Quantifying Fluctuations of Average Solvent Environments for Embedding Calculations

Cristina E. González-Espinoza,[†] Christopher A. Rumble,[*,†,¶] Daniel Borgis,[‡] and Tomasz A. Wesolowski[*,†]

†*Université de Genève, Département de Chimie Physique 30, Quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland.*

‡*Ecole Normale Supérieure, Département de Chimie, UMR 8640 ENS-CNRS-UPMC, 24 Rue Lhomond, 75005 Paris, France*

¶*Current address: Department of Chemistry, The Pennsylvania State University, Altoona College, 3000 Ivyside Park, Altoona, Pennsylvania 16601, United States*

E-mail: crumble@psu.edu; tomasz.wesolowski@unige.ch

Abstract

In the context of employing embedding methods to study spectroscopic properties, the viability and effectiveness of replacing an ensemble of calculations by a single calculation using an average description of the system of study are evaluated. This work aims to provide a baseline of the expected fluctuations in the average description of the system obtained in the two cases: from calculations of an ensemble of geometries, and from an average environment constructed with the same ensemble. To this end, the classical molecular dynamics simulation of a very simple system was used: a rigid molecule of acetone in a solution of rigid water. We perform a careful numerical analysis of the fluctuations of the electrostatic potential felt by the solute, as well as the fluctuations in the effect on its electronic density, measure through the dipole moment and the atomic charges derived from the corresponding potential. At the same time, we inspect the accuracy of the methods used to construct average environments. Finally, the proposed approach to generate the embedding potential from an average environment density is applied to estimate the solvatochromic shift of the first excitation of

acetone. In order to account for quantum-confinement effects that may be important in certain cases, the fluctuations on the shift due to the interaction with the solvent are evaluated using Frozen-Density Embedding Theory. Our results demonstrate that, for normally-behaved environments, the constructed average environment is a reasonably good representation of a discrete solvent environment.

1 Introduction

Dealing with the electronic properties of molecules requires the use of quantum mechanics. Given the ubiquity of the condensed phase in laboratory chemistry, the calculation of electronic properties in condensed matter is indispensable for connecting theoretical calculations with experimental observables. The large size of such condensed phase systems makes the electronic-structure calculations computationally challenging and approximations are essential for maximizing computational efficiency. A commonly encountered approach for reducing the cost of condensed phase quantum calculations are multi–level methods where the full system is partitioned into two pieces treated with varying levels of complexity. The first partition is a small subsystem of interest, here termed fragment A, which is treated with high–level quantum chemical methods. This subsystem is then embedded within a surrounding environment, here called fragment B, to which various approximations and lower–level methods are applied to minimize computational cost.

An important application of multi-level methods is to the computational study of liquid-phase photochemistry, where a quantum mechanical chromophore is embedded in an approximation of the solvent environment. In such studies, one is typically interested in changes in solute electronic excitation energy brought about by the solvent environment, which is characterized by

$$\delta\nu_{\text{ex}} = \nu_{\text{emb}} - \nu_{\text{iso}}, \tag{1}$$

where ν_{iso} is the excitation energy of the isolated solute in vacuum, ν_{emb} is the excitation energy of the solute embedded in the solvent environment, and $\delta\nu_{\text{ex}}$ the 'solvent shift' of the excitation energy. The quality of the excitation energy calculation is usually evaluated by comparing ν_{emb} to the maxima of measured electronic absorption spectra, whereas the shift is compared to the shift in the chromophore's absorption maximum between the solvent of interest and the maximum in a non-polar solvent such as n-hexane, where the non-polar solvent measurement acts as a stand–in for the more demanding gas–phase measurements. Single-reference quantum chemical methods, such as time-dependent Density Functional

Theory (TDDFT),[1,2] configuration interaction (CI),[3,4] equation of motion coupled-cluster (EOM-CC),[5] among others, are used to model the solute electronic transitions, and a number of different approximations can be employed to describe the solvent environment.

One method for treating the solvent is to determine a single *average solvent environment* in which to embed the chromophore. This approach has the obvious benefit of requiring only two calculations for determining $\delta\nu_{\mathrm{ex}}$: one for ν_{iso} and one for ν_{emb}. In the polarizable contiuum model (PCM),[6,7] the environment is described as a polarizable dielectric continuum. In the simplest implementation of PCM, the solvent is modeled as a linear isotropic continuum characterized by the static dielectric constant, ϵ, of the bulk solvent.[6] In more sophisticated formulations, the solute cavity is given a molecular shape and the reaction field is described in terms of polarization charges or reaction field factors included in the solute Hamiltonian. This approach allows for self-consistency to be reached between the solute wavefunction and the solvent polarization by means of iterative procedures. In particular, the conductor-like variant of PCM (C-PCM),[8] based on the COSMO methods,[9] has been become very popular due to its ease–of–use, low computational cost, reasonable agreement with experiment, and widespread implementation in quantum chemistry software packages.

Instead of embedding the solute in a dielectric continuum, one can determine an effective solvent potential which acts on the solute. If this potential is constructed using only solvent electrostatics it is called the 'averaged solvent electrostatic potential' (ASEP).[10] Such a potential was constructed in Ref. 10 by placing point-charges at grid points to approximate the average solvent configuration obtained from molecular dynamics (MD) simulations. The values of the charges are fitted to reproduce the ASEP at the position of the solute by means of the ChELPG procedure,[11] while the position and the number of point-charges is optimized. A second method for constructing the ASEP utilizes scaled charges in all the positions of solvent atoms generated by an MD simulation, as was done by Coutinho *et al.*[12] These workers constructed an 'average single configuration' using 60 000 scaled charges from 100 superposed configurations of 200 water molecules centered around a single acetone solute. An important aspect of this procedure, termed as ASEC, is the selection of the configurations, which ensures convergent results by using statistically uncorrelated configurations that extend to outer layers of the solvent. The ASEC method has been used to model the effect of the solvent on spectroscopic properties,[13] such as NMR shielding parameters, absorption spectra,[14,15] electronic hyperpolarizabilities,[16,17] and more sophisticated protocols to optimize the free energy of the full system.[18]

Another category of continuum models of the solvent is based on Frozen-Density Embedding Theory (FDET).[19] In these methods, the embedding potential includes a non-linear term in the solute electron density, which is responsible for describing the electronic exchange-

correlation and Pauli repulsion between the fragments. Previous studies done with FDET–based methods have used the average solvent density to describe solvatochromic shifts for aminocoumarin, coumarin 153 and acetone in several solvents.[20–24]

Comparisons between these average environment calculations and experimental spectra are, on the surface, simple and straight–forward. Two single–point calculations are performed to determine $\delta\nu_{\text{ex}}^{(\text{calc})}$, two absorption spectra measured to determine $\delta\nu_{\text{ex}}^{(\text{exp})}$, and the resulting maxima and shifts compared. In reality, condensed phase spectra are not the result of a single chromophore in a static average environment, but of a fluctuating distribution of flexible chromophore conformations in an ensemble of solvent configurations, each contributing a distinct excitation energy to the observed spectrum. The resulting distribution of excitations, in conjunction with other factors such as vibronic transitions and finite excited state lifetimes, introduces substantial broadening (on the order of 0.34–0.40 eV for coumarin 153, a common solvatochromic probe, in dipolar liquids)[25] and asymmetry to the measured spectra. The broadening and asymmetry complicates determination of the experimental $0\rightarrow0$ solute transitions for comparison with theory, hence the use of the spectral maximum as a proxy. The single–point nature of average environment methods does not allow for straight–forward determination of spectral broadening and makes it difficult to evaluate differences between $\delta\nu_{\text{ex}}^{(\text{calc})}$ and $\delta\nu_{\text{ex}}^{(\text{exp})}$.

One way for accessing these fluctuations is to use molecular dynamics (MD) simulations (or Monte Carlo sampling[26,27]) to generate ensembles of discrete solute/solvent configurations on which to perform an ensemble of single–point quantum mechanical (QM) calculations.[28,29] Due to the large number of solvent nuclei needed for simulating realistic dynamics, the solvent cannot be treated explicitly, and approximate descriptions of the discrete solvent environment are needed to make the QM computations tractable. The most simplest method for describing a discrete solvent environment is point–charge embedding (PCE), where the solute is embedded the electric field generated by the surrounding solvent atoms. In order to capture the effects due to mutual polarization or other non-electrostatic interactions, the assignment of charges values for the environment and the embedding process can be made self-consistent. Examples of such polarizable embedding methods[30] are the ASEP-MD methods[31–33] , the Effective Fragment Potential (EFP)[34] and X-Pol.[35] An alternative to electrostatic–only embedding is FDET, where the two independent variables, namely: an auxiliary N_A-electrons wavefunction Ψ_A and the electron density of the environment $\rho_B(\mathbf{r})$, are both generated from the individual solute/solvent configurations.[36,37]

These *ensemble* methods provide access to the inhomogeneous broadening of spectra, *i.e* broadening due to the distribution of solute/solvent configurations, and give some insight into the range of $\delta\nu_{\text{ex}}$ expected in the system that single–point calculations are unable to provide.

These approaches are still not perfect, as most applications do not realistically account for homogenous broadening effects such as the aforementioned vibronic transitions and finite excited state lifetimes. Even though they provide valuable information on fluctuations, ensemble calculations can become prohibitively expensive when the embedded system is large or many solute/solvent configurations must be considered, necessitating a fall–back to single–point approaches.

The goal of the present study is assess the validity of substituting a single averaged solvent environment for an ensemble of environments. We wish to determine the fluctuations and uncertainties in properties derived from such calculations, and provide guidance on how to realistically construct an average environment. In essence, we wish to test whether or not an averaged environment can capture the features of an ensemble of calculations, an approximation we summarize as:

$$\left\langle \hat{O}\big[\Psi_A(\mathbf{R}_A)[B]\big] \right\rangle \approx \hat{O}\big[\Psi_A(\bar{\mathbf{R}}_A)[\bar{B}]\big]. \tag{2}$$

In this expression, we define $\hat{O}[\Psi]$ as the quantum mechanical expectation value, $\langle\Psi|\hat{O}|\Psi\rangle$ of some system property described by $\hat{O}$. On the left–hand side, we take $\Psi_A(\mathbf{R}_A)[B]$ to be the wavefunction of fragment A, with instantaneous nuclear coordinates $\mathbf{R}_A$, embedded in instantaneous environment B, which could be described using any of the previously discussed methods. The $\langle\rangle$ indicates averaging over an ensemble of realizations of $\mathbf{R}_A$ and B, and we will refer to this quantity as the *ensemble average* of $\hat{O}$. On the right–hand side we have the same operator $\hat{O}$, but instead of operating on every instantaneous realization of $\mathbf{R}_A$ and B, $\hat{O}$ is evaluated once for a single wavefunction, $\Psi_A(\bar{\mathbf{R}}_A)[\bar{B}]$. Here, the instantaneous coordinates $\mathbf{R}_A$ and environment B are replaced by an averaged set of subsystem coordinates $\bar{\mathbf{R}}_A$ embedded in the averaged environment $\bar{B}$. This quantity will be referred to as the *environmental average* of $\hat{O}$. The quantities $\bar{\mathbf{R}}_A$ and $\bar{B}$ may be ensemble averages, but are not required to be, as would be the case in a method such as PCM.

If Eq. 2 is a reasonable approximation, we expect the environmental averages to fall well within the fluctuations of the ensemble calculations and be reasonably close to the actual ensemble average. To evaluate Eq. 2, we have performed classical MD simulations of the acetone (ACE) and water solute/solvent system in order to generate a realistic set of solute/solvent configurations for determining ensemble and environmental averages of various system properties as well as the first electronic transition of ACE. We chose to use the ACE/water system because the first excitation of ACE in water has been used extensively by our group and others as a benchmark for solvation models.[20,24,38–40] The ensemble averages will be calculated from N independent MD frames, which can be written using the notation

of Eq. 2 as:

$$\left\langle \hat{O}\big[\Psi_A(\mathbf{R}_A)[B]\big] \right\rangle = \frac{1}{N}\sum_{i=1}^{N} \hat{O}\big[\Psi_A(\mathbf{R}_A)[B_i]\big] = \langle O \rangle_N, \tag{3}$$

where B_i represents the environment of MD frame i out of a total of N frames, and $\langle O \rangle_N$ a shorthand for the ensemble average which we will use for the remainder of this work. We consider the ensemble average the 'true' average of $\hat{O}$ and will characterize its fluctuations using the standard deviation, σ:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \Big(\langle O \rangle_N - \hat{O}\big[\Psi_A(\mathbf{R}_A)[B_i]\big]\Big)^2}. \tag{4}$$

In order to keep our analysis focused on Eq. 2 as closely as possible, all molecules will be simulated with fixed molecular geometries. This allows us to replace $\bar{\mathbf{R}}_A$ in Eq. 2 with a single $\mathbf{R}_A$ for the static ACE geometry and all the fluctuations we observe can be attributed to reorientation of the solvent environment. Complicating factors such as chromophore and solvent flexibility will be left for future studies. We then express the environmental averages as

$$\hat{O}\big[\Psi_A(\bar{\mathbf{R}}_A)[\bar{B}]\big] = \hat{O}\big[\Psi_A(\mathbf{R}_A)[\bar{B}]\big] = \bar{O}_N, \tag{5}$$

with $\bar{O}_N$ being our abbreviation for the environmental averages. The various strategies used for determining $\bar{B}$ will be discussed as they appear in the text.

In our statistical analysis of Eq. 2, we will consider the ensemble and environmental averages of four different system properties that are influenced by their solvent environment. The most fundamental of these is the electrostatic potential (ESP) generated by the solvent, which we study using the ESP evaluated at the nuclei of the ACE solute. We will then move on to three properties derived from the ESP: the atomic point-charges and dipole moments of ACE embedded in the solvent ESP determined using the ChELPG method,[11] and the first electronic transition of ACE. In each case we will compare the ensemble averages and their fluctuations with the single value generated by an averaged environment calculation. As mentioned previously, the fluctuations we observe will be a consequence of only the rotation and translation of solvent molecules due to our use of rigid solute and solvent molecules in our simulations.

Our results suggest that the ensemble averages of these four properties are well– re-

produced by our environmental averaging methods. Although the average properties of the system are captured, properties of a single solvent configuration can become strongly decorrelated when different ESP approximations are applied to the system. We therefore advise caution when comparing the quality of calculations performed on a single solvent configuration under differing sets of approximations, as these differences between different approximations on the same configuration can become much larger than the differences between the ensemble and environmental averages. We conclude that that Eq. 2 is a valid approximation when the fluctuations of the properties in question are normally distributed, as we observed for the rigid solute/solvent simulations performed here.

2 Methods

2.1 The Embedding Energy

The embedded quantum system A (the solute) is characterized by the position of its M nuclei, $\mathbf{R}_A$, and their momentum, $\mathbf{p}_A$. Any instantaneous configuration of the mobile environment B (the solvent) is characterized by the position of all its atoms, $\mathbf{R}_B$, and their momentum, $\mathbf{p}_B$. The energy of the total $A + B$ system can be written within the Born-Oppenheimer approximation as

$$\begin{aligned} E(\mathbf{R}_A, \mathbf{R}_B, \mathbf{p}_A, \mathbf{p}_B) =& K(\mathbf{p}_A, \mathbf{p}_B) + V_B(\mathbf{R}_B) \\ &+ E_{\mathrm{emb}}[v_{emb}](\mathbf{R}_A, \mathbf{R}_B) \end{aligned} \tag{6}$$

with $K(\mathbf{p}_A, \mathbf{p}_B)$ being the classical kinetic energy of the nuclei, $V_B(\mathbf{R}_B)$ is the self-energy of fragment B, and $E_{\mathrm{emb}}[v_{emb}](\mathbf{R}_A, \mathbf{R}_B)$ is the energy of embedding A in B. Note that from now on, the dependence on the nuclear coordinates $\mathbf{R}_A$ and $\mathbf{R}_B$ will be kept implicit.

To be consistent, the integration of the solvent equations of motion should be done with the full Hamiltonian, including both QM and molecular mechanics (MM) parts.[41] Additionally, one should thus include the true AB quantum mechanical coupling using the full electronic density ρ_A. This could be done in future studies, but for the time being we apply a simplified strategy. Instead of following the standard QM/MM approaches where both levels are evaluated sequentially or simultaneously with the proper partitioning of the MM and QM energies,[41] we will perform classical MD simulations first, followed by QM calculations on the structures generated by classical simulations. As a consequence, the kinetic energy of the nuclei will be 0, and the total energy will depend only on V_B and $E_{\mathrm{emb}}[v_{emb}]$. A proper way to make the statistics generated this way compatible with the appropriate statistics in

the canonical ensemble will be given in Section 3.1.4.

Here, we suppose that the environmental self-energy term, V_B, can be reduced to a molecular mechanics force field, *i.e*, $V_B = V_B^{MM}$. Furthermore, to compute the solute-solvent electrostatic interaction, the electron density carried by each solvent molecule is gathered with their nuclei point charges to yield a simplified, effective point charge representation $\hat{\sigma}_B(\mathbf{r})$, that given by the MM force field, *i.e.*,

$$\hat{\sigma}_B(\mathbf{r}) = \sum_{j=1}^{N_{\text{solv}}} q_j \, \delta(\mathbf{r} - \mathbf{R}_{B,j}), \tag{7}$$

where the hat symbol designates discrete distributions. Then, the embedding potential is defined as the electrostatic interaction with these discrete charges

$$v_{emb}(\mathbf{r}) = v_{elst}(\mathbf{r}) = \int d\mathbf{r}' \frac{\hat{\sigma}_B(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \tag{8}$$

which yields the only-electrostatics embedding energy

$$\begin{aligned} E_{\text{emb}} =& H_A + V_{elst} \\ =& \left\langle \Psi_A | \hat{H}_A | \Psi_A \right\rangle \\ &+ \int\int d\mathbf{r} d\mathbf{r}' \, \frac{(\hat{\rho}_{N_A}(\mathbf{r}) + \rho_A(\mathbf{r}))\hat{\sigma}_B(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \end{aligned} \tag{9}$$

with the solute nuclei charge density given by $\hat{\rho}_{N_A}(\mathbf{r}) = \sum_{k=1}^{M} Z_k \, \delta(\mathbf{r} - \mathbf{R}_{A,k})$.

2.2 Frozen-Density Embedding Theory

The ultimate goal is to go beyond electrostatic-only embedding to describe the interaction between fragment A and B, especially when studying electronic properties, as quantum-confinement effects may play an important role. For that, we turn into methods based on Frozen-Density Embedding Theory (FDET).[19]

Contrary to the electrostatics-only case, the embedding potential,

$$v_{\text{emb}}[\rho_A, \rho_B; v_{\text{elst}}](\mathbf{r}) = v_{\text{elst}}(\mathbf{r}) + v_{xcT}^{\text{nad}}[\rho_{\text{A}}, \rho_{\text{B}}](\mathbf{r}), \tag{10}$$

due to the non-additive exchange-correlation and kinetic term, $v_{xcT}^{nad}[\rho_A, \rho_B](\mathbf{r})$, is not homogeneous with the density ρ_A, *i.e.*, the energy is not equal to the expectation value of the potential. The embedding energy must then be evaluated with Ψ_A^{FDET}, the variational

solution of the equation

$$\left(\hat{H}_A + v_{emb}\right)\Psi^{\mathrm{FDET}} = \lambda\Psi^{\mathrm{FDET}}. \tag{11}$$

The FDET embedding potential given in Eq. 10 and the total energy ($E^{\mathrm{FDET}}_{v_{AB}}[\Psi^{\mathrm{FDET}}, \rho_B]$) satisfy by construction the condition of self-consistency with the Hohenberg-Kohn energy functional ($E^{HK}_{v_{AB}}[\rho]$),[42,43] meaning that

$$E^{\mathrm{FDET}}_{v_{AB}}[\Psi^{\mathrm{FDET}}, \rho_B] = E^{HK}_{v_{AB}}[\rho^{\mathrm{FDET}}_A + \rho_B] \tag{12}$$

where ρ^{FDET}_A is the density corresponding to the wavefunction obtained in Eq. 11, for any admissible $\rho_B(\mathbf{r})$.

2.2.1 FDET interpretation of Eq. 2

In FDET, the expectation value of any quantum mechanical operator evaluated at the geometry $\{\mathbf{R}_A\}, \{\mathbf{R}_B\}^i$ is a functional of ρ_{B_i}:

$$\begin{aligned}\left\langle \Psi^{\mathrm{FDET}(\mathbf{R}_A,\mathbf{R}^i_B)}_A[\rho_{B_i}]\middle|\hat{O}\middle|\Psi^{\mathrm{FDET}(\mathbf{R}_A,\mathbf{R}^i_B)}_A[\rho_{B_i}]\right\rangle \\ \equiv \langle \hat{O}_{\mathbf{R}_A,\mathbf{R}^i_B}\rangle[\rho_{B_i}]\end{aligned} \tag{13}$$

where $\Psi^{\mathrm{FDET}(\mathbf{R}_A,\mathbf{R}^i_B)}_A[\rho_{B_i}]$ denote the optimal embedded FDET wavefunction for a given $\rho_{B_i}(\mathbf{r})$ and nuclear positions $\{\mathbf{R}_A\}, \{\mathbf{R}_B\}^i$. The index i here indicates the i-th configuration of fragment B, from an ensemble of N geometries. The square brackets in $\Psi^{\mathrm{FDET}(\mathbf{R}_A,\mathbf{R}^i_B)}_A[\rho_{B_i}]$ indicate that, up to the unitary transformation, the FDET embedded wavefunction is uniquely determined by ρ_{B_i} for a given $\{\mathbf{R}_A\}, \{\mathbf{R}_B\}^i$.

Therefore, $\bar{B}$ from Eq. 2 is defined in FDET by the set of $\{\mathbf{R}_B\}^i$ and ρ_{B_i}, which determine the local fields (functions of $\mathbf{r}$ in 3D), $\langle\rho_B\rangle(\mathbf{r})$ and $\langle\rho^{nuc}_B\rangle(\mathbf{r})$. The fields represent the probability to find the negative (electrons) or positive (nuclei) charge in a given position $\mathbf{r}$, respectively. The field $\langle\rho^{nuc}_B\rangle(\mathbf{r})$ can be easily transformed to the field of the corresponding electrostatic potential which is denoted with $\langle v^{nuc}_B(\mathbf{r})\rangle$.

Hence, equation 2 takes the following form:

$$\left\langle\langle \hat{O}_{\mathbf{R}_A,\mathbf{R}^i_B}\rangle[\rho_{B_i}]\right\rangle_N \approx \langle \hat{O}_{\mathbf{R}_A,\langle v^{nuc}_B\rangle}\rangle[\langle\rho_B\rangle] \tag{14}$$

The approximation in Eq. 14 has been used previously in combination with 3D-RISM and MD simulations to study solvatochromism in absorption[20,21,24] and emission.[23] For the sake of this publication, we unify the notation and rewrite this key approximation in the same format used in the introduction. In this particular case, as mentioned before, we use a fixed

geometry of the environment, which means that $\Psi_A(\mathbf{R}_A) = \Psi_A(\bar{\mathbf{R}}_A)$, so equation 14 can be expressed as:

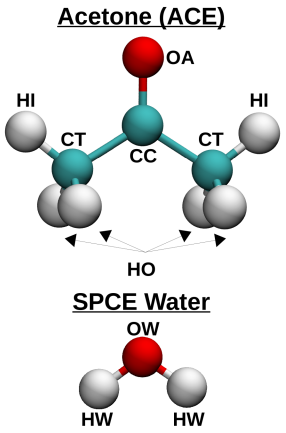
$$\left\langle \hat{O}\big[\Psi_A(\mathbf{R}_A)[\rho_B]\big] \right\rangle \approx \hat{O}\big[\Psi_A(\bar{\mathbf{R}}_A, \bar{v}_B^{nuc})[\bar{\rho}_B]\big] \tag{15}$$

2.3 Computational Technical Details

2.3.1 Molecular Dynamics

All classical molecular dynamics simulations were performed using DL_POLY_4.10.[44,45] The intermolecular interactions were treated using a simple Lennard-Jones plus Coulomb force-field, and the simulated system consisted one acetone (ACE) molecule dissolved in 1000 SPCE waters.[46] The ACE structure and point charges were taken from Aidas *et. al*[39] and the Lennard-Jones parameters from the ketone and methyl group parameters of the OPLS-AA force-field.[47] Lennard-Jones parameters for cross-species interactions were calculated using the Lorentz-Berthelot mixing rules. Electrostatics were handled using an Ewald summation and electrostatics and Lennard-Jones forces were cutoff at 15 Å. Trajectories were integrated using the velocity Verlet algorithm with a 2 fs timestep. All molecules were held rigid using the SHAKE algorithm with the default tolerance settings of DL_POLY_4.10.

Initial conformations were generated using the `fftool` and `PACKMOL`[48] packages. The systems contained one ACE and 1000 SPCE molecules in an initial 32 Å box with cubic periodic boundary conditions. The following equilibration procedure was then followed. First, forces were minimized using the conjugate gradient method followed by a 25 ps simulation in the NVT ensemble. The temperature was 298.15 K and controlled using a Nosé–Hoover thermostat with a relaxation time of 0.5 ps. Next, a 50 ps simulation in the NPT ensemble was conducted in order to relax the system volume using a Nosé–Hoover barostat with 1 atm of pressure a relaxation time of 1.5 ps. Following the NPT simulation, the volume of the system was then fixed followed by a final 25 ps equilibration simulation in the NVT ensemble. A production simulation of 5 ns was then performed with frames saved every 1 ps, resulting in 5000 frames for later analysis. The production frames were then centered and wrapped around the ACE center-of-mass using the MDAnalysis package.[49] All images generated from the MD simulations were created using the VMD 1.9.3 package[50] and the Tachyon ray tracing library.[51]



Scheme 1: Space-filling models and atom labels of all simulated systems.

2.3.2 Embedding Calculations

For the FDET calculations in this work, the energy and density of the core fragment A is evaluated with `PySCF`[52] using HF and the aug-cc-pVDZ basis set. The local-density approximation (LDA) was used for the non-additive functional and functional derivatives, $E_{xcT}^{\text{nad}}[\rho_A, \rho_B]$, $v_{xcT}^{\text{nad}}[\rho_A, \rho_B](\mathbf{r})$, meaning the Thomas-Fermi kinetic energy functional,[53,54] the Slater exchange functional[55] and the Vosko-Wilk-Nusair fit of Eq. 4.4 for the correlation energy of the uniform electron gas (VWN5).[56] The AO contraction of the embedding potential was done using functions from `PySCF` and our Python libraries `FDET-Average`[57] and `FDETaco`.[58] The ChELPG charges discussed in subsection 3.1.3 were calculated with the new single-fragment module in Q-Chem 5.4[59] from densities obtained from doing linearized-FDET[60] embedding with HF/aug-cc-pVDZ and adding a Møller-Plesset first-order correction to them. Finally, the excitation energies were obtained also with the linearized-FDET approximation, with the reference density, ρ_A^{ref}, being the ground-state HF/aug-cc-pVDZ density of the isolated acetone. For frame-wise calculations with FDET, the density of the environment was obtained from the superposition of molecular densities at the HF/aug-cc-pVDZ level of theory. All excitation energies were computed with the ADCC python module,[61] following the FDE-ADC(2) procedure as described in Prager *et al.*[60]

3 Results and Discussion

3.1 Electrostatic Properties

3.1.1 Solvent Systems and Environmental Averaging Methods

We take this opportunity to not only test Eq. 2, but to also evaluate different sets of approximations used for building the solvent environment from the MD simulation frames which we will call **Direct/Cube**, **Direct/Supercube**, **Grid/Cube**, and **Grid/Sphere**. In the **Direct/Cube** method, each MD frame is first centered around the ACE center-of-geometry[1] and the solute coordinates wrapped so that all atoms of the cubic system are within a box centered at ACE and maintaining the system's periodicity. Next, we directly calculate the ESP at the nucleus of each ACE atom using a simple pair-wise sum over all solvent atoms:

$$v_{elst}(\mathbf{r}_n) = \sum_{j=1}^{N_{\text{solv}}} \frac{q_j}{|\mathbf{r}_n - \mathbf{r}_{B,j}|}, \tag{16}$$

where $\mathbf{r}_n$ in this case is the position of a particular solute nucleus, N_{solv} the number of solvent atoms, and q_j the charge of solvent atom j at position $\mathbf{r}_{B,j}$. This process is repeated for every frame and the values of $v_{elst}(\mathbf{r}_n)$ from each frame are used to build probability distributions and ensemble averages. Using this method, the ensemble and environmental average ESPs are by definition exactly the same. Because there are no approximations made in calculating the ESP in this method, we consider **Direct/Cube** our reference method. In order to determine if the 1000 SPCE water system was sufficiently large for the ESP to converge, we created the **Direct/Supercube** system, in which 26 copies of the solvent box were tiled around a central box. The ESP experienced by the central ACE molecule was then calculated and compared to that of the single 1000 solvent molecule box.

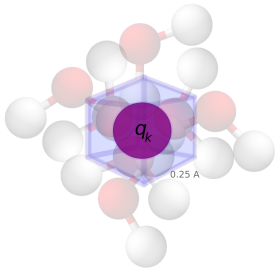
The pair-wise calculation in **Direct/Cube** can be extremely expensive due to poor scaling of Eq. 16 with N_{solv} and the number of grid-points $\{\mathbf{r}_n\}$, particularly when evaluating v_{elst} on a dense integration grid, as it is done in the next subsections to calculate derivative properties. To address this limitation, we add one approximation in what we call the **Grid/Cube** method. The MD frames are prepared in the same manner as in **Direct/Cube**, but are then rotated such that each frame is aligned to the ACE nuclei. Next, solvent atoms and their charges are compacted inside 0.025 nm cubic volume elements using a 3D histogram, illustrated in Scheme 2. The ESP at the solute nuclei is then calculated according

[1] Although our solute is rigid, it is permitted to freely translate and rotate during the simulation

to

$$v_{elst}(\mathbf{r}_n) = \sum_{l=1}^{N_{\text{grid}}} \frac{q_l}{|\mathbf{r}_n - \mathbf{r}_l|}, \tag{17}$$

where N_{grid} is the number of grid points, q_l the average of all solvent charges compacted into into volume element l, and $\mathbf{r}_l$ the center of the volume element. This cubic grid ESP is then extrapolation over the molecular Becke-Lebedev grid using the nearest-neighbor method. The cubic grid is dense enough such that each point typically contains only 1 atom per MD frame, therefore no efficiency is gained using the grid method for determining the ensemble average. On the other hand, the environmental average ESP calculation is much more efficient, as all solvent atoms from the 5000 frames are compacted into grid points which significantly reduces the total number of electrostatic interactions to calculate. There are two main differences in this approximation compared to the that in Ref. 10. First, we build an evenly distributed cubic grid instead of using the minimal number of charges placed on spherical shells around the solute with optimized radius. Secondly, we do not optimize the charges, they are only compacted according to their position.



Scheme 2: The **Grid** method for averaging. For each volume element, the compacted charge q_k used to evaluate the averaged environment density in Eq. 25, is the combination of all the charges that appear inside the volume element v_α throughout the trajectory. The size of the cube allows only one atom per volume element in each frame.

Unfortunately, the rotation required to align each frame in **Grid/Cube** results in a non-uniform solvent density for distances more than 1/2 a box length away from the ACE nuclei while also destroying the periodicity of the system. Therefore, physically meaningful environmental averages cannot be extracted from the **Grid/Cube** system, and consequently, they are not calculated. To alleviate this problem, we employ the **Grid/Sphere** method, were we only consider a sphere of water molecules whose oxygen atom is $\leq$ ½ a box length from the ACE center-of-geometry. Our hope is that these spherical clusters exhibit the same properties as the **Grid/Cube** reference system. Illustrations of **Direct/Cube**, **Di-**

rect/Supercube, and **Grid/Sphere** are shown in Scheme 3. Note that all *environmental averages* require alignment of the individual frames, and for the **Direct** approaches, the electrostatic potential is evaluated as in Ref. 12, using the charges of all water molecules over the 5000 frames, scaled by a factor of 1/5000.

3.1.2 Electrostatic Potentials

Distributions of the ESP at the ACE nuclei, $\mathbf{r}_n = \mathbf{R}_{A,n}$, and the corresponding environmental averages from the **Direct/Supercube**, **Direct/Cube**, **Grid/Cube**, and **Grid/Sphere** methods are shown in Figure 1. The ensemble averages, standard deviations, and environmental averages are plotted in Figure 2 and tabulated in Table **??**. These data demonstrate that the ESP distributions of all the ACE nuclei are roughly Gaussian and all chemically equivalent atoms have the same distributions and ensemble average. We also observe negligible differences between the **Direct/Supercube** and **Direct/Cube** methods, allowing us to conclude that the 1000 water molecule box was sufficiently large to converge the system. The standard deviations of the **Grid/Cube** distributions are roughly 30% larger than the **Direct/Cube** reference system, whereas the **Grid/Sphere** distributions are only 20% larger. We attribute this discrepancy to the zero-th order extrapolation scheme that is used to project the discrete solvent charges on the 3D-grid; higher order schemes like B-splines would be more appropriate. The environmentally averaged ESPs are in strong agreement with their ensemble counterparts and fall well within the standard deviation of the ensemble averages.

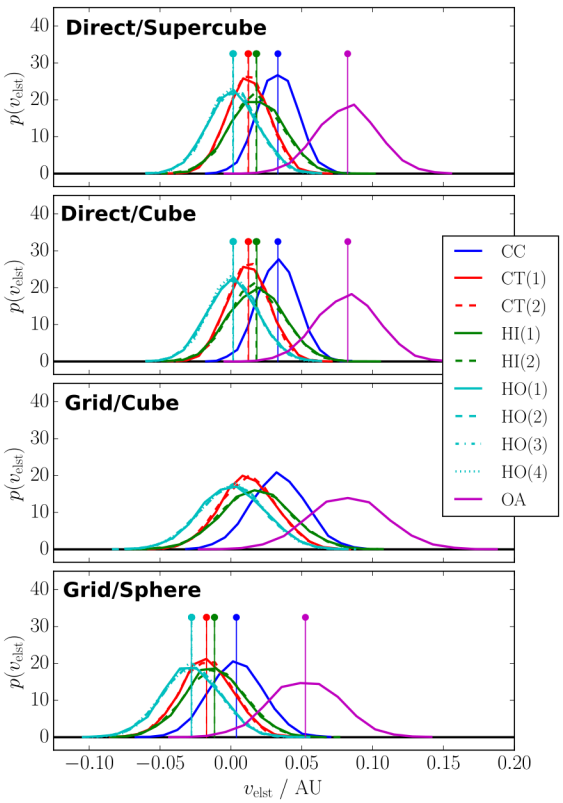


Figure 1: Probability distributions and and environmental averages (sticks) of the ESP experienced by the ACE nuclei for the **Direct/Supercube**, **Direct/Cube**, **Grid/Cube**, and **Grid/Sphere** systems.

We also found that the ESP distributions, ensemble averages, and environmental averages of the **Grid/Sphere** method were all shifted by about -0.29 AU from those of the **Direct/Cube** and **Grid/Cube** methods. This offset can be attributed to the differences in boundary conditions between the periodic cubic and non-periodic spherical systems, which results in a constant shift in the ESP. As we shall see, the ESP offset is constant, *i.e.* it is the same at every position, explaining why this difference does not affect any of the derivative properties. See the Appendix A for a complete explanation of the origin of the shift.

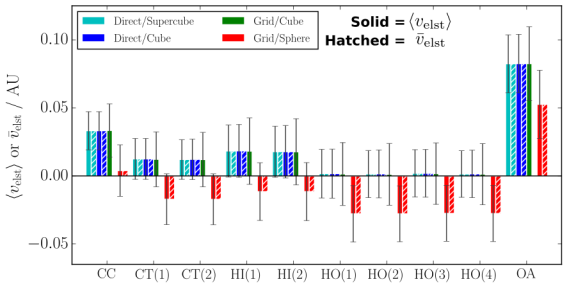
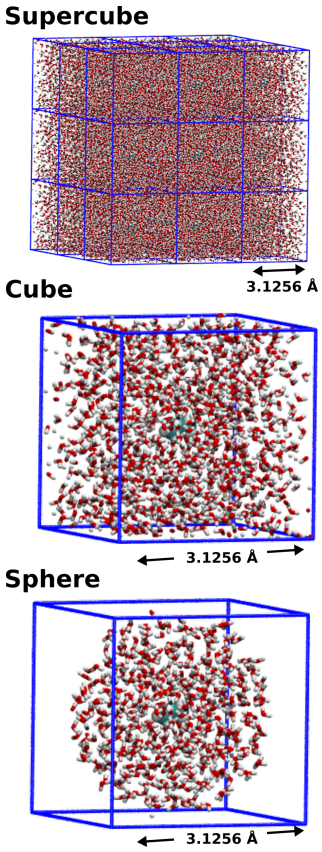


Figure 2: Ensemble and environmental averages of the ESP at all ACE nuclei. Error bars are the standard deviation.

These results demonstrate that there is no significant difference between the ensemble and environmentally averaged ESPs derived from **Direct/Cube** reference method and those of the **Grid/Sphere** method results, which allows us to conclude that both the grid approximation and the spherical cluster are acceptable approximations for determining the ESP experienced by a solute. These approximations effect only a minor increase in the width of the ESP probability distributions while giving the same ensemble averages. The environmental averaging techniques we employed resulted in average ESPs that are equivalent to the full ensemble calculations.



Scheme 3: Cubic (top) and spherical (bottom) systems.

3.1.3 ChELPG Charges and Molecular Dipole Moments

The first derivative properties we will consider are the ChELPG charges and dipole moments of the ACE solute determined in the presence of the solvent ESP. The solute electron densities used to evaluate the electrostatic potential in the ChELPG procedure are obtained from a Hartree-Fock optimization of ACE in the presence of the solvent ESP, $v_{elst}(\mathbf{r})$, as in Eq. 9. We then performed the ChELPG fitting procedure for each individual frame of the MD simulations in order to determine the probability distributions and ensemble averages, $\langle Q_A^{\mathrm{ChELPG}} \rangle$, of the point–charge of each ACE atom. This procedure was carried out for ESPs derived using the **Direct/Cube Grid/Cube** and **Grid/Sphere** methods described in Section 3.1.2. The environmentally averaged charges, $\bar{Q}_A^{\mathrm{ChELPG}}$, were determined from a single ChELPG calculation with ACE embedded in the environmentally averaged ESP determined using either the **Direct/Cube** or **Grid/Sphere** methods.

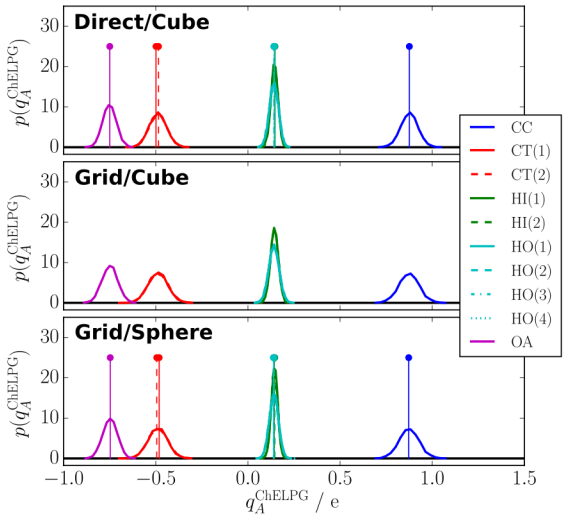


Figure 3: Probability distributions and and environmental averages (vertical lines) of the ChELPG charges of all ACE atoms for the **Direct/Cube**, **Grid/Cube**, and **Grid/Sphere** systems.

Probability distributions, ensemble averages, and environmental averages of the ACE ChELPG charges are shown in Figure 3. Their ensemble averages, standard deviations, and environmental averages are plotted in Figure 4 and tabulated in Table **??**. As with the ESPs, all ChELPG charge distributions are Gaussian and chemically equivalent atoms have the same distributions of point charges. The ensemble and environmental point charges from all systems are practically identical, just as was seen for the ESPs, but the standard deviation of the **Grid/Cube** distributions are, for most atoms, a factor of 2 larger than **Direct/Cube**. This is in contrast to the **Grid/Sphere** standard deviations which are nearly the same as **Direct/Cube**. The offset in the ESP observed in the **Grid/Sphere** method did not have an effect on the distribution of ChELPG charges. We now can be confident that ESPs derived from the grid and spherical solvent cluster approximations do not siginficantly change a solute's ensemble and environmental average point–charges compared to a direct ESP evaluation.

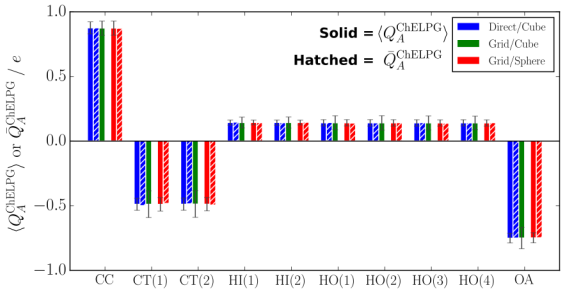


Figure 4: Ensemble and environmental averages of the ChELPG charges for all ACE nuclei. Error bars are the standard deviation.

As another test of the quality of the ChELPG method and our averaging procedures, we calculated the dipole moments generated from the quantum mechanical (QM) charge density resulting from the Hartree-Fock optimization of ACE and the classical (CL) dipole moments calculated from the fitted ChELPG point charges. The distribution of dipole moments from the ensemble of calculations and the corresponding environmental averages using the **Direct/Cube** **Grid/Cube** and **Grid/Sphere** methods are shown in Figure 5. The ensemble averages, standard deviations, and environmental averages can be found in Figure 6 and Table 1. The ensemble and environmental average dipole moments are nearly identical in all systems, whereas the standard deviation of the **Grid/Sphere** distributions are 15% narrower than their **Direct/Cube** and **Grid/Cube** counterparts. We can conclude from these data that the ChELPG method does an excellent job of reproducing the average quantum mechanical solute dipoles as well as their distributions. Additionally, the grid and sphere approximations do not significantly affect the dipole moment of the solute.

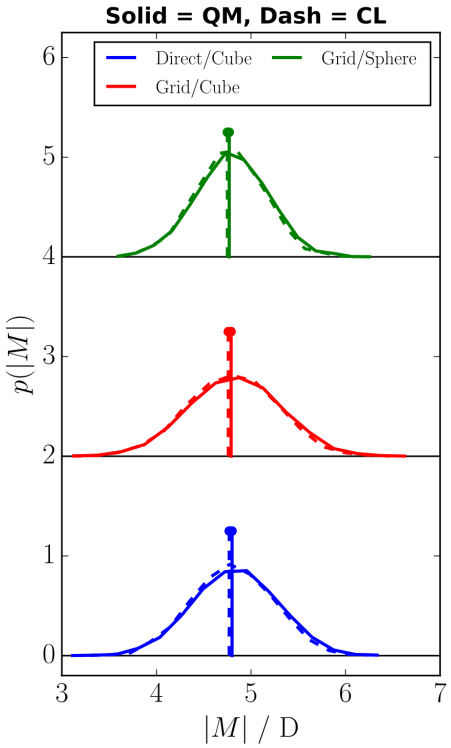


Figure 5: Distribution of the ACE quantum mechanical (QM, solid lines) dipole moments and the classical dipole moment derived from the ChELPG point-charges (CL, dashed lines) of the ACE nuclei from the **Direct/Cube**, **Grid/Cube**, and **Grid/Sphere** systems. The distributions have been offset vertically for clarity. Vertical lines are the environmental averages.

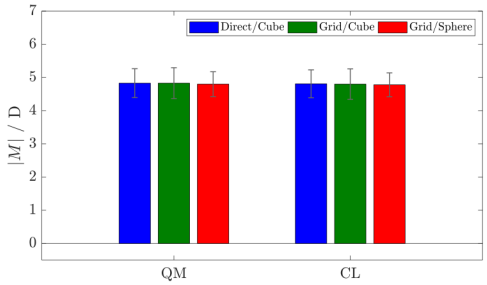


Figure 6: Ensemble and environment averages of the quantum mechanical (QM) and classical (CL) dipole moments. Error bars are the standard deviation.

Although the QM and CL dipole moment averages and distributions are the same for all three systems, that does not guarantee that the dipole moments of individual frames are also the same. Confirming this correlation is critical for evaluating differences between calculations performed with different approximations on the same MD frame. The top panel of

Figure 7 shows the correlation between the QM and CL dipole moments for **Direct/Cube**, **Grid/Cube**, and **Grid/Sphere**. These data confirm that the ChELPG procedure faithfully reproduces the dipole moments for every frame/ESP method combination considered here. On the other hand, correlations between frames treated with different ESP approximations are not as strong (bottom panel Figure 7). These data show that, for a single frame, the **Direct/Cube Grid/Cube** and **Grid/Sphere** methods result in significantly different dipole moments and, by inference, ChELPG charges and ESPs. We attribute this decorrelation to two factors: 1) the slight movement of solvent atoms from their 'true' positions when the gridding procedure is applied and 2) the removal of solvent molecules when constructing the spheres from the cubes. When either procedure 1) or 2) are applied to the original solvent cube they produce a 'new' set of solvent coordinates that are not exactly the same as the original cube of solvent. Due to the sensitive nature of the ESP, frame-to-frame correlation is lost, which manifests later in decorrelated ChELPG charges and dipole moments. Even though the within-frame correlation is lost, no effect is seen in the average properties or their distributions, suggesting that the ensemble properties of the system are conserved even though individual frame correlation is not. This effect must always be kept in mind when applying different approximations to individual frames. Although these approximations do not affect the ensemble and environmental averages, evaluating approximations by comparing results from a single MD frame can be misleading.

Table 1: Ensemble, $\langle|M|\rangle$, and environmental, $|\bar{M}|$, average ACE dipole moments (vertical lines in Figure 5 and bars in Figure 6). The columns labeled 'σ / D' are the standard deviations of the ensemble averages (error bars in Figure 6). Rows labeled 'QM' are the dipole moments calculated from the quantum mechanical charge density and 'CL' from the CHELPG charges.

	Direct/Cube			**Grid/Cube**			**Grid/Sphere**		
	$\langle\|M\|\rangle$ / D	σ / D	$\|\bar{M}\|$ / D	$\langle\|M\|\rangle$ / D	σ / D	$\|\bar{M}\|$ / D	$\langle\|M\|\rangle$ / D	σ / D	$\|\bar{M}\|$ / D
QM	4.83	0.44	4.80	4.83	0.47	–	4.80	0.38	4.77
CL	4.81	0.43	4.77	4.80	0.47	–	4.78	0.37	4.75

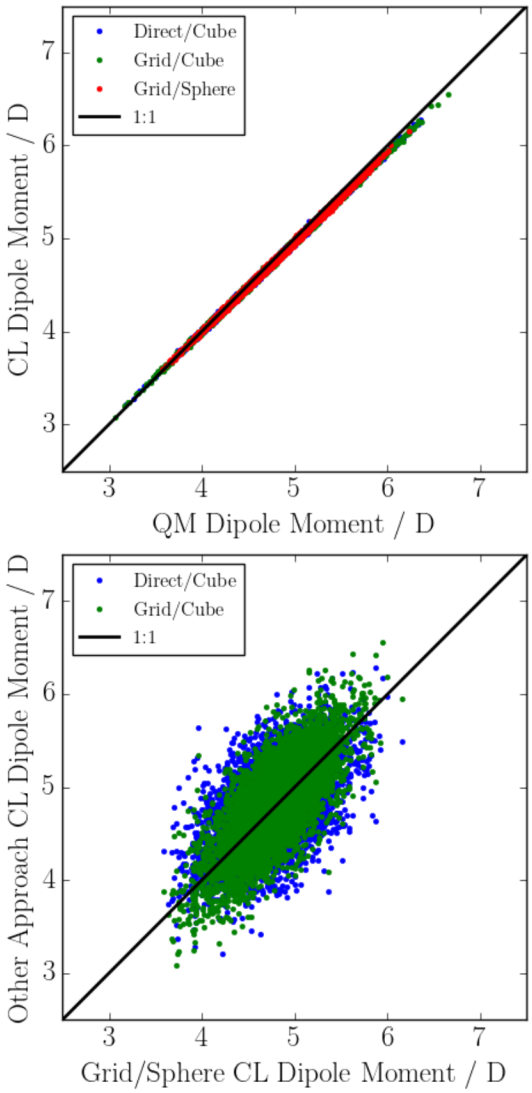


Figure 7: Top: Correlation between the QM and CL dipole moments in the **Direct/Cube**, **Grid/Cube**, and **Grid/Sphere** systems. Bottom: Correlation between the **Grid/Sphere** CL dipole moments and the **Direct/Cube** and **Grid/Cube** systems.

3.1.4 The Separated QM/MM Time-Averaging Scheme

The electrostatic interaction energy defined in Eq. 9 can be written in two equivalent ways, as a function of the embedding electrostatic potential created by B on A, $v_{els}(\mathbf{r})$ in our previous

notations, or reversely the external electrostatic potential created by A on B, $v_A(\mathbf{r};[\rho_A])$, *i.e.*,

$$\begin{aligned} V_{elst} &= \int d\mathbf{r}\, v_{elst}(\mathbf{r})\,(\hat{\rho}_{N_A}(\mathbf{r}) + \rho_A(\mathbf{r})) && (18)\\ &= \int d\mathbf{r}\, v_A(\mathbf{r};[\rho_A])\,\hat{\sigma}_B(\mathbf{r}) \\ &\simeq \int d\mathbf{r}\, v_A(\mathbf{r};Q_k^{ChELPG}[\rho_A])\,\hat{\sigma}_B(\mathbf{r}) && (19)\\ &\simeq \sum_{j=1}^{N} q_j \sum_{k=1}^{M} \frac{Q_k^{ChELPG}}{|\mathbf{R}_{B,j} - \mathbf{R}_{A,k}|}. && (20) \end{aligned}$$

In the near equality, the electrostatic potential due to the embedded system charge density has been approximated by the one created by a set of point charges located on the nuclei, that are best fitted according to an appropriate scheme. As mentioned above, we have chosen the electrostatic-potential based ChELPG scheme.

Note that if the QM method used to obtain Ψ_A is variational, it can be shown easily (Appendix B) that the appropriate molecular dynamics scheme to handle the Hamiltonian defined by Eq. 9, with energy conservation only limited by the near equality of Eq. 19, is typically that for a polarizable system, obtained with the following procedure: 1) for a given configuration $\mathbf{R}_B$, compute Ψ_A and ρ_A by minimization of Eq. 9 using Eq. 18, monitor the 'gas phase + polarization' energy $\left\langle \Psi_A | \hat{H}_A | \Psi_A \right\rangle$, and extract the set of charges $\{Q_k^{ChELPG}[\rho_A]\}$, 2) compute the forces on the environment atoms using those point charges and Eq. 20, and 3) update the environment positions $\mathbf{R}_B$ and iterate. Such algorithm is indeed standard in the literature, but its implementation in our analysis will be left for future investigations.

In the case of a completely separated MM and QM calculation scheme, as in this work, the MM simulations are performed with a reference non-polarizable system to generate afterwards the correct statistics for the polarizable system itself. To this end, one can define a reference, fixed set of charges $\{Q_k^0\}$, and a reference Hamiltonian H_0 where the electrostatic interaction seen by the environment in Eq. 20 is replaced by that with the fixed charges $\{Q_k^0\}$. The polarizable system Hamiltonian can be written as

$$\begin{aligned} H &= H_0 + \Delta V_{elst} && (21)\\ &= H_0 + \sum_{j=1}^{N} q_j \sum_{k=1}^{M} \frac{Q_k^{ChELPG} - Q_k^0}{|\mathbf{R}_{B,j} - \mathbf{R}_{A,k}|}. && (22) \end{aligned}$$

Then, in the canonical thermodynamic ensemble, the canonical average value of any observ-

able of the polarizable system can be computed as

$$\left\langle \hat{O}(\mathbf{R}_B) \right\rangle = \frac{\left\langle \hat{O}(\mathbf{R}_B) \exp(-\beta \Delta V_{elst}(\mathbf{R}_B)) \right\rangle_0}{\langle \exp(-\beta \Delta V_{elst}(\mathbf{R}_B)) \rangle_0}, \tag{23}$$

where the subscript 0 indicates a dynamics performed with the reference Hamiltonian H_0, *i.e.*, with the same self-consistent scheme as above but with the forces computed with the fixed charges $\{Q_k^0\}$ in step 2.

The biased ensemble averages (calculated while neglecting the exponentials in Eq. 23) and unbiased ensemble average (calculating using Eq. 23) for the ChELPG charges of all ACE atoms from the 5000 frame **Direct/Cube** and **Grid/Sphere** systems are provided in Table 2, in addition to the initial ChELPG charges used in the MD simulations. The variations we see here between the biased and unbiased ChELPG charges are smaller than the fluctuations due to the environment. Therefore, we are confident in only considering the biased averages in future calculations. This conclusion by be different, though, if the initial ChELPG charges were drastically different from the final charges.

Table 2: Initial, biased, and unbiased ensemble average ChELPG charges of ACE from the **Direct/Cube** and **Grid/Sphere** systems.

		Direct/Cube			**Grid/Sphere**		
atom	**Initial / e**	**Biased / e**	**Unbiased / e**	**% Diff.**	**Biased / e**	**Unbiased / e**	**% Diff.**
CC	0.7800	0.8759	0.8883	1.4%	0.8743	0.8823	0.9%
CT1	-0.4926	-0.4885	-0.4706	-3.7%	-0.4890	-0.4743	-3.0%
CT2	-0.4926	-0.4880	-0.5210	6.8%	-0.4878	-0.5148	5.5%
HI1	0.1338	0.1437	0.1398	-2.7%	0.1441	0.1385	-3.9%
HI2	0.1338	0.1440	0.1503	4.4%	0.1437	0.1506	4.8%
HO1	0.1423	0.1412	0.1344	-4.8%	0.1412	0.1350	-4.4%
HO2	0.1423	0.1410	0.1550	9.9%	0.1408	0.1442	2.5%
HO3	0.1423	0.1407	0.1281	-8.9%	0.1404	0.1404	0.0%
HO4	0.1423	0.1409	0.1439	2.1%	0.1404	0.1558	11.0%
OA	-0.6317	-0.7508	-0.7481	-0.4%	-0.7482	-0.7577	1.3%

3.1.5 Conclusions

We find that the grid method for determining the ESP and spherical solvent cluster approximation have no significant effect on the ensemble and environmental average ESP, ChELPG charges, or solute dipole moments. On the other hand, these approximations are less effective when applied within the same frame, which was attributed to differences in the solvent cluster size and shifting of solvent nuclei after applying the grid procedure. This means that the quality of the approximations can only be evaluated by looking at ensembles of calculations. Single-frame agreement should be improved by using a higher-order charge extrapolation

scheme and/or finer grids. This would be at the expense of an increased computation time for the 3D electrostatic embedding potential acting on the solute. Additionally, unbiasing the trajectories did not effect the average system properties, but this may not hold for systems in which the simulated solute charges are drastically different than those of the embedded solute. Therefore, we conclude that Eq. 2 holds for the ESP, ChELPG charges, and dipole moments and that **Grid/Sphere** set of approximations reasonably reproduces the electrostatic properties of the **Direct/Cube** reference system. We can now confidently use ESPs determined using the **Grid/Sphere** method for computing electronic structure properties.

3.2 Acetone Electronic Transitions

Next we will test the applicability of Eq. 2 to solute electronic transitions by calculating the first electronic transition of ACE embedded in different descriptions of the solvent environment. For all transition energy calculations, all obtained with the ADC(2) method,[62] the reference is the first excitation energy of an isolated ACE molecule in vacuum.

One set of calculations will be performed using point charge embedding (PCE), where the ACE solute is embedded in the ESP generated by the solvent point charges taken directly from the MD forcefield (that of SPCE water). For the ensemble averages, the embedding ESP is determined from a sphere of solvent molecules with a radius of 1/2 the box length. The ESP is then determined using the pair-wise method of Eq. 16. This method combines the direct pair-wise evaluation of the ESP in the **Direct/Cube** method of Section 3.1.1 with the spherical solvent cluster approximation. Excitation energies for each frame are then calculated. For the environmental average, the embedding ESP was calculated using the **Grid/Sphere** method and a single-point excitation energy calculation. These procedures for determining the ensemble and environmental averages were chosen in order to compare our computationally cheap **Grid/Sphere** method of ESP determination with a very expensive, yet very exact, implementation of the PCE procedure.

We also calculated second set of ACE excitation energies using FDET embedding. This method combines the electrostatic contribution from the solvent with the non-electrostatic (*i.e.* Pauli exclusion) portion of the energy. The ensemble averages are determined by first taking a sphere of solvent molecules with a radius of 1/2 a box length centered at the ACE solute. A quantum mechanical charge density is then determined at the HF/aug-cc-pVDZ level of theory from an inner-sphere of solvent molecules 1/4 of a box length from the ACE. This superposed density is then used in the evaluation of the FDET embedding potential. All solvent molecules farther than 1/4 a box length distance from the ACE only contribute to the electrostatic portion of the embedding energy. This procedure is performed for each

of the 5000 MD frames in order to construct the ensemble average FDET excitation energy.

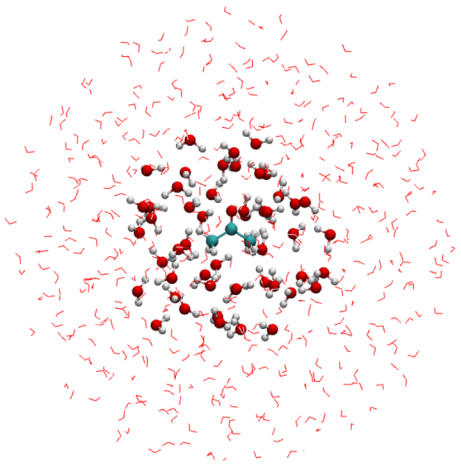


Figure 8: System used in the frame-wise FDET-embedding calculations. The internal sphere of water molecules (ball-and-stick models) are treated as densities and used to evaluate the FDET-embedding potential. Water molecules beyond this layer are treated as point-charges.

For the FDET environmental averages, the embedding charge density is evaluated using only the positions of the solvent nuclei and their forcefield charges; no quantum mechanical embedding density is used in order to maximize computational efficiency. We begin by gridding the solvent nuclei as in the **Grid/Sphere** method of ESP determination. The electronic density is obtained by dividing the effective electronic charge q^e_k, that is the effective number of electrons associated to the atom type k, by the volume element formed by the grid. As we need to account for all atoms over the ensemble of solvent geometries, instead of tracking individual atoms we group them into atom types. Here, it is important to stress that q_j from the definition of $\hat{\sigma}_B$ (Eq. 7), is the total effective charge of the j-th atom, taken from the force field. So, if the j-th atom is of the type k, then

$$q^e_{j(k)} = -q_j + Z_{j(k)} \tag{24}$$

where $Z_{j(k)}$ is the nuclear charge of the atom j, of type k. To simplify the notation, from now on we only use k.

The average solvent density at the position of the point with index α ($\mathbf{r}_\alpha = \{x_\alpha, y_\alpha, z_\alpha\}$), as in our previous work,[24] is evaluated as

$$\langle \rho_B \rangle(\alpha) = \sum_k q^e_k g_k(\alpha), \tag{25}$$

here, the division over the volume elements is contained inside the solvent distribution function $g_k(\alpha)$. For a given atom type k in the volume element corresponding to the point α:

$$g_k(\alpha) = \frac{n_{\alpha,k}}{N\nu_\alpha}; \quad \sum_{k,\alpha} g_k(\alpha)\nu_\alpha = N_{\text{nuc}}, \tag{26}$$

where N is the number of configurations (MD-frames), N_{nuc} being the total number of nuclei in the box, $n_{\alpha,k}$ the number of occurrences of an atom of type k in the volume element ν_α, and the total volume of the box $V = \sum_\alpha \nu_\alpha$. Therefore, the average electrostatic potential is computed as

$$v_{elst}[\langle\rho_B\rangle](\mathbf{r}) = \sum_{k,\alpha} \frac{(-q_k^e + Z_k)g_k(\alpha)\nu_\alpha}{|\mathbf{r} - \mathbf{r}_\alpha|} \tag{27}$$

We emphasize that this ESP is equivalent to the **Grid/Sphere** method described in Section 3.1.1. Whereas for the PCE the embedding potential is just $v_{elst}(\mathbf{r})$, for the FDET calculations the embedding potential is:

$$\begin{aligned} v_{emb}[\rho_A, \langle\rho_B\rangle; v_{elst}](\mathbf{r}) =& v_{elst}[\langle\rho_B\rangle](\mathbf{r}) \\ &+ v_{xcT}^{\text{nad(lin)}}[\rho_A, \langle\rho_B\rangle](\mathbf{r}). \end{aligned} \tag{28}$$

The environmental average excitation from FDET is then obtained from a single-point calculation in the presence of the potentials of Eq. 28.

For comparison with the single-point environmental averages, we calculate the ensemble average transition energies using two approaches. First, we take the weighted average of the transition energies:

$$\langle\nu_{\text{ex}}\rangle = \sum_{i=1}^{N} w_i \nu_{\text{ex}}^{(i)}, \tag{29}$$

where $\nu_{\text{ex}}^{(i)}$ is the excitation energy of frame i, and the weight, w_i, is determined from the oscillator strength, f_i, according to

$$w_i = \frac{f_i}{\sum_{i=1}^{N} f_i}. \tag{30}$$

The weighted standard deviation of $\langle\nu_{\text{ex}}\rangle$ is calculated as:

$$\sigma_{\text{wt}} = \sqrt{\sum_{i=1}^{N} w_i \big(\nu_i - \langle\nu\rangle\big)^2}. \tag{31}$$

Note that Eq. 3.2 is only valid when $\sum_i w_i = 1$. We also calculate the unweighted average of ν_{ex},

$$\langle \nu_{\text{ex}} \rangle_d = \frac{1}{N} \sum_{i=1}^{N} \nu_{\text{ex}}^{(i)}. \tag{32}$$

which is the version of an ensemble average closest to the environmental average, where a single calculations is performed. Note that this expression assumes that the excitation energies can be decoupled from their oscillator strength (hence the subscript d). If this assumption holds we expect to observe $\langle \nu_{\text{ex}} \rangle = \langle \nu_{\text{ex}} \rangle_d$.

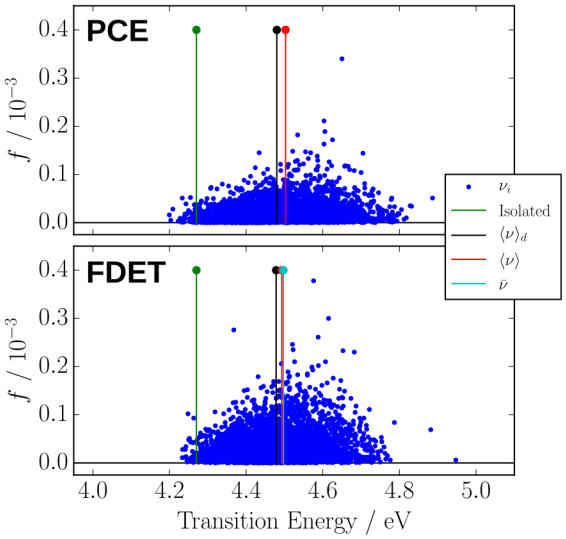


Figure 9: Electronic transitions of ACE in SPCE water evaluated using PCE and FDET embedding methods. Due to the symmetry of the transition, the oscillator strength of the isolated ACE and environmentally averaged calculations is effectively 0.

Table 3: Ensemble and environmental first excitation energies of ACE. The ensemble averages (Eq. 3) are weighted by the oscillator strength. Shifts are calculated with respect to the isolated ACE. All values are in eV.

	$\langle \nu_{\text{ex}} \rangle_d$	$\langle \delta\nu_{\text{ex}} \rangle_d$	$\langle \nu_{\text{ex}} \rangle$	$\langle \delta\nu_{\text{ex}} \rangle$	σ_{wt}	$\bar{\nu}_{\text{ex}}$	$\delta\bar{\nu}_{\text{ex}}$
Isolated	–	–	–	–	–	4.270	–
PCE	4.480	0.210	4.503	0.233	0.105	4.493	0.223
FDET	4.478	0.208	4.493	0.223	0.093	4.497	0.227

The ACE transitions from these three sets of calculations are plotted in Figure 9, and all transition energy averages and single-point calculation results are given in Table 3. The environmental average transition energies are in excellent agreement with the ensemble aver-

ages, to within 0.2%. One can also observe that the difference between the PCE and FDET methods is rather small. This is consistent with previous observations which found that the magnitude of the shift can be primarily attributed to the solute-solvent electrostatic interaction[24,39]. Additionally, the small difference between $\langle \nu_{\mathrm{ex}} \rangle_d$ and $\langle \nu_{\mathrm{ex}} \rangle$ shows that decoupling the oscillator strengths and excitation energies is a good approximation, applicable to model the average behavior of the transitions. However, in terms of frame-to-frame correlation, similar to the dipole moments, the single-frame oscillator strengths and transition energies are not well correlated with each other when different sets of approximations are applied (see Figure 10).

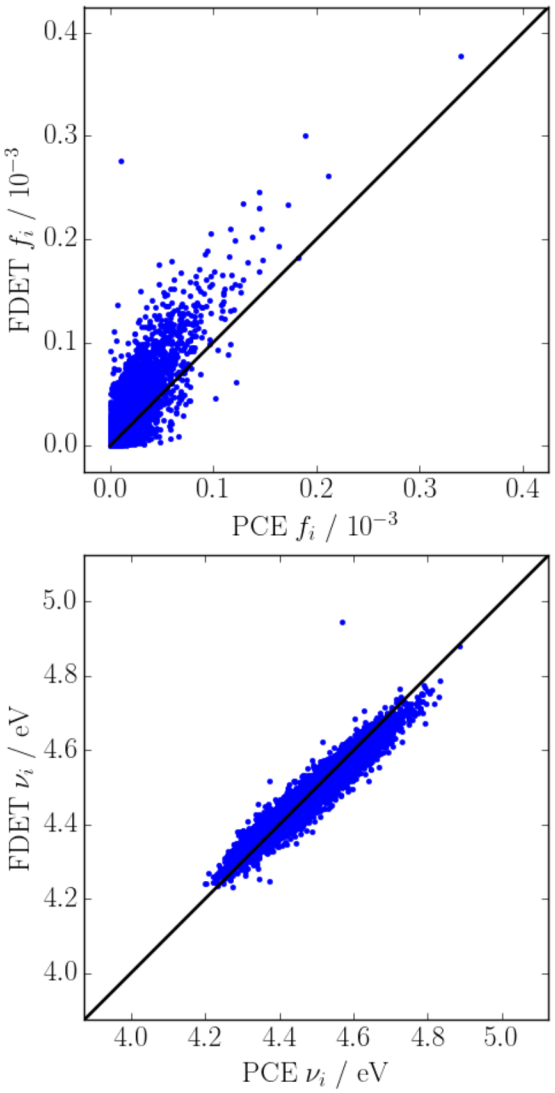


Figure 10: The correlation between the PCE and FDET oscillator strengths (top) and transition energies (bottom) for the first ACE electronic excitation. The black line represents a perfect 1:1 correlation.

3.3 Sample Size Dependence

At the outset of this study, we arbitrarily chose to use 5000 MD frames in the hope that this was a sufficiently large sample for properly converging the ensemble and environmental averages. In future applications involving high level theories, 5000 frames may be computationally intractable and a smaller number of calculations must be considered. We now look to determine the number of calculations needed in order to be 'reasonably' confident in the average excitation shift generated by an ensemble of calculations. For this section, we will consider only the solvatochromic shift, $\delta\nu_{\text{ex}} = \nu_{\text{emb}} - \nu_{\text{iso}}$, *i.e.* the shift between the isolated and embedded transition energies. We use this quantity, instead of the excitation energies, as we are interested in the accuracy of the description of the effect of the environment, and not the accuracy of the QM method itself.

First we will consider the number of frames required to converge an ensemble average shift, $\langle\delta\nu_{\text{ex}}\rangle$. We begin by splitting the 5000 frame FDET/Sphere trajectory into n_{samp} samples consisting of n_{frame} sequential frames, where $n_{\text{samp}} * n_{\text{frame}} = 5000$. The sample's mean excitation shifts were calculated as in Section 3.2, together with the weighted standard deviation. We will call these individual sample calculations the 'sample means' and 'sample standard deviations.' After processing each sample, the mean of the sample means and its 95% confidence interval were calculated, which we will refer to as the 'aggregate mean' and 'aggregate standard deviation.' The aggregate standard deviation gives insight into the uncertainty of $\langle\delta\nu_{\text{ex}}\rangle$ when it is calculated from n_{frame} individual calculations.

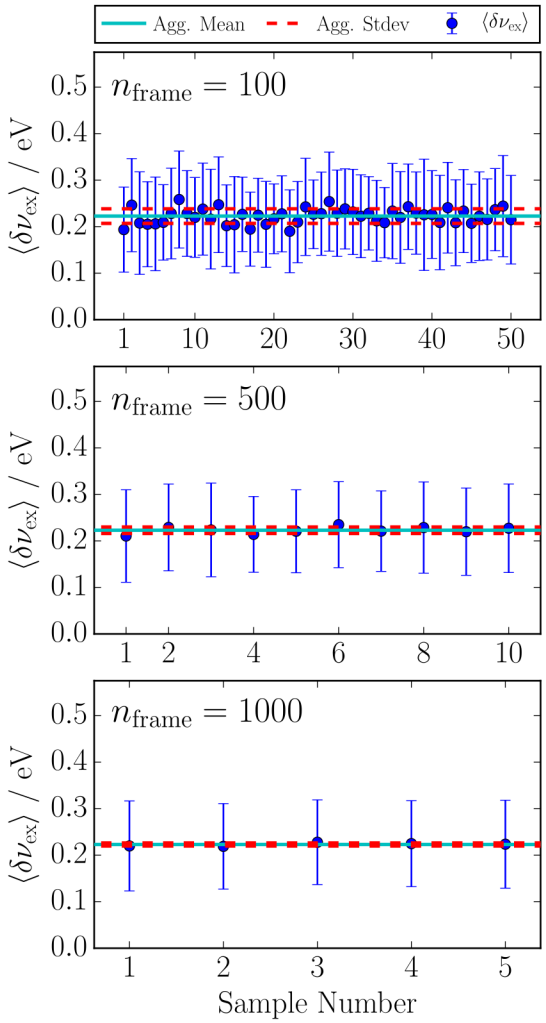


Figure 11: Break down of the full 5000 FDET/Sphere transition energy shifts, $\langle \nu \rangle = \sum_{i=1}^{N} w_i \nu_i$, into smaller sequential samples of varying size.

Such data for $n_{\text{frame}} = 100,\ 500,$ and 1000 are plotted in Figure 11, where the individual sample means are plotted as blue circles with blue error bars representing the sample standard deviation. The aggregate means and standard deviations are plotted as cyan and red lines, respectively. These demonstrate that the fluctuations within a sample (blue error bars) are reasonably well captured by as few as 100 MD frames. Therefore, if one wishes only to characterize the fluctuations in the shift itself, only 100 or so calculations are needed. The relationship between the uncertainty in the ensemble average shift and n_{frame} is shown in Figure 12. The aggregate standard deviations fit well to $a/\sqrt{n_{\text{samp}}}$ law, as expected for a Gaussian random variable. Based on this fit, at least 775 individual transitions must be calculated in order to achieve an aggregate standard deviation in $\langle \nu_{\text{ex}} \rangle$ of ≤ 5 meV.

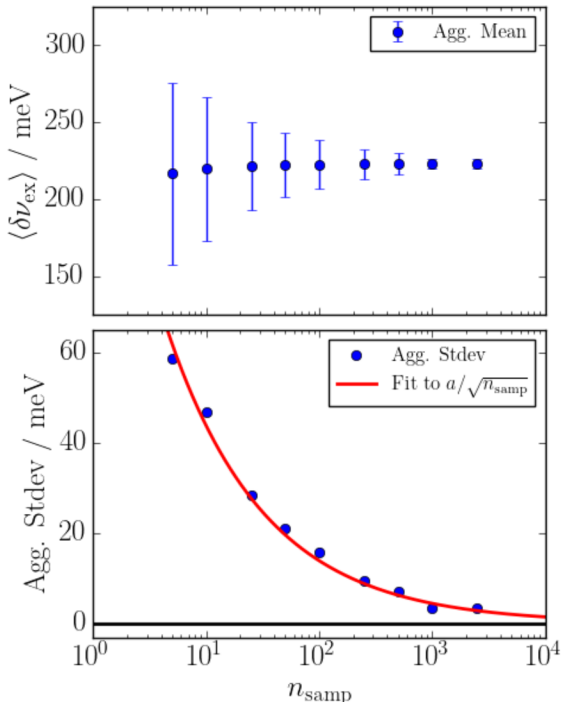


Figure 12: Top: Aggregate means of the ensemble average transition energy shift $\langle \delta\nu_{ex} \rangle$ vs. n_{samp}. Error bars are the aggregate standard deviation. Bottom: Aggregate standand deviations vs. n_{samp}. The red curve is a fit to $a/\sqrt{n_{samp}}$ where $a = 139$ meV.

Although we have confirmed that the ensemble calculations behave normally, this does not necessarily mean that the same conclusion can be made for the environmental averages.To characterize the uncertainty in $\bar{\nu}_{ex}$, we broke down the 5000 frame MD trajectory into samples of 100, 500, and 1000 sequential frames. An average environment using the **Grid/Sphere** method from each sample and an FDET calculation was performed in the same manner as in Section 3.2. Figure 13 shows the resulting environmentally averaged transition energy shift from each of these samples, and the aggregate means and standard deviations of $\delta\bar{\nu}_{ex}$ are given in Table 4. We find that in order to ensure an uncertainty in $\delta\bar{\nu}_{ex}$ of ≤ 5 meV, a minimum of 500 MD frames must be used to construct the average solvent environment. Note that this is a lower bound on the uncertainty, as we have not included any solute or solvent flexibility or polarizability in our simulations. Our result is roughly in agreement with the coupled-cluster with single- and double-excitations reported in Ref. 39, where it was shown that from 400 MD configurations, the standard deviation of ensemble averages of the first excitation energy of acetone was smaller then 100 cm^{-1} (approx. 12 meV).

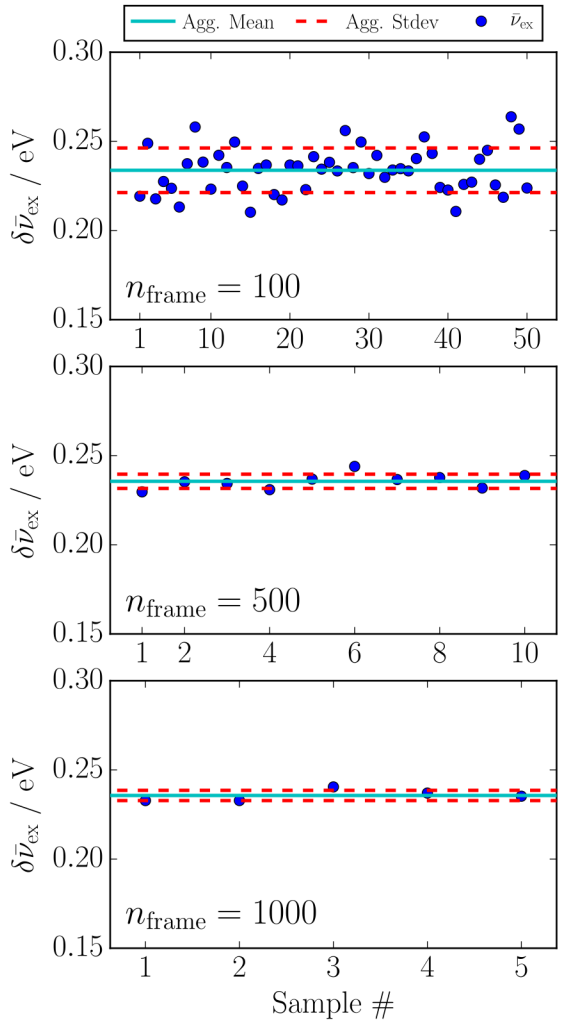


Figure 13: Environmentally averaged transition energy shifts, $\delta\bar{\nu}_{\mathrm{ex}} = \bar{\nu}_{\mathrm{emb}} - \nu_{\mathrm{iso}}$, constructed from a varying n_{frame} number of sequential simulation frames.

Table 4: Aggregate mean and standard deviation of the environmentally averaged transition energy shifts.

n_{samp}	n_{frame}	$\delta\bar{\nu}_{\mathrm{ex}}$ Agg. Mean / meV	$\delta\bar{\nu}_{\mathrm{ex}}$ Agg. Stdev / meV
50	100	233.8	12.5
10	500	235.6	4.0
5	1000	235.7	2.9

4 Conclusions

The results of our study demonstrate that Eq. 2 is a valid approximation for a rigid solute dissolved in a rigid solute. The the three different methods used for treating the solvent

environment, **Direct/Cube Grid/Cube** and **Grid/Sphere**, were found to give the same ensemble averages of the ESP, ChELPG charges, and solute dipole moments. Strong correspondence was found between these ensemble averages and their corresponding environmental averages. When reducing the system from a cube of solvent to a sphere, we observed a uniform shift of -0.29 eV in the ESP due to the change in the boundary conditions. Because this shift was constant for all points in space, it did not affect any properties derivative of the ESP. We found that the most efficient method to get environmental averages, **Grid/Sphere**, is valid for determining averages, but frame-to-frame correlation is lost due to the shifting of solvent atoms inherent in the grid approximation. This correspondence can only be recovered by using a very fine grid, so that the potential converges to the exact solution. Therefore, the **Grid/Sphere** method is not recommended when looking at properties that depend on the instantaneous fluctuations of the solvent or the solute. Because we chose to perform uncoupled QM/MM calculations, the effect of the choice in the charges used for the classical simulation was explored. We found that there was no significant difference in the ensemble averages before and after an unbiasing procedure was applied.

When studying the first vertical excitation of acetone, as in this particular case, the shift in the excitation energy due to the solvent is mainly the result of the electrostatic interactions, it was not a surprise finding that the results from PCE and FDET were very similar. The ensemble and environmental averages were found to be very close to each other in both cases: 0.233 eV and 0.223 eV with PCE, and 0.223 eV and 0.227 eV with FDET, proving once more the validity of Eq. 2. Finally, we looked at the dependence of the shift in the excitation energy on the number of independent frames used for constructing the ensemble and environmental averages. We found that at least 775 frames are needed for the ensemble averages to have an uncertainty of < 5 meV, whereas the environmental averages required at least 500 frames.

The immediate following steps in this project are to: 1) explore other approaches to construct environmental averages, for instance molecular density functional theory, and 2) investigate the applicability of environmental averages in cases where the solute is flexible and where the construction of an average geometry is required.[63]

Appendix A: Potential shift due to the change from P-type to M-type boundary conditions

Solvent molecules located at far distances from the solute molecule perspective, where the solute electric field is small, appear to be rotating freely. Therefore, on average, solvent molecules at the borders can be seen as averaged charges or charged spheres. For instance, water molecules are reduced to an isotropic quadrupole ($\gamma = 2q_H d^2$) and higher isotropic multipoles.

In the language of Kastenholz,[64] the periodic boundary conditions (PBC) used in MD and Molecular Density Functional Theory (MDFT) are labeled as P-boundary type. In this case, the average charge density at every point in the box is zero, and there are no discontinuities at the borders. On the contrary, M-type boundary conditions, where molecules are taken as a whole at the boundaries, give a different picture – due to the split of molecules there is a surface polarization P_M at the borders. According to Ref. 64, the polarization at the surface

$$P_M - \frac{1}{2}\rho\gamma \tag{33}$$

where ρ is the liquid density, creates inside the box a uniform shift $\Delta\phi$ of the potential,$\phi_M = \phi_{\text{out}} - \Delta\phi$. In the infinite fluid case,

$$\Delta\phi = \frac{2\pi}{3}\rho\gamma. \tag{34}$$

For SPCE water model, $d = 1$ Å, $\rho = 0.03328$ mol /Å^2, $q_H = 0.4238$ e, and $\gamma = 0.8476$, therefore

$$\Delta\phi = \frac{2\pi}{3}\rho\gamma \approx 0.059 \text{ e/Å} \approx 0.031 \text{ e/Bohr}. \tag{35}$$

Appendix B: The Separated QM/MM Time-Averaging Scheme

Supposing for simplicity that the solute A is kept immobile and the embedding potential reduced to its electrostatic component, and adopting here a DFT framework, the total energy of the $A + B$ system in Eq. 6 can be expressed as

$$\begin{aligned} H(\mathbf{R}_A, \mathbf{R}_B, \mathbf{p}_B) =& K(\mathbf{p}_B) + H_A[\rho_A(\mathbf{r})] \\ &+ V_{elst}[\rho_A(\mathbf{r}), \mathbf{R}_A, \mathbf{R}_B] \\ &+ V_{MM}(\mathbf{R}_A, \mathbf{R}_B), \end{aligned} \tag{36}$$

changing here E in H, and knowing indeed that there is a parametric dependence of $\rho_A(\mathbf{r})$ on both $\mathbf{R}_A$ and $\mathbf{R}_B$, *i.e.* $\rho_A(\mathbf{r}) \equiv \rho_A(\mathbf{r}; \mathbf{R}_A, \mathbf{R}_B)$. $V_{MM}(\mathbf{R}_A, \mathbf{R}_B)$ is the MM force field that describes the mutual interaction of the solvent molecules and their non-electrostatic interaction with the solute, modelled by site-site Lennard-Jones contributions. The force on any atom j of the solvent can be written as

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{R}_{B,j}} &= -\int d\mathbf{r} \left[\frac{\delta H_A}{\delta \rho_A(\mathbf{r})} + v_{elst}(\mathbf{r}) \right] \frac{\partial \rho_A(\mathbf{r})}{\partial \mathbf{R}_{B,j}} \\ &\quad - \int d\mathbf{r} \left(\rho_A(\mathbf{r}) + \hat{\rho}_{N_A}(\mathbf{r}) \right) \frac{\partial v_{elst}(\mathbf{r})}{\partial \mathbf{R}_{B,j}} - \frac{\partial V_{MM}}{\partial \mathbf{R}_{B,j}} \end{aligned} \tag{37}$$

The bracket in the first term is just the Euler-Lagrange equation for the quantum system and it vanishes for all $\mathbf{r}$'s. So does the whole integral. The second term is the electrostatic contribution to the force, $\mathbf{F}^{elst}_{B,j}$. Accounting from the equalities 17-19 in the main text it can be rewritten as

$$\begin{aligned} \mathbf{F}^{elest}_{B,j} &\simeq \int d\mathbf{r} \sum_{k=1}^{M} Q_k^{ChELPG} \delta(\mathbf{r} - \mathbf{R}_{A,k}) \sum_{j=1}^{N} q_j \frac{(\mathbf{R}_{B,j} - \mathbf{r})}{|\mathbf{r} - \mathbf{R}_{B,j}|^3} \\ &= \sum_{j=1}^{N} q_j \sum_{k=1}^{M} Q_k^{ChELPG} \frac{(\mathbf{R}_{B,j} - \mathbf{R}_{A,k})}{|\mathbf{R}_{A,k} - \mathbf{R}_{B,j}|^3} \\ &= -\frac{\partial}{\partial \mathbf{R}_{B,j}} \sum_{j=1}^{N} q_j \sum_{k=1}^{M} \frac{Q_k^{ChELPG}}{|\mathbf{R}_{A,k} - \mathbf{R}_{B,j}|} \\ &= -\frac{\partial}{\partial \mathbf{R}_{B,j}} \sum_{j=1}^{N} q_j v_A(\mathbf{R}_{B,j}; Q_k^{ChELPG}[\rho_A]) \end{aligned} \tag{38}$$

This implies that once the quantum Hamiltonian has been minimized, yielding the ground state electronic density ρ_A^{min}, the total energy of the QM/MM system is given by

$$\begin{aligned} H(\mathbf{R}_A, \mathbf{R}_B, \mathbf{p}_B) =& K(\mathbf{p}_B) + H_A[\rho_A^{min}(\mathbf{r})] \\ &+ \sum_{j=1}^{N} q_j v_A(\mathbf{R}_{B,j}; Q_k^{ChELPG}[\rho_A^{min}]) \\ &+ V_{MM}(\mathbf{R}_A, \mathbf{R}_B), \end{aligned} \tag{39}$$

and the force are given consistently by Eq. 38 above. This leads to the self-consistent molecular dynamics algorithm described in the text. As mentioned there too, one can define at this point a reference classical Hamiltonian H_0 for which the energy is defined exactly as above but with a fixed set of solute charges $\{Q_k^0\}$. One can then decompose the total energy into $H = H_0 + \Delta V_{elst}$ as in Eqs. 20-21. The following Eq. 22 is a standard relationship, exact in the canonical thermodynamic ensemble, which enables to compute the property of

the system under study from a dynamics performed with the reference Hamiltonian. Even if exact in theory, the formula is only applicable in practice if the reference system remains close to the exact one (here if the reference charges are a good guess with respect to the ChELPG charges measured on average), or if the statistics in ΔV_{elst} is Gaussian (a seemingly good approximation in our study). In that case, the following formula can be used

$$\left\langle \hat{O}(\mathbf{R}_B) \right\rangle = \left\langle \hat{O}(\mathbf{R}_B) \right\rangle_0 - 2\beta \left\langle \delta\Delta V_{elst}(\mathbf{R}_B)\delta\hat{O}(\mathbf{R}_B) \right\rangle_0 , \tag{40}$$

where $\delta X = X = \langle X \rangle$ indicate the fluctuation.

Acknowledgement

D.G. thanks the Department of Physical Chemistry at University of Geneva for invited Professor position. All calculations were performed on the Baobab and Yggdrasil clusters at the University of Geneva.

References

(1) Hirata, S.; Head-Gordon, M. Time-dependent density functional theory within the Tamm-Dancoff approximation. *Chem. Phys. Lett.* **1999**, *314*, 291–299.

(2) Dreuw, A.; Head-Gordon, M. Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chem. Rev.* **2005**, *105*, 4009–4027.

(3) Del Bene, J. E.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. X. Molecular Orbital Studies of Excited States with Minimal and Extended Basis Sets. *J. Chem. Phys.* **1971**, *55*, 2236–2241.

(4) Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. Toward a systematic molecular orbital theory for excited states. *J. Phys. Chem.* **1992**, *96*, 135–149.

(5) Stanton, J. F.; Barlett, R. J. The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *J. Chem. Phys.* **1993**, *98*, 7029–7039.

(6) Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94*, 2027–2094.

(7) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999–3094.

(8) Barone, V.; Maurizio, C. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

(9) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.

(10) Sanchez, M. L.; A., A. M.; Olivares del Valle, F. J. Study of solvent effects by means of averaged solvent electrostatic potentials obtained from molecular dynamics data. *J. Comput. Chem.* **1997**, *18*, 313–322.

(11) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361–373.

(12) Coutinho, K.; Georg, H.; Fonseca, T.; Ludwig, V.; Canuto, S. An efficient statistically converged average configuration for solvent effects. *Chem. Phys. Lett.* **2007**, *437*, 148–152.

(13) de Almeida, A. R.; Oliveira, L. B. A.; Colherinhas, G. Solvent effects on the spectroscopic properties of Damascone derivatives: A sequential Monte Carlo/Quantum Mechanics study. *Chemical Physics Letters* **2019**, *730*, 531–537.

(14) Bistafa, C.; Georg, H. C.; Canuto, S. Combining ab initio multiconfigurational and Free Energy Gradient methods to study the $\pi \rightarrow \pi^*$ excited state structure and properties of uracil in water. *Computational and Theoretical Chemistry* **2014**, *1040-1041*, 312–320, Excited states: From isolated molecules to complex environments.

(15) Jaramillo, P.; Coutinho, K.; Cabral, B. J.; Canuto, S. Explicit solvent effects on the visible absorption spectrum of a photosynthetic pigment: Chlorophyll-c2 in methanol. *Chemical Physics Letters* **2011**, *516*, 250–253.

(16) Torres, E.; Georg, H.; Fonseca, T.; Castro, M. First hyperpolarizability of isomers of pyridinium N-phenoxide betaine dye in solution using the ASEC-FEG method. *Chemical Physics Letters* **2018**, *699*, 261–266.

(17) Brandão, I.; Fonseca, T. L.; Georg, H. C.; Castro, M. A.; Pontes, R. B. Assessing the structure and first hyperpolarizability of Li@B10H14 in solution: a sequential QM/MM study using the ASEC–FEG method. *Phys. Chem. Chem. Phys.* **2020**, *22*, 17314–17324.

(18) Orozco-Gonzalez, Y.; Manathunga, M.; Marín, M. d. C.; Agathangelou, D.; Jung, K.-H.; Melaccio, F.; Ferré, N.; Haacke, S.; Coutinho, K.; Canuto, S.; Olivucci, M. An Average Solvent Electrostatic Configuration Protocol for QM/MM Free Energy Optimization: Implementation and Application to Rhodopsin Systems. *J. Chem. Theory Comput.* **2017**, *13*, 6391–6404.

(19) Wesołowski, T. A. Embedding a multideterminantal wave function in an orbital-free environment. *Phys. Rev. A* **2008**, *77*, 012504.

(20) Kaminski, J. W.; Gusarov, S.; Wesolowski, T. A.; Kovalenko, A. Modeling Solvatochromic Shifts Using the Orbital-Free Embedding Potential at Statistically Mechanically Averaged Solvent Density. *J. Phys. Chem. A* **2010**, *114*, 6082–6096.

(21) Zhou, X.; Kaminski, J. W.; Wesolowski, T. A. Multi-scale modelling of solvatochromic shifts from frozen-density embedding theory with non-uniform continuum model of the solvent: the coumarin 153 case. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10565.

(22) Zhou, X.; Wesolowski, T. A.; Tabacchi, G.; Fois, E.; Calzaferri, G.; Devaux, A. First-principles simulation of the absorption bands of fluorenone in zeolite L. *Phys. Chem. Chem. Phys.* **2013**, *15*, 159–167.

(23) Shedge, S. V.; Wesolowski, T. A. Nonuniform Continuum Model for Solvatochromism Based on Frozen-Density Embedding Theory. *ChemPhysChem* **2014**, *15*, 3291–3300.

(24) Laktionov, A. A.; Chemineau-Chalaye, E.; Wesolowski, T. A. Frozen-density embedding theory with average solvent charge densities from explicit atomistic simulations. *Journal of Chemical Theory and Computation* **2016**, *18*, 2829–2844.

(25) Horng, M. L.; Gardecki, J. A.; Papazyan, A.; Maroncelli, M. Subpicosecond Measurements of Polar Solvation Dynamics: Coumarin 153 Revisited. *The Journal of Physical Chemistry* **1995**, *99*, 17311–17337.

(26) Acevedo, O.; Jorgensen, W. L. Quantum and Molecular Mechanical (QM/MM) Monte Carlo Techniques for Modeling Condensed-Phase Reactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 422–435.

(27) Kawashima, Y.; Dupuis, M.; Hirao, K. Monte Carlo microsolvation simulations for excited states using a mixed-Hamiltonian model with polarizable and vibrating waters: Applications to the blueshift of the H2CO excitation. *The Journal of Chemical Physics* **2002**, *117*, 248–257.

(28) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–1229.

(29) Nochebuena, J.; Naseem-Khan, S.; Cisneros, G. A. Development and application of quantum mechanics/molecular mechanics methods with advanced polarizable potentials. *WIREs Computational Molecular Science* **2021**, *11*, e1515.

(30) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *Journal of Chemical Theory and Computation* **2010**, *6*, 3721–3734.

(31) Sanchez Mendoza, M. L.; Aguilar, M. A.; Olivares Del Valle, F. J. A mean field approach that combines quantum mechanics and molecular dynamics simulation: The water molecule in liquid water. *Journal of Molecular Structure: THEOCHEM* **1998**, *426*, 181–190.

(32) Martín, M. E.; Sánchez, M. L.; Olivares del Valle, F. J.; Aguilar, M. A. A multi-configuration self-consistent field/molecular dynamics study of the (n→π*)1 transition of carbonyl compounds in liquid water. *The Journal of Chemical Physics* **2000**, *113*, 6308–6315.

(33) Sánchez, M. L.; Martín, M. E.; Aguilar, M. A.; Olivares Del Valle, F. J. Solvent Effects by Means of Averaged Solvent Electrostatic Potentials: Coupled Method. *Journal of Computational Chemistry* **2000**, *21*, 705–715.

(34) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632–672.

(35) Gao, J.; Truhlar, D. G.; Wang, Y.; Mazack, M. J. M.; Löffler, P.; Provorse, M. R.; Rehak, P. Explicit Polarization: A Quantum Mechanical Framework for Developing Next Generation Force Fields. *Accounts of Chemical Research* **2014**, *47*, 2837–2845, PMID: 25098651.

(36) Neugebauer, J.; Jacob, C. R.; Wesolowski, T. A.; Baerends, E. J. An Explicit Quantum Chemical Method for Modeling Large Solvation Shells Applied to Aminocoumarin C151. *J. Phys. Chem. A* **2005**, *109*, 7805–7814.

(37) Sapana, S.; Wesolowski, T. A. Nonuniform Continuum Model for Solvatochromism Based on Frozen-Density Embedding Theory. **2014**, *15*, 3291–3300.

(38) Bernasconi, L.; Sprik, M.; Hutter, J. Time dependent density functional theory study of charge-transfer and intramolecular electronic excitations in acetone–water systems. *The Journal of Chemical Physics* **2003**, *119*, 12417–12431.

(39) Aidas, K.; Kongsted, J.; Osted, A.; Mikkelsen, K. V. Coupled Cluster Calculation of the $n \rightarrow \pi^*$ Electronic Transition of Acetone in Aqueous Solution. *J. Phys. Chem.* **2005**, *109*, 8001–8010.

(40) Neugebauer, J.; Louwerse, M. J.; Baerends, E. J.; Wesolowski, T. A. The merits of the frozen-density embedding scheme to model solvatochromic shifts. *J. Chem. Phys.* **2005**, *122*, 094115.

(41) Cao, L.; Ryde, U. On the Difference Between Additive and Subtractive QM/MM Caltulations. *Front. Chem.* **2018**, *6*, 89.

(42) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.

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

(44) Todorov, I. T.; Smith, W.; Trachenko, K.; Dove, M. T. DL_POLY_3: New Dimensions in Molecular Dynamics Simulations via Massive Parallelism. *J. Mater. Chem.* **2006**, *16*, 1911–1918.

(45) Bush, I. J.; Todorov, I. T.; Smith, W. A DAFT DL_POLY Distributed Memory Adaptation of the Smoothed Particle Mesh Ewald Method. *Comput. Phys. Commun.* **2006**, *175*, 323–329.

(46) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(47) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(48) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.

(49) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.

(50) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(51) Stone, J. *An Efficient Library for Parallel Ray Tracing and Animation*. M.Sc. thesis, Computer Science Department, University of Missouri-Rolla, 1998.

(52) Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K. PySCF: the Python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2017**, *8*, e1340.

(53) Thomas, L. H. The calculation of atomic fields. *Math. Proc. Cambridge Philos. Soc.* **1927**, *23*, 542.

(54) Fermi, E. Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente. *Z. Phys.* **1928**, *48*, 73–79.

(55) Slater, J. C. The Theory of Complex Spectra. *Phys. Rev.* **1929**, *34*, 1293–1322.

(56) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(57) González-Espinoza, C. E.; Ricardi, N. FDET-Averaged. `https://github.com/crisely09/fdeta`, 2020.

(58) González-Espinoza, C. E.; Scheurer, M. FDETaco. `https://github.com/crisely09/taco-1`, 2021.

(59) Epifanovsky, E. et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *The Journal of Chemical Physics* **2021**, *155*, 084801.

(60) Prager, S.; Zech, A.; Aquilante, F.; Dreuw, A.; Wesolowski, T. A. First time combination of frozen density embedding theory with the algebraic diagrammatic construction scheme for the polarization propagator of second order. *J. Chem. Phys.* **2016**, *144*, 204103.

(61) Herbst, M. F.; Scheurer, M.; Fransson, T.; Rehn, D. R.; Dreuw, A. adcc: A versatile toolkit for rapid development of algebraic-diagrammatic construction methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2020**, e1462.

(62) Dreuw, A.; Wormit, M. The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2015**, *5*, 82–95.

(63) Ricardi, N.; González-Espinoza, C. E.; Adam, S.; Church, J. C.; Schapiro, I.; Wesolowski, T. A. Embedding non-rigid solutes in an averaged environment: a case study on rhodopsins. *Submitted* **2021**,

(64) Kastenholz, M. A.; Hünenberger, P. H. Development of a lattice-sum method emulating nonperiodic boundary conditions for the treatment of electrostatic interactions in molecular simulations: a continuum-electrostatics study. *J. Chem. Phys.* **2006**, *124*, 124108.

Graphical TOC Entry