, A. A.; Payne, M. C. Minimal parameter implicit solvent model for *ab initio* electronic-structure calculations. *Europhys. Lett.* **2011**, *95*, 43001.

(14) Dziedzic, J.; Fox, S. J.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. Large-scale DFT calculations in implicit solvent—A case study on the T4 lysozyme L99A/M102Q protein. *Int. J. Quantum Chem.* **2013**, *113*, 771–785.

(15) Todorova, N.; Makarucha, A. J.; Hine, N. D. M.; Mostofi, A. A.; Yarovsky, I. Dimensionality of Carbon Nanomaterials Determines the Binding and Dynamics of Amyloidogenic Peptides: Multiscale Theoretical Simulations. *PLoS Comput. Biol.* **2013**, *9*, e1003360.

(16) Tuckerman, M. E. *Statistical Mechanics: Theory and Molecular Simulation*; Oxford University Press: Cambridge, U.K., 2010; Chapter 14, pp 533–556.

(17) Ramírez, R.; López-Ciudad, T.; Kumar, P. P.; Marx, D. Quantum corrections to classical time-correlation functions: Hydrogen bonding and anharmonic floppy modes. *J. Chem. Phys.* **2004**, *121*, 3973–3983.

(18) Hernandez de la Peña, L. On the zero temperature limit of the Kubo-transformed quantum time correlation function. *Mol. Phys.* **2014**, *112*, 929–936.

(19) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(20) Kohn, W. Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms. *Phys. Rev. Lett.* **1996**, *76*, 3168–3171.

(21) Prodan, E.; Kohn, W. Nearsightedness of electronic matter. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11635–11638.

(22) Taraskin, S. N.; Drabold, D. A.; Elliott, S. R. Spatial Decay of the Single-Particle Density Matrix in Insulators: Analytic Results in Two and Three Dimensions. *Phys. Rev. Lett.* **2002**, *88*, 196405.

(23) Cloizeaux, J. D. Analytical Properties of *n*-Dimensional Energy Bands and Wannier Functions. *Phys. Rev.* **1964**, *135*, A698–A707.

(24) Ismail-Beigi, S.; Arias, T. A. Locality of the Density Matrix in Metals, Semiconductors, and Insulators. *Phys. Rev. Lett.* **1999**, *82*, 2127–2130.

(25) He, L.; Vanderbilt, D. Exponential Decay Properties of Wannier Functions and Related Quantities. *Phys. Rev. Lett.* **2001**, *86*, 5341–5344.

(26) Goedecker, S. Decay properties of the finite-temperature density matrix in metals. *Phys. Rev. B* **1998**, *58*, 3501–3502.

(27) McWeeny, R. Some Recent Advances in Density Matrix Theory. *Rev. Mod. Phys.* **1960**, *32*, 335–369.

(28) Skylaris, C.-K.; Mostofi, A. A.; Haynes, P. D.; Diéguez, O.; Payne, M. C. Nonorthogonal generalized Wannier function pseudopotential plane-wave method. *Phys. Rev. B* **2002**, *66*, 035119.

(29) Hine, N. D. M.; Robinson, M.; Haynes, P. D.; Skylaris, C.-K.; Payne, M. C.; Mostofi, A. A. Accurate ionic forces and geometry optimization in linear-scaling density-functional theory with local orbitals. *Phys. Rev. B* **2011**, *83*, 195102.

(30) Marzari, N.; Mostofi, A. A.; Yates, J. R.; Souza, I.; Vanderbilt, D. Maximally localized Wannier functions: Theory and applications. *Rev. Mod. Phys.* **2012**, *84*, 1419–1475.

DOI: 10.1021/acs.jctc.5b00391
J. Chem. Theory Comput. 2015, 11, 3321–3332

(31) O'Regan, D. D.; Payne, M. C.; Mostofi, A. A. Generalized Wannier functions: A comparison of molecular electric dipole polarizabilities. *Phys. Rev. B* **2012**, *85*, 193101.
(32) Stone, A. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.* **1981**, *83*, 233−239.
(33) Stone, A.; Alderton, M. Distributed multipole analysis. *Mol. Phys.* **1985**, *56*, 1047−1064.
(34) Cohen, M. H.; Wasserman, A. On the Foundations of Chemical Reactivity Theory. *J. Phys. Chem. A* **2007**, *111*, 2229−2242.
(35) Rein, R. On Physical Properties and Interactions of Polyatomic Molecules: With Application to Molecular Recognition in Biology. *Adv. Quantum Chem.* **1973**, *7*, 335−396.
(36) Stone, A. J. *GDMA: Distributed Multipoles from Gaussian98 Wavefunctions*, 2nd Edition; University of Cambridge: Cambridge, U.K., 1998.
(37) Stone, A. J. Distributed Multipole Analysis: Stability for Large Basis Sets. *J. Chem. Theory Comput.* **2005**, *1*, 1128−1132.
(38) Dziedzic, J.; Hill, Q.; Skylaris, C.-K. Linear-scaling calculation of Hartree−Fock exchange energy with non-orthogonal generalised Wannier functions. *J. Chem. Phys.* **2013**, *139*, 214103.
(39) Hill, Q. *Development of more accurate computational methods within linear-scaling density functional theory*. Ph.D. thesis; University of Southampton: Southampton, U.K., 2010.
(40) Haynes, P. D.; Payne, M. C. Localised spherical-wave basis set for *O*(*N*) total-energy pseudopotential calculations. *Comput. Phys. Commun.* **1997**, *102*, 17−32.
(41) Case, D.; Darden, T.; Cheatham, T., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A.; Kolossváry, I.; Wong, K.; Paesani, F.; Vanicek, J.; Wolf, R.; Liu, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D.; Mathews, D.; Seetin, M.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. *AMBER12*; University of California: San Francisco, CA, 2012.
(42) Wilkinson, K. A.; Hine, N. D. M.; Skylaris, C.-K. Hybrid MPI-OpenMP Parallelism in the ONETEP Linear-Scaling Electronic Structure Code: Application to the Delamination of Cellulose Nanofibrils. *J. Chem. Theory Comput.* **2014**, *10*, 4782−4794.
(43) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(44) Hill, Q.; Skylaris, C.-K. Including dispersion interactions in the ONETEP program for linear-scaling density functional theory calculations. *Proc. R. Soc. London Ser. A* **2009**, *465*, 669−683.
(45) Nunes, R. W.; Vanderbilt, D. Generalization of the density-matrix method to a nonorthogonal basis. *Phys. Rev. B* **1994**, *50*, 17611−17614.
(46) Millam, J. M.; Scuseria, G. E. Linear scaling conjugate gradient density matrix search as an alternative to diagonalization for first principles electronic structure calculations. *J. Chem. Phys.* **1997**, *106*, 5569−5577.
(47) Haynes, P. D.; Skylaris, C.-K.; Mostofi, A. A.; Payne, M. C. Density kernel optimization in the ONETEP code. *J. Phys.: Condens. Matter* **2008**, *20*, 294207.
(48) Kwac, K.; Lee, K.-K.; Han, J. B.; Oh, K.-I.; Cho, M. Classical and quantum mechanical/molecular mechanical molecular dynamics simulations of alanine dipeptide in water: Comparisons with IR and vibrational circular dichroism spectra. *J. Chem. Phys.* **2008**, *128*, 105106.
(49) Gaigeot, M.-P. Alanine Polypeptide Structural Fingerprints at Room Temperature: What Can Be Gained from Non-Harmonic Car−Parrinello Molecular Dynamics Simulations. *J. Phys. Chem. A* **2008**, *112*, 13507−13517.
(50) Gaigeot, M.-P. Infrared spectroscopy of the alanine dipeptide analog in liquid water with DFT-MD. Direct evidence for $P_{II}/[\beta]$ conformations. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10198−10209.
(51) Fox, S. J.; Pittock, C.; Tautermann, C. S.; Fox, T.; Christ, C.; Malcolm, N. O. J.; Essex, J. W.; Skylaris, C.-K. Free Energies of Binding from Large-Scale First-Principles Quantum Mechanical Calculations: Application to Ligand Hydration Energies. *J. Phys. Chem. B* **2013**, *117*, 9478−9485.
(52) Linstrom, P.; Mallard, W. *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; National Institute for Standards and Technology (NIST): Gaithersburg, MD (available via the Internet at: http://webbook.nist.gov, accessed April 2015).
(53) Lemus, R. Vibrational excitations in $\{H_2O\}$ in the framework of a local model. *J. Mol. Spectrosc.* **2004**, *225*, 73−92.
(54) Yu, Y.; Lin, K.; Zhou, X.; Wang, H.; Liu, S.; Ma, X. New C−H Stretching Vibrational Spectral Features in the Raman Spectra of Gaseous and Liquid Ethanol. *J. Phys. Chem. C* **2007**, *111*, 8971−8978.
(55) Santra, B.; Michaelides, A.; Scheffler, M. Coupled cluster benchmarks of water monomers and dimers extracted from density-functional theory liquid water: The importance of monomer deformations. *J. Chem. Phys.* **2009**, *131*, 124509.
(56) Kolb, B.; Thonhauser, T. van der Waals density functional study of energetic, structural, and vibrational properties of small water clusters and ice I_h. *Phys. Rev. B* **2011**, *84*, 045116.
(57) Kühne, T. D. Second generation Car−Parrinello molecular dynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 391−406.

DOI: 10.1021/acs.jctc.5b00391
J. Chem. Theory Comput. 2015, 11, 3321−3332