

PCCP

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Frequency-dependent force fields for QMMM calculations

Ignath Harczuk,* Olav Vahtras, and Hans Ågren

KTH Royal Institute of Technology, School of Biotechnology, Division of Theoretical Chemistry and Biology, SE-106 91 Stockholm, Sweden

E-mail: harczuk@kth.se

Abstract

We outline the construction of frequency-dependent polarizable force fields. The force fields are derived from analytic response theory for different frequencies using a generalization of the LoProp algorithm giving a decomposition of a molecular dynamical polarizability to localized atomic dynamical polarizabilities. These force fields can enter in a variety of applications - we focus on two such applications in this work: Firstly, they can be incorporated in a physical, straightforward, way for current existing methods that use polarizable embeddings, and we can show, for the first time, the effect of the frequency dispersion within the classical environment of a quantum mechanics - molecular mechanics (QMMM) method. Our methodology is here evaluated for some test cases comprising water clusters and organic residues. Secondly, together with a modified Silberstein-Applequist procedure for interacting inducible point-dipoles, these frequency-dependent polarizable force fields can be used for a classical determination of frequency-dependent cluster polarizabilities. We evaluate this methodology by comparing with corresponding results obtained from quantum mechanics or QMMM where the absolute mean $\bar{\alpha}$ is determined with respect to the size of the QM and MM parts of the total system.

*To whom correspondence should be addressed

Keywords: LoProp, Frequency Dependent Polarizability, Molecular Dynamics, Polarizable Force Field

Introduction

Force fields constitute crucial quantities in classical molecular dynamics simulations as they dictate the precision and the cost, and therefore the limitations and applicability of such calculations. It follows that force fields have been the subject of ever ongoing efforts of refinement and analysis where both the results of more accurate models. i.e. quantum mechanics based¹, and of fitting to experimental data are employed². Largely, such force fields divide into electrostatic³, polarizable⁴ and reactive force fields⁵, lately also so-called capacitance force fields⁶, regulating charge transfer in the molecular mechanics medium. With the implementation of quantum-classical hybrid models, like quantum mechanics - molecular mechanics, QMMM⁷, the quality of force fields are also scrutinized in terms of the interaction between the quantum and classical parts⁸. This puts additional demands in that the quality of the force fields should balance the precision of the quantum part for the evaluation of the structural dynamics or properties. More recent implementations^{9,10} of properties and spectroscopy in the QMMM framework often demand full electronic as well as polarizable electronic embeddings and furthermore complete granulation of such embeddings in terms of separate atomic contributions, sometimes even covering atom - bond partitioning and multipole expansions of the charges¹¹⁻¹³. An often attended vehicle to fulfill this demand is the so-called LoProp method¹⁴ for extracting atomic and interatomic contributions to molecular properties. It was originally formulated as a sequence of transformations of the atomic overlap matrix, giving a localized orthonormal basis that depends only on molecular structure and the initial atomic orbital basis set. The localized properties are then obtained by forming partial traces over basis functions associated with one chosen center or a pair of centers. These localized properties have indeed been of much use as granulated force-fields in QMMM calculations of properties, ranging from NMR, EPR over to optical properties and X-ray spectroscopy¹⁵⁻¹⁹.

In a previous work²⁰ we showed how a LoProp polarizable force field can be derived quantum mechanically for extended environments by using analytical response theory. The theory, where the change of a wave function due to an external field is represented by a unitary transformation, derives from the classical work by Olsen and Jørgensen²¹ for MCSCF theory, and later generalized for DFT by Salek et al.²² In the present work we extend the quantum mechanical response theory approach to construct the localized polarizable force fields to be frequency dependent. The possibility to do so follows from the fact that analytic response theory, in contrast to the finite field algorithm used in the original LoProp approach, can determine the first order response for frequencies which directly match those of the externally perturbing field. This effectively produces polarizable force fields which include the quantum mechanical density perturbation caused by an external field. This extension of the MM region in typical QMMM calculations allows for a more physical description of properties in the QM region, since the interaction between the MM and the QM regions now includes the frequency dependence of the classical region.

Our work is motivated by a few facts - frequency-dependent molecular-mechanics force fields would allow for an instantaneous, or optical, MM polarization by the electronic degrees of freedom in the quantum part of a QMMM model. Such a model has to our knowledge not yet been implemented: Secondly, it makes it possible to extend the QMMM properties to cover the full QMMM simulation volume. This is thus an extension of current QMMM property calculations where the embedding MM part acts as a perturber of the property determined by the QM: Thirdly, it allows for a calculation of the property for very large clusters that are evaluated classically. We address the latter aspect by implementing the Silberstein-Applequist model²³ for many interacting induced dipoles for the linear frequency dependent polarizability. As MM clusters are sampled by molecular dynamics at a given temperature and pressure, the possibility to follow the evolution of the cluster property in time and size is presented. To demonstrate the performance of our model we compute the static and dynamic polarizability of the TIP3P²⁴ water model and study the size-dependent behavior and convergence of the mean polarizability for different ensembles of water clusters. We present two different cases for these clusters, where we obtain configurations

of varying size by the means of MD simulations. We first calculate the error in the mean polarizability using a modified Silberstein-Applequist model for water in gas phase with respect to QM response theory, and then utilize the QMMM response method to obtain the absolute mean polarizability. For the QMMM part, we directly incorporate the LoProp frequency dependent properties in the MM region, and study their effect. We further apply this method to the calculation of the frequency dependence in the amino acid acid tryptophan situated in a protein. We choose the tryptophan model because of its size and conjugated aromatic system in the side-chain, to show that the localized frequency-dependent properties can be modeled in more complex molecules, and also be derived as force fields for biological systems. The tryptophan configurations are taken directly from MD trajectories, following a comparison of the convergence of the properties with respect to the simulation time.

A frequency-dependent LoProp approach

In a previous paper²⁰ we demonstrated an alternative derivation of the LoProp approach which was based on analytical response theory. As response theory provides a formulation of perturbation theory which handles time-independent and time-dependent cases within the same formalism, we now investigate the effects of projecting the dynamical polarizability onto atomic contributions. This can be expected to be meaningful provided that the frequencies involved are well below the first resonance. To outline the steps in single-determinant time-dependent response theory (see e.g. Salek et al²²) we have a state $|\tilde{0}\rangle$ that evolves in time

$$|\tilde{0}\rangle \equiv e^{-\hat{\kappa}(t)}|0\rangle \quad (1)$$

where the exponential operator is a parameterized time-evolution operator, $\hat{\kappa}$ a real anti-hermitean operator

$$\hat{\kappa} = \kappa_{pq}(t)a_p^\dagger a_q \quad (2)$$

of which the matrix elements $\{\kappa_{pq}\}$ form the parameters of the theory. A time-dependent variational principle based on the Ehrenfest theorem provides a solution to the time-independent Schrödinger equation in this space of parameters $\{\kappa_{pq}\}$, i.e. arbitrary static operators \hat{O} satisfy

$$\delta\langle 0 | [\hat{O}, e^{\hat{\kappa}(t)} (\hat{H} - i \frac{\partial}{\partial t}) e^{-\hat{\kappa}(t)}] | 0 \rangle = 0 \quad (3)$$

A well-chosen set of operators $\{\hat{O}\}$ provides linear systems of equations for the parameters $\{\kappa_{pq}\}$ to various orders in the perturbation. If we consider monochromatic perturbations of frequency ω we can extract the dynamical dipole polarizability from the linear response function

$$\delta\langle e^{\hat{\kappa}} \hat{r} e^{-\hat{\kappa}} \rangle(\omega) = \langle [\delta\hat{\kappa}, \hat{r}] \rangle(\omega) = \sum_{pq} \vec{r}_{pq} \delta D^{pq}(\omega) \quad (4)$$

$$\delta D^{pq}(\omega) = [\delta\kappa^T(\omega), D]^{pq} \quad (5)$$

The dipole moment is expanded in terms of atomic- and bond contributions

$$\langle -\vec{r}_C \rangle = - \sum_{AB} \sum_{\substack{l \in A \\ m \in B}} (\vec{r}_{AB})_{lm} D^{lm} + \sum_A Q_A (\vec{R}_A - \vec{R}_C) \quad (6)$$

with the localized LoProp basis, where \vec{r}_{AB} is the electronic coordinate with respect to the midpoint of the bond $A - B$ and the frequency-dependent variation can be written

$$\delta\langle -\vec{r}_C \rangle(\omega) = \sum_{AB} \sum_{l \in A, m \in B} -(\vec{r}_{AB})_{lm} \delta D^{lm}(\omega) + \sum_{AB} \Delta Q_{AB}(\omega) (\vec{R}_A - \vec{R}_B) \quad (7)$$

ΔQ_{AB} is a charge-transfer matrix satisfying

$$\sum_B \Delta Q_{AB}(\omega) = \delta Q_A(\omega) \quad (8)$$

where the local charge response to the external perturbation is

$$\delta Q_A(\omega) = - \sum_{l \in A} \delta D^l(\omega) \quad (9)$$

This leads to localized dynamical polarizability of the form

$$\alpha_{AB}(\omega) = - \sum_{\substack{l \in A \\ m \in B}} \vec{r}_{lm} \delta D^{lm}(\omega) + \Delta Q_{AB}(\omega) (\vec{R}_A - \vec{R}_B) \quad (10)$$

completely analogous to the static case.

Computational Details

The ground state and the frequency dependent first order perturbed electronic densities are calculated for the TIP3P water model using the program DALTON²⁵ by the means of analytic response theory at the TDHF²⁶ level. The LoProp approach is then used to calculate the charges q , dipole moments \mathbf{p} and polarizabilities $\alpha(\omega)$ of the atomic sites using a program implemented in Python.²⁷ Since the molecular polarizability is sensitive to the choice of the basis set, we do an initial study of the molecular $\alpha^{\text{Molecular}}(\omega)$ using different basis sets and frequencies. For the evaluation of the properties in the water clusters $\alpha^{\text{Cluster}}(\omega)$, we take interest in the average (isotropic) polarizability

$$\bar{\alpha} = \frac{1}{3} \sum_i \alpha_{ii} \quad (11)$$

In our Silberstein-Applequist implementation, we represent each water molecule using a set of inducible point dipoles. The point dipoles belonging to the same molecule can not interact with each other because their interaction has already been accounted for at the time the localized properties were derived. Furthermore, the localized properties are additive and sum up to the molecular net property. Previous works²⁸ that implement the Silberstein-Applequist model assign

the atomic polarizabilities semi-empirically from experimental data, and fit these to the molecular polarizabilities derived from QM. In addition these models incorporate a scaling of the interatomic distances. In this way the polarizability can be obtained for large molecules such as proteins if fitting is performed to large sets of experimentally available data. Our model differs in that the intermolecular properties are evaluated from the *ab initio* derived properties of each isolated system.

Water

To obtain configurations of the water clusters, we perform an NVT ensemble molecular dynamics calculation for the TIP3P water model by employing the AMBER03²⁹ force field in the GROMACS computational package³⁰. The TIP3P model is parameterized to reproduce several experimental data of water and is routinely used as the explicit solvent representation in standard simulations of biological systems. We solvate 340 water molecules in a 21.64 \AA^3 periodic box and perform a steepest descent energy minimization using 200 steps, following an equilibrium run of 200 ps using temperature and pressure coupling. The temperature is 298.15 K and we use the leapfrog integrator with a 2 fs time-step. Finally, the pressure coupling is turned off and the system runs for a full 20 ns. We pick up snapshots during the last 10 nanosecond run at a 100 ps spacing, ensuring no statistical correlation between the configurations.

To pick N water molecules from a trajectory point P , we use the following systematic approach. Firstly, all water molecules are ordered in ascending order with respect to the closest distance to the center of the simulation box, and then the first N molecules are chosen in that ordered list. For the QMMM calculations, we thus take N water molecules in the QM region, and the remaining M molecules in the sorted list which creates a QMMM setup where the core QM region is embedded by the MM region which is systematically varied by choosing M .

The localized properties are derived for the TIP3P model with a predefined geometry $r_{\text{O-H}} = 0.9572 \text{ \AA}$ and $\theta_{\text{H-O-H}} = 104.52^\circ$ and used as a template for all other water molecules in a cluster. We thus transform all the properties from the template to the target geometries in the water cluster

by summing over the components of the dipole moment vector and polarizability tensor multiplied by the indices of the transformation matrix obtained from the Euler angles between the template and the target molecule. Using Einstein notation, for e.g. the α tensor we have:

$$\alpha_{ij}^{Transformed} = R_{ix}R_{jy}\alpha_{xy}^{Template} \quad (12)$$

where R is the rotation matrix that transforms the template into the target molecules coordinate frame.

$$R = R_z R_y R_x \quad (13)$$

For simplicity, the template model is situated in the xz -plane with the dipole moment vector pointing along the z -axis as seen in Figure 1.

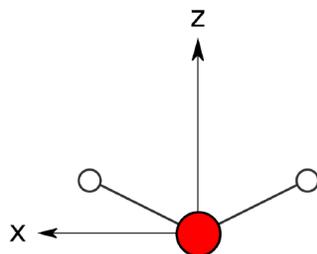


Figure 1: Template orientation in space.

We use two different point-dipole types to represent the water molecules in our classical Silberstein-Applequist implementation. For the first type, referred to as the Oxygen-centered type, the net properties of the water molecule obtained from a QM calculation are put on the oxygen atom. The second type is referred to as the LoProp type, where the corresponding QM calculated properties are put on each atom in the water molecule with its own local charge, dipole moment, and frequency-dependent polarizability, as obtained by the LoProp transformation. For the water calculations, the frequencies used are $\omega(\lambda) = 0.0(\infty)$, 0.0239 (1907 nm), 0.0428 (1064 nm) and 0.0774 (589 nm).

Tryptophan

The protein *cyclooxygenase-2* with the PDB entry 4COX.pdb³¹ is used to collect the configurations of the tryptophan residues. We use the same ensemble, parameters and procedure as in the water simulation, with the addition of counter-ions appropriate for the physiological pH-level, and a total simulation time of 4 ns. Although in reality a small area of the protein is absorbed onto the lipid membrane, for our purposes, the MD simulation only serves as a tool to gather a data set for the statistical analysis of $\alpha^{\text{LoProp}}(\omega)$, and thus the membrane-protein interaction is irrelevant. We take the configurations from the trajectory for all 6 tryptophan residues at two different time intervals $\Delta t = 1$ ps, and $\Delta t = 200$ ps, respectively, and thus compare the $\alpha^{\text{LoProp}}(\omega)$ as a function of both frequency and correlation time averaging. Since the amino acids need to fulfill valency, we add the HCO- group with a hydrogen from the preceding residue and the -NH₂ group capped with a hydrogen in the following group, respectively, for all residues. We furthermore calculate the QMMM polarizability $\alpha^{\text{QMMM}}(\omega)$ of one chosen tryptophan residue, by explicitly including all the residues within a 10 Å center-of-mass distance of the chosen tryptophan residue. For each residue in the MM region, we represent each atom with a charge and isotropic polarizability. We perform the full polarizable force field derivation with the means of the MFCC^{32,33} method, where we use the con-cap group H₃C-CO-NH-CH₃, by taking the H₃C-CO- atoms from preceding amino acids, and the -NH-CH₃ from the following amino acids. For all residues included in the MM-region. For the tryptophan residue in the QM-region, we use the TDHF method in DALTON for the calculation of polarizabilities, using the ANO³⁴ basis set with a $2s1p(\text{H})/3s2p1d(\text{C/N/O})$ contraction. The frequencies used for the tryptophan calculations are $\omega(\lambda) = 0.0(\infty), 0.0428(1064\text{ nm}), 0.0911(500\text{ nm}), 0.114(400\text{ nm})$.

Results and Discussion

Polarizable force fields form a prerequisite for accurate MD and quantum-classical QMMM simulations alike. In the case of QMMM property calculations it seems that the LoProp approach,¹⁴

originally presented by Gagliardi et al, has become a most important vehicle for generating force fields that include polarization. While the LoProp scheme is most flexible for the construction of decomposed force fields, five main types has emerged in QMMM applications³⁵: i) atom distributed charges; ii) atom distributed charges with a central polarizability; iii) atom distributed charges and polarizabilities; iv) atom and bond distributed charges and polarizabilities; v) atom and bond distributed charges and polarizabilities with multipoles expanded at each atomic site (normally truncating at octupoles). These different MM models have been particularly well tested in case of the water solvent³⁶⁻³⁸. The precision of the more refined models comes with a computational cost, and is also directed by the type of property considered in the QMMM calculations. An often made observation is that the third level is needed for obtaining the necessary precision, *i.e.* decomposing the polarizabilities into atomic contributions. It has also been noted on several occasions that a good MM parametrization of water (and possibly any other solvent), relieves the need to include waters in the QM box, thereby making the QMMM partitioning remain well-defined over simulation time, and substantially reducing cost³⁹. We demonstrate the performance of our model with the following studies: the basis set dependence of the polarizability for a single water molecule; the relative error in our point-dipole model for gas phase water clusters with respect to the QM results; QMMM treatment of the same water clusters embedded in a polarizable MM environment where each particle is represented with our derived frequency dependent force fields, and lastly; the frequency dependent polarizability in a tryptophan residue, both evaluated in gas phase, and as a property determined with QMMM where we include the closest amino acid residues, represented by atomic charges and isotropic polarizabilities derived with our method described in the theory section.

In the theory section it was stated that the polarizable frequency dependent properties are only meaningful when evaluated well below resonance. In Figure 2, the mean polarizability $\bar{\alpha}(\omega)$ was calculated quantum mechanically, and using the classical model, for a random cluster consisting of 10 water molecules. It can be seen that the polarizability diverges and further oscillates at field strengths larger than 0.3 A.U. (152 nm). It can also be noted that the frequency dependent

polarizabilities capture the dispersion of the quantum mechanical result up until 0.25 A.U., where the relative error is 5.7% and 6.8% for the Oxygen and LoProp model, respectively. Since the electric field strengths chosen for the calculations are below 0.0774 A.U. in the case of water, and 0.114 A.U. for the amino acids of the protein, we thus conclude that the frequency dependent properties can accurately capture the dispersion of the MM-region, and are physically meaningful.

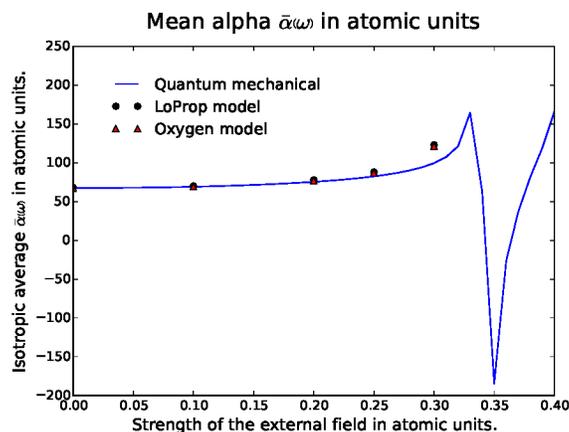


Figure 2: The mean polarizability evaluated for a random water cluster configuration consisting of 10 molecules at different frequencies of the external field. Atomic units are used.

In Figure 3 we plot the average value of the total molecular dynamic polarizability $\alpha^{\text{Mol}}(\omega)$, evaluated at different frequencies using different basis sets. It is seen that the $\alpha^{\text{Mol}}(\omega)$ increases systematically with respect to the chosen basis set and frequency. Larger ANO basis sets in general yield more accurate molecular properties, and are therefore more feasible for small models such as water. In large systems a big basis set can be computationally expensive, which is why a 6-31+G* type of contraction is commonly used for amino acids, as it is a good compromise between accuracy and computational resources required. The ANO basis set corresponding to the $2s1p$ H / $3s2p1d$ O contraction yields similar molecular polarizabilities as the cc-pVTZ one, but comes with the drawback of added computational time. However, since we are interested in benchmarking our point-dipole model using localized frequency-dependent properties, we choose to only evaluate the ANO type basis functions shown in figure 3 for the water clusters. This is because the transformation of the properties to the LoProp basis is most conveniently done if one starts with an ANO

basis set directly.

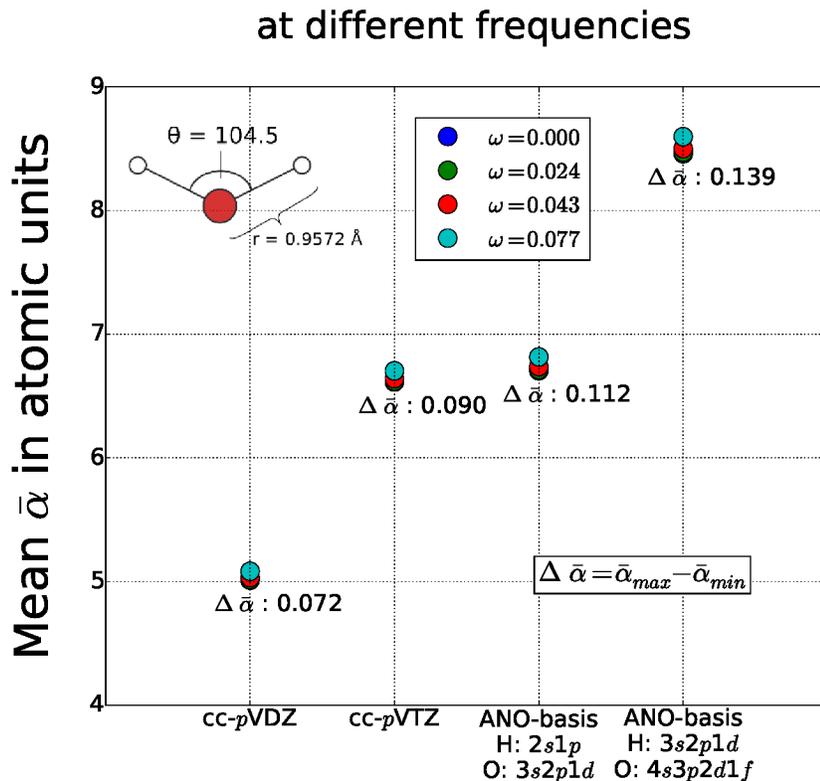


Figure 3: The mean polarizability evaluated at different frequencies for the TIP3P water model.

To compare how the chosen basis set affects the frequency dependent properties of water clusters, we calculate the relative error using our point dipole model defined as:

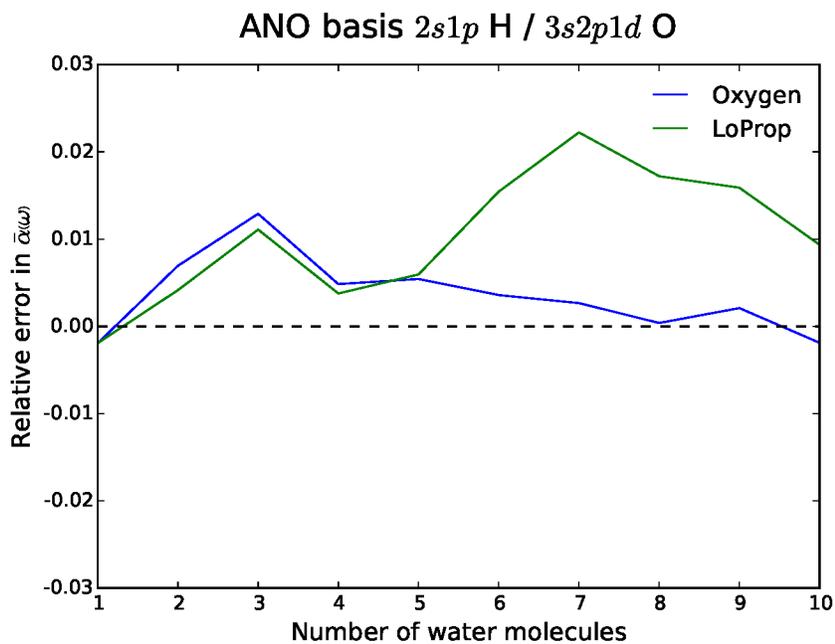
$$\bar{\alpha}^{\text{Error}} = \frac{\bar{\alpha}^{\text{Model}} - \bar{\alpha}^{\text{QM}}}{\bar{\alpha}^{\text{QM}}} \quad (14)$$

where $\bar{\alpha}^{\text{Model}}$ and $\bar{\alpha}^{\text{QM}}$ are the average polarizabilities determined with our point-dipole model and TDHF, respectively.

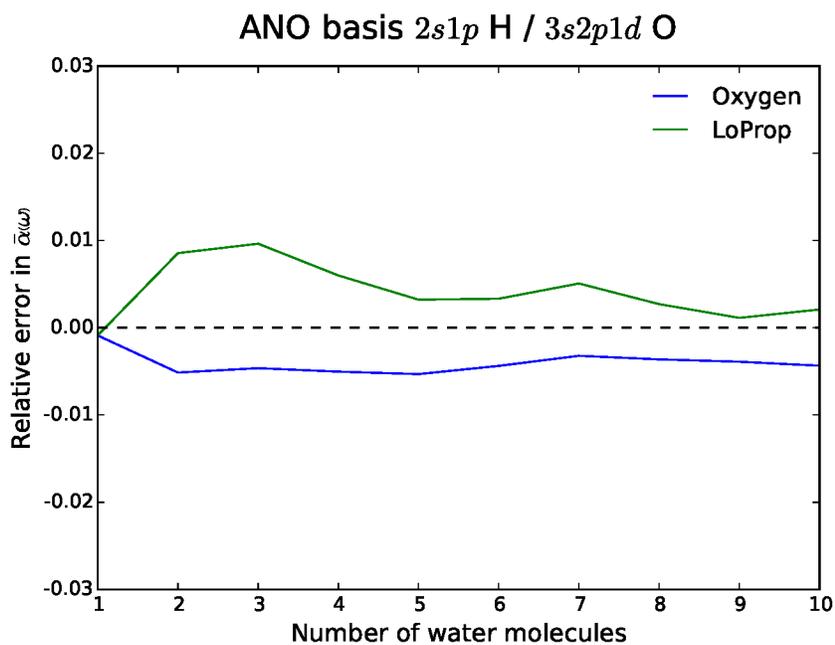
In Figure 4, we plot the total static cluster polarizability $\alpha^{\text{Cluster}}(\omega = 0)$ for one random snapshot taken from the MD simulation and for an averaging over 10 uncorrelated snapshots, respectively, evaluated both for the oxygen-centered and the LoProp type as a function of the cluster size. From Figures 4a–4b it is seen that the classical point-dipole model predicts the QM result rather well for this particular basis set, with a maximum error around 2 % for the LoProp type

water. Moreover, by performing an average over uncorrelated MD configurations, we see in Figure 4b that both the oxygen-centered and LoProp type converge to the TDHF result for 7-10 water molecule cluster sizes with an error of $< 0.5\%$.

In Figure 5, we plot the frequency dependent $\bar{\alpha}^{Cluster}(\omega)$, with $\omega = 0.0774$, evaluated with properties obtained at $\omega = 0.0$, and $\omega = 0.0774$, respectively. This means that in our point-dipole model, the point polarizabilities of the oxygen-centered and the LoProp type are derived for two different frequencies, labeled as O-centered ω and LoProp ω , respectively, in Figure 5. Both the blue (O-centered) and the red (LoProp) lines shifts upward when including the frequency dependence of the polarizability in the model, which shows that both models capture the frequency dispersion of the water clusters as predicted by TDHF. Furthermore, if the frequency dependence of $\bar{\alpha}^{Molecule}(\omega)$ is not included in the point-dipole model, the error is twice as large for the oxygen-centered model compared to the LoProp model. This is due to the LoProp model slightly overestimating the total $\bar{\alpha}^{Cluster}(\omega)$, as compared to the Oxygen-centered type.



(a) One random MD snapshot.



(b) Averaging over 10 uncorrelated MD snapshots.

Figure 4: The relative error of the mean $\bar{\alpha}^{Cluster}(\omega = 0)$.

For the larger ANO basis set corresponding to a $3s2p1d(H)/4s3p2d1f(O)$ contraction, we plot

the relative error of the static cluster polarizability $\bar{\alpha}^{Cluster}(\omega = 0)$ averaged over 10 snapshots in Figure 6. For this basis set, both the oxygen-centered type and LoProp type overestimate the polarizability with respect to the TDHF method. In Figure 7, analogous with the smaller ANO $2s1p$ (H) / $3s2p1d$ (O), contraction, we plot the frequency dependent $\bar{\alpha}^{Cluster}(\omega)$ as a function of the cluster size, to see how the point-dipole model behaves when the properties are frequency dependent. It can be seen that for this basis set, the relative error of the frequency dependent $\bar{\alpha}^{Cluster}(\omega)$ does not converge as in the case of the smaller basis set and already for clusters larger than 6 water molecules in size the relative error is in the 4 % and 5 % range for the oxygen and LoProp type, respectively. The added diffuse functions on the oxygen and hydrogen atoms for the larger basis set thus snowballs when more and more water molecules interact, causing the point-dipole model to overestimate the total $\bar{\alpha}^{Cluster}(\omega)$. We see, however, that the overestimation does not diverge and can be assumed to be stable even for larger clusters.

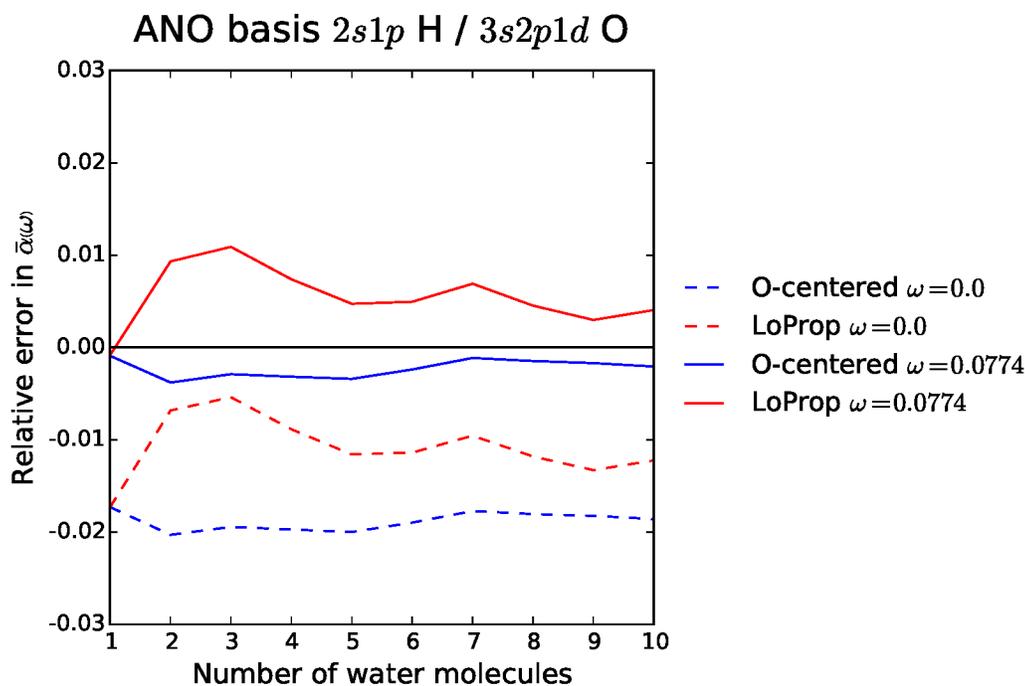


Figure 5: The relative error of the cluster polarizability $\bar{\alpha}^{Cluster}(\omega = 0.0774)$ using the $2s1p$ (H) / $3s2p1d$ (O) ANO basis set.

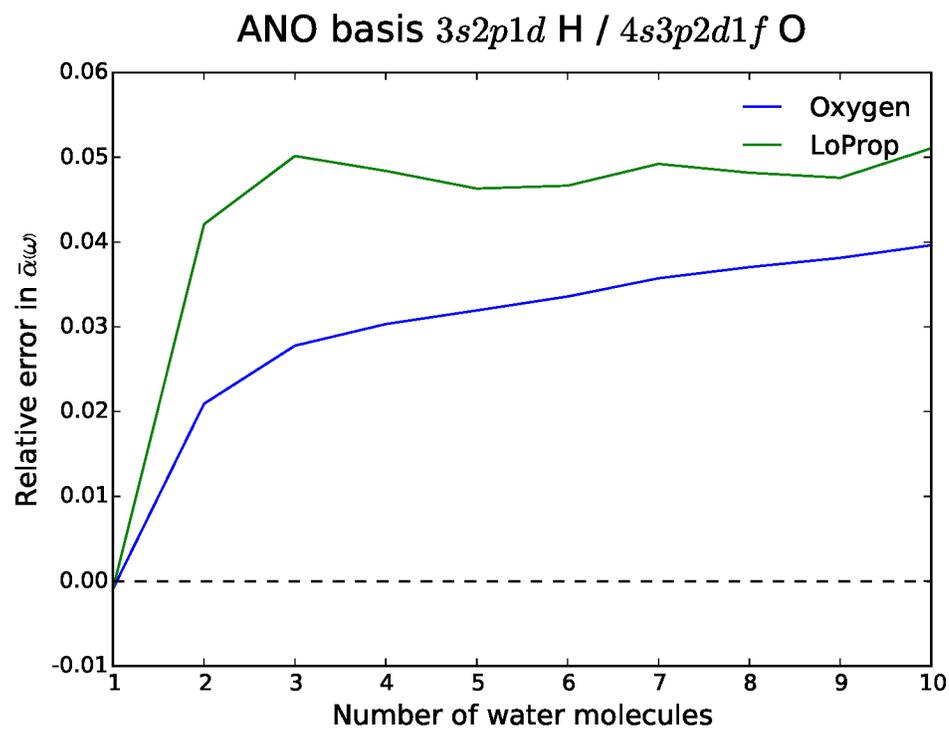


Figure 6: The relative error of the static cluster polarizability $\bar{\alpha}^{Cluster}(\omega = 0)$ using the $3s2p1s$ (H) / $4s3p2d1f$ (O) ANO basis set.

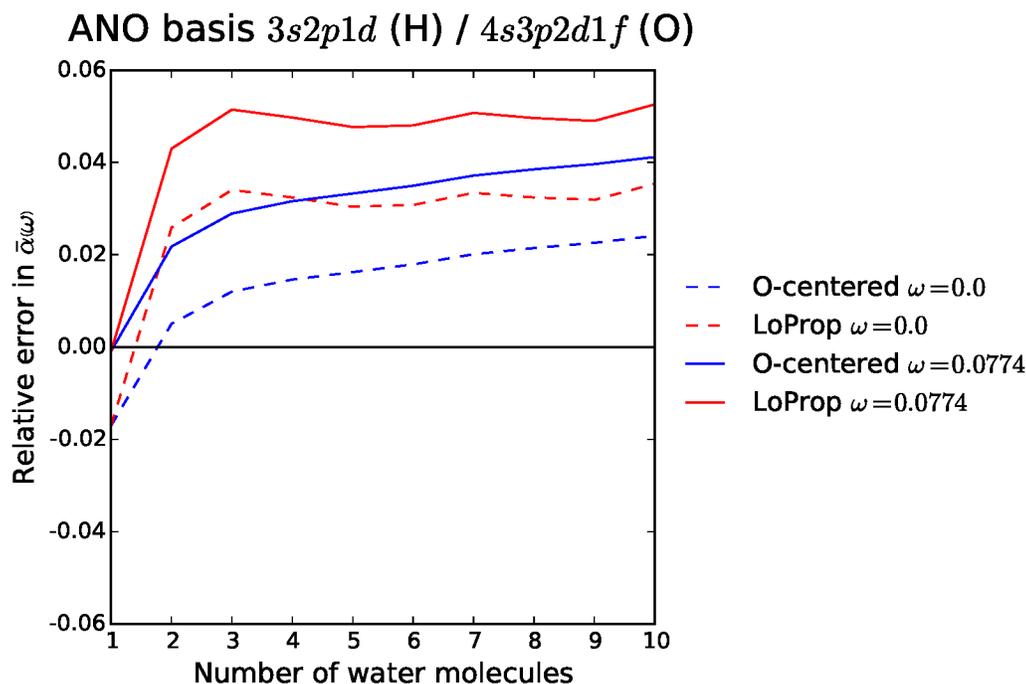


Figure 7: The relative error of the cluster polarizability $\bar{\alpha}^{Cluster}(\omega = 0.0774)$ using the $3s2p1s/4s3p2d1f$ ANO basis set.

We also calculated the relative error with respect to the choice of basis by using the larger $3s2p1s$ (H) / $5s4p3d2d$ (O) ANO-contraction, where more basis functions have been added on the oxygen atom. The results are shown in Figure 8. It is evident that increasing the basis set size does not necessarily introduce a larger error for the frequency dependent polarizability evaluated with the classical model. The accuracy is improved when one utilizes more basis functions on the oxygen relative to the hydrogen atoms, as one can see by comparing Figures 6-8.

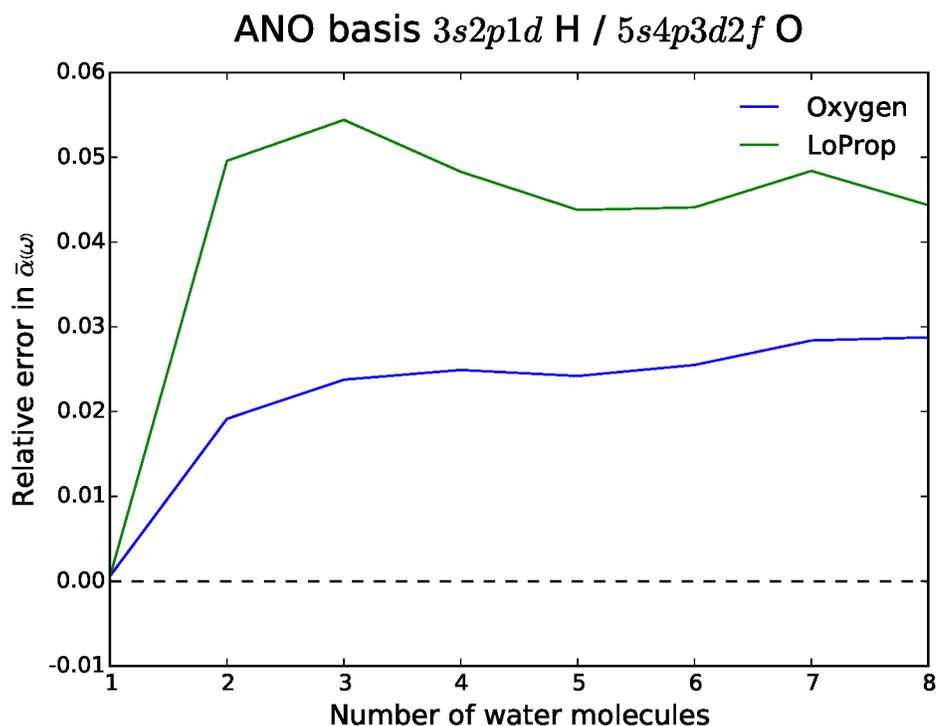


Figure 8: The relative error of the static cluster polarizability $\bar{\alpha}^{Cluster}(\omega = 0)$ using the $3s2p1s$ (H) / $5s4p3d2f$ (O) ANO basis set.

For the QMMM results, we use the $2s1p / 3s2p1d$ basis set to compute the total $\bar{\alpha}^{Cluster}(\omega)$ for a QM system inside an MM polarizable embedding, with properties derived using the same basis set. The QMMM results for the water clusters are summarized in Figures 9–10. It is seen that the effect of including frequency dependent polarizabilities in the MM region does not affect the $\bar{\alpha}^{Cluster}(\omega = 0.0774)$ with any significance as the dotted circles overlap the lines. If the amount of water molecules in the MM-region is small, then the mean polarizability for one QM molecule is larger than for the gas-phase, but as the MM-region is systematically expanded, the mean polarizability approaches the gas-phase value for the oxygen-centered type, and a slight underestimation of the gas-phase value for the LoProp type is seen.

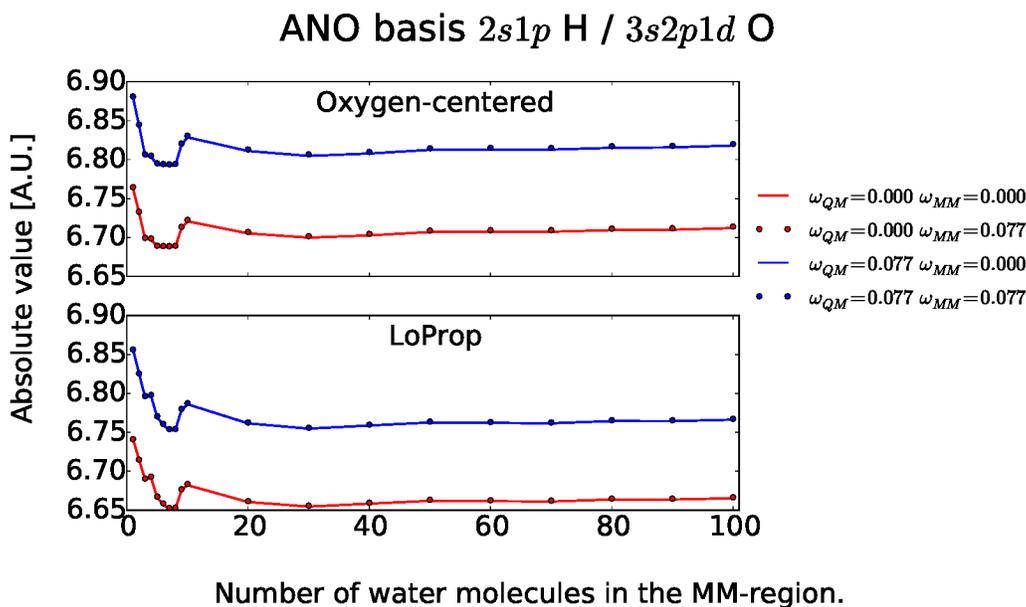


Figure 9: $\bar{\alpha}(\omega)$ obtained by varying the size of the MM region for one QM water molecule.

In Figure 10 we plot the mean $\bar{\alpha}^{Cluster}(\omega)$ as a function of varying the amount of QM water molecules, by keeping the MM-region fixed at 100 molecules. Two things can be noted here. Firstly, as in the change of the MM-region, there is no significant difference of calculating the QMMM property by having the MM properties derived at different frequencies. This is expected, as the polarizability for the basis set $2s1p / 3s2p1d$ varies at most 1.7% at the chosen frequencies, and thus barely perturbs the QM wave-function. This is highlighted by the dots overlapping the lines Figure 10b. Secondly, the frequency dependent $\bar{\alpha}(\omega)$ varies nearly linearly with respect to the size of the QM region. Since the polarizability of water is not highly dependent on the field applied, it is safe to assume that the larger basis set will behave in the same qualitative manner, highlighting the fact that for water, the frequency dependent polarizability is not significant in QMMM type calculations.

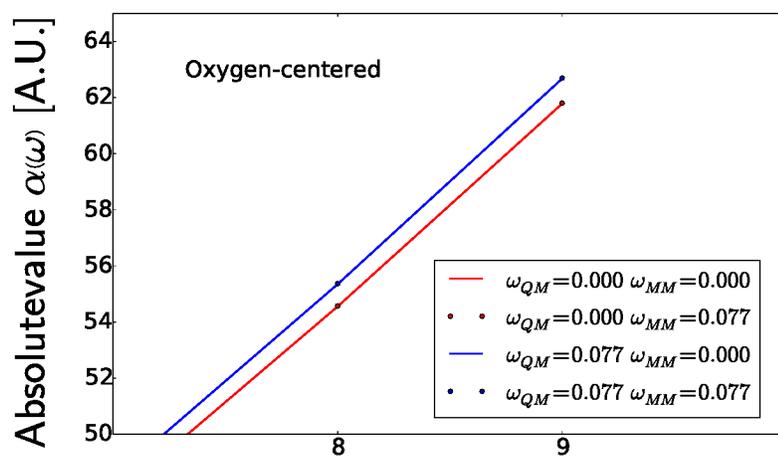
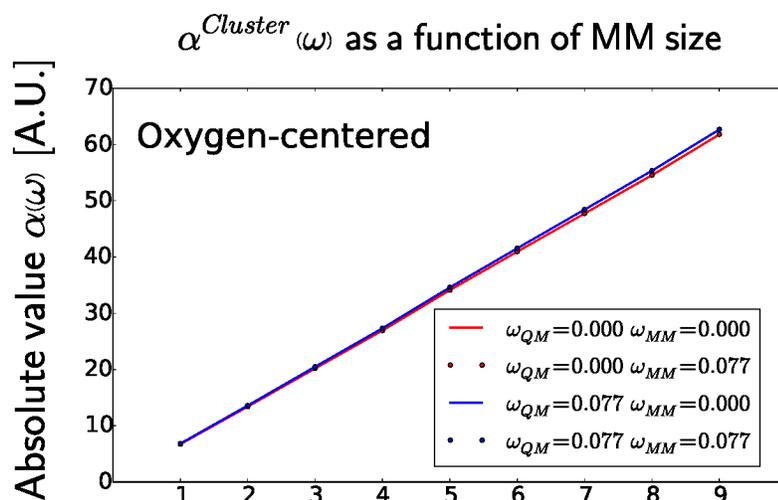


Figure 10: $\bar{\alpha}(\omega)$ obtained by varying the size of the QM region and using 100 MM water molecules.

The frequency dependence of the isotropic $\alpha^{\text{LoProp}}(\omega)$ for all the atoms in tryptophan is summarized in Figures 12 - 13. The atom types marked in red in Figure 11 denote the atoms which participate in the aromatic bonds formed by the planar side-chain.

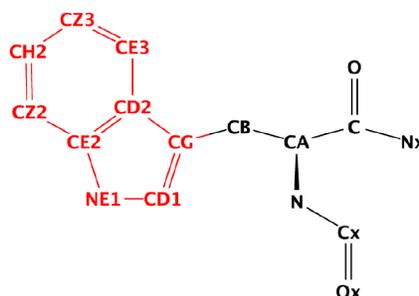


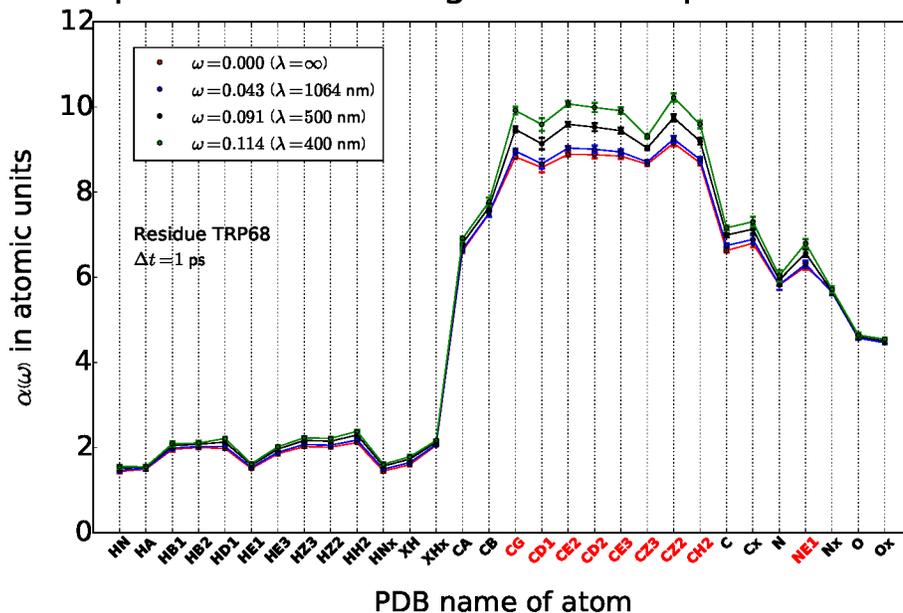
Figure 11: The tryptophan residue capped with $HCO-$ and $-NH_2$ groups. The atom types in red denote aromatic the ones.

The frequency dependence of each atom is indeed significantly higher for the aromatic atomic types, independent of which time separation the configurations are chosen at or which particular tryptophan residue the atom belongs to. Since the electrons in the aromatic system will be delocalized, and more easily perturbed, the polarizability will be higher there than for other atoms. This is captured by our model, seen as the polarizability of the NE1 atom is more easily perturbed by an external field than the N atom. In Figures 12a and 12b, the difference in properties is seen by comparing the time separation of each frame from the MD simulation. Not surprisingly, the deviation is almost 0 for the 1 ps case for all but the CD1 and CD2 atom types. This is because the motions of the atoms are highly correlated and each property will vary by a small amount. By taking the configurations at 200 ps apart, however, a larger fluctuation is seen for the localized properties.

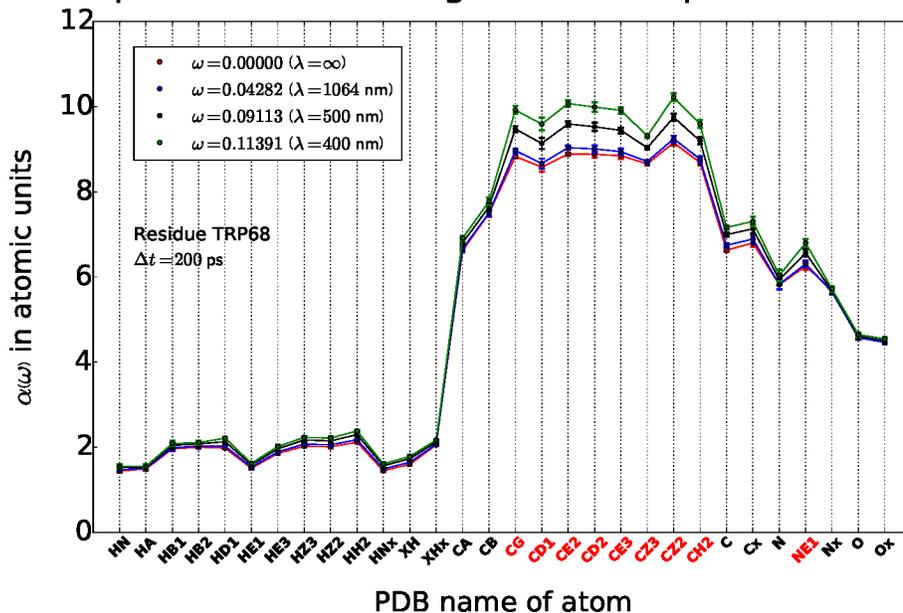
An averaging over all residues for both time separations is depicted in Figure 13. The difference between two random snapshots becomes evident, which shows that the orientation of the residues and the chemical surroundings are more important than the averaging for one residue over time. This means that the force fields obtained are quite stable over time with respect to each residue, and moreover depend on the configuration of close lying neighboring residues. A total averaging over all 6 residues and 10 snapshots is presented in Figure 14. The average deviation in atomic units for the aromatic atoms, at the frequency $\omega = 0.0774$ is 0.13 and 0.12, for the 1 ps, and 200 ps time separation, respectively. The isotropic character of the localized frequency dependent $\alpha(\omega)$

can be approximated as an average over a few residues and snapshots. This result suggests that the frequency dependent polarizabilities may also be transferable between different systems.

In Figure 15, we plot the molecular frequency dependent mean polarizability for a tryptophan residue, embedded in an MM polarizable environment, represented by point charges and isotropic polarizabilities on each atomic site in the MM-region, for residues within a center-of-mass distance of 10 Å from tryptophan. The effect of having a frequency dependent polarizable embedding shifts the results of about 2 % for the frequency dependent polarizability determined by QMMM with respect to using a polarizable embedding obtained at the static frequency. Our studies on the water molecule shows that the frequency dependent polarizability varies more when using a larger, more diffuse basis set. Furthermore, The inclusion of the closest MM environment had the most significant effect on the QM subsystem, while an extended MM environment predicted the gas-phase value for the polarizability of the Oxygen-centered type water and a slight underestimation of the gas-phase value for the LoProp type.

Isotropic $\alpha^{LoProp}(\omega)$ averaged over snapshots for TRP68

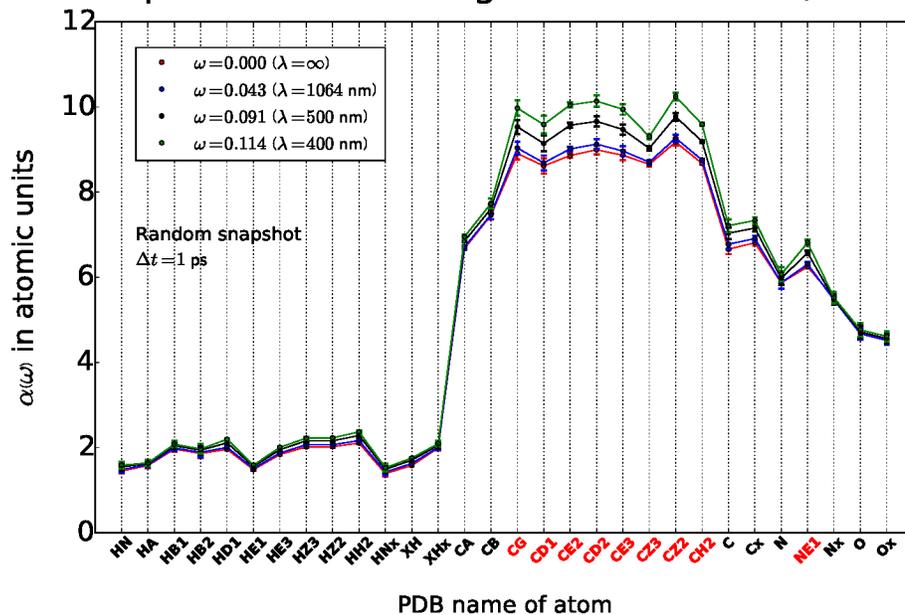
(a) 1 ps spacing.

Isotropic $\alpha^{LoProp}(\omega)$ averaged over snapshots for TRP68

(b) 200 ps spacing.

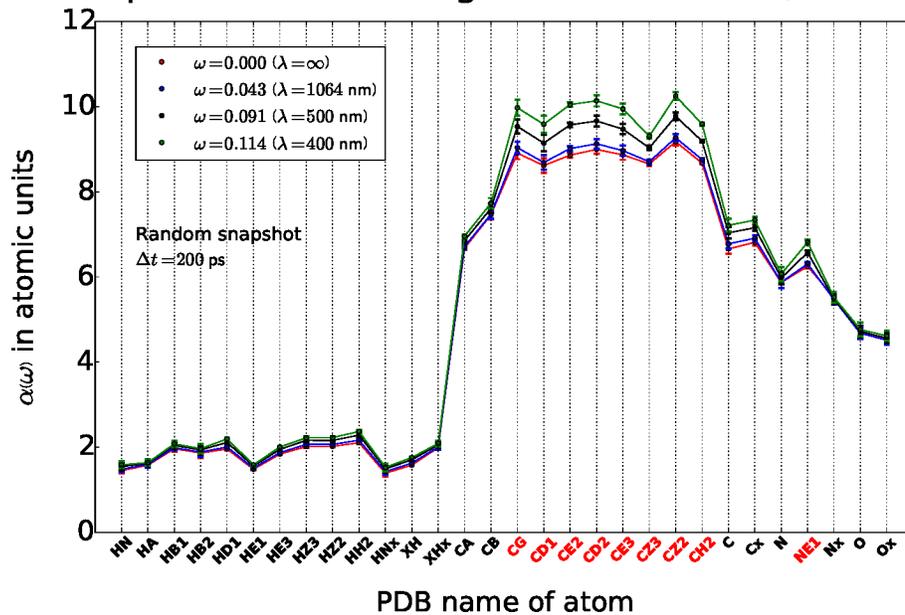
Figure 12: $\alpha^{LoProp}(\omega)$ averaged over 10 snapshots for the residue TRP68.

Isotropic $\alpha^{LoProp}(\omega)$ averaged over residues, $\Delta t = 1$ ps



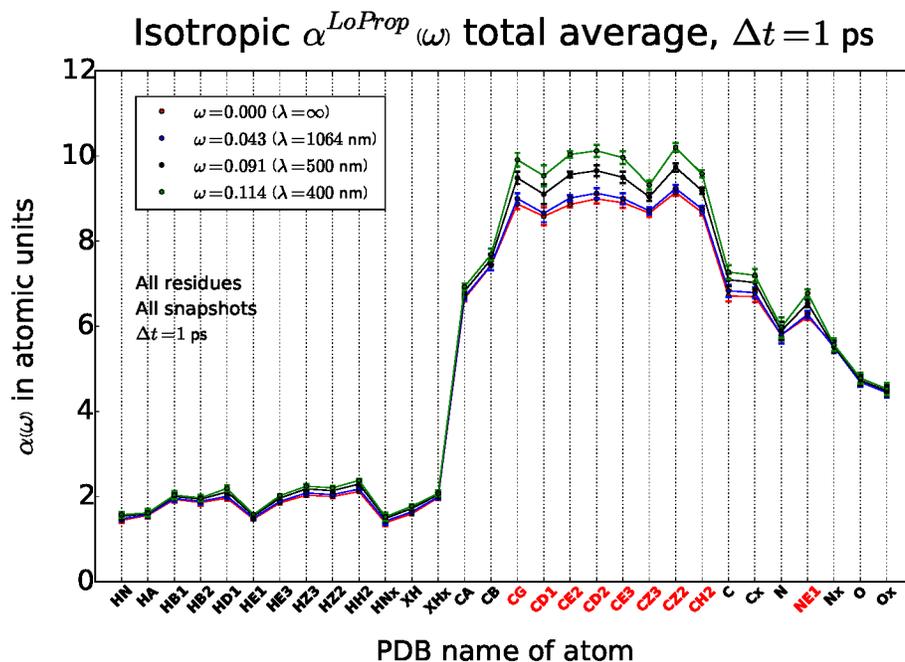
(a) 1 ps spacing.

Isotropic $\alpha^{LoProp}(\omega)$ averaged over residues, $\Delta t = 200$ ps

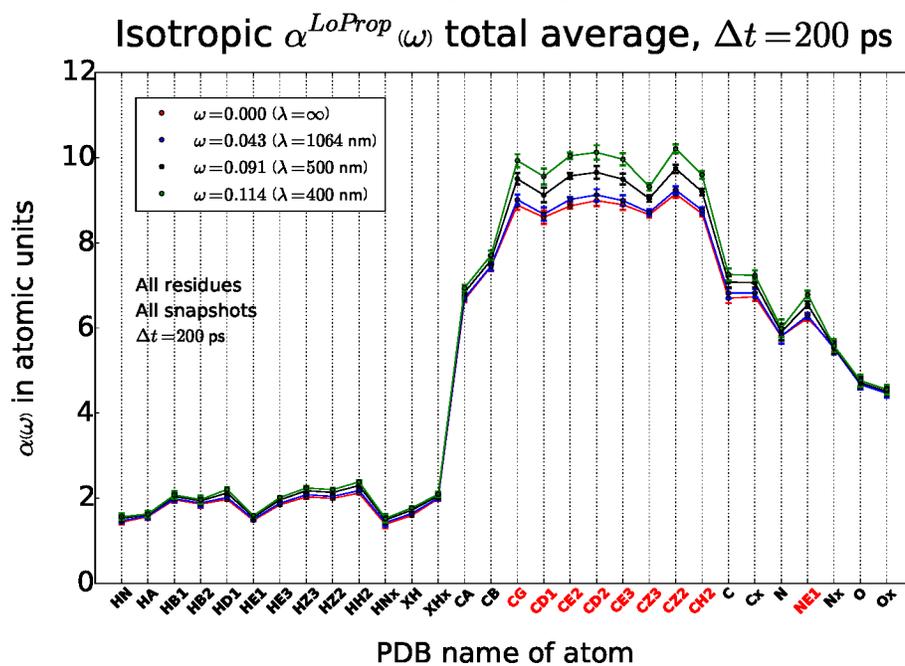


(b) 200 ps spacing.

Figure 13: $\alpha^{LoProp}(\omega)$ averaged over all residues for one random snapshot.



(a) 1 ps spacing.



(b) 200 ps spacing.

Figure 14: $\alpha^{LoProp}(\omega)$ averaged over all residues and 10 snapshots.

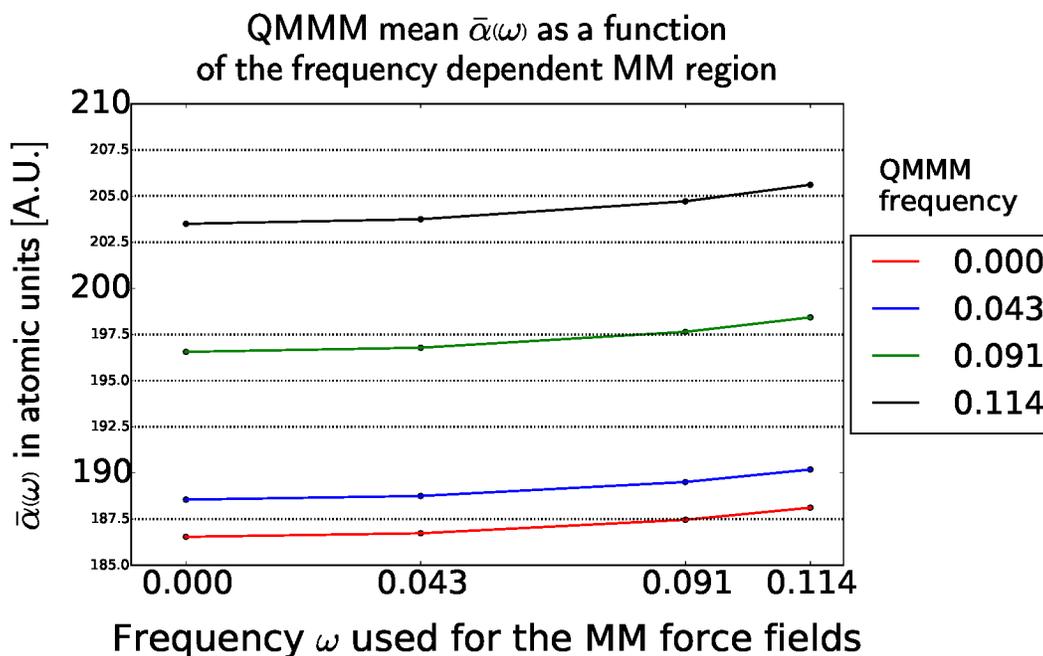


Figure 15: The average polarizability of the TRP residue evaluated at 4 different frequencies, where the x -axis corresponds to different polarizable embeddings, derived at the corresponding frequencies.

Conclusions

We have presented for the first time a way of incorporating quantum mechanical granulation of molecular mechanics environment through the localization of frequency-dependent polarizabilities. Our work was motivated by that such frequency-dependent molecular mechanics force fields allow for an instantaneous, or optical, polarization in the molecular mechanics part by the electronic degrees of freedom in the quantum part of a quantum mechanics - molecular mechanics (QMMM) model and makes it possible to extend the QMMM properties to cover the full QMMM simulation volume. This is thus a generalization of current QMMM technology where the embedding MM part acts as a perturber of the property determined by the QM. Furthermore, these force fields allow for calculations of frequency-dependent properties for very large clusters that are evaluated classically. We address the latter aspect by implementing the Silberstein-Applequist model for many interacting induced dipoles for the linear frequency-dependent polarizability. The performance and precision of the method is evaluated through studying a few selected cases. Specifically,

we investigated the purely classical intermolecular polarizability using *ab-initio* derived properties, the effect caused by frequency dependency in the MM environment in a QMMM study of water clusters, and the effect of the dynamic local polarizability in larger delocalized aromatic systems. The outcome from incorporating frequency-dependent MM properties shows no significant effect in the case of water molecules for the QMMM evaluated property $\bar{\alpha}(\omega)$, since the polarizability is not highly sensitive to the applied field frequency. For larger systems, and especially polarizable systems, the effect of frequency dependence in the MM environment becomes non-negligible. We have shown that our LoProp implementation can extract the frequency-dependent properties in complex structures, and furthermore be used in general QMMM calculations for the determination of properties within the polarizable embedding framework. We tested the methodology for a small number of systems and parameters, but as the method can directly be implemented with any quantum mechanical wave-function and basis set, an extension to the study of any other systems is possible without the need of re-writing existing code. We can thus foresee that frequency-independent force fields in QMMM calculations can routinely be replaced by frequency-dependent force fields with gain of generality and without loss of efficiency.

Acknowledgments

The simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at National Supercomputing Center (NSC) and High Performance Computing Centre North (HPC2N), projects “Multiphysics Modeling of Molecular Materials”, SNIC 023/07-18, and “Design of Force Fields for Theoretical Spectroscopy”, SNIC-2014-1-179.

References

- (1) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force fields: Application to liquid water. *The Journal of Chemical Physics* **1994**, *101*, 6141.
- (2) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926.
- (3) Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics* **2000**, *112*, 8910.
- (4) Lamoureux, G.; MacKerell, A. D.; Roux, B. A simple polarizable model of water based on classical Drude oscillators. *The Journal of Chemical Physics* **2003**, *119*, 5185.
- (5) van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *The Journal of Physical Chemistry A* **2001**, *105*, 9396–9409.
- (6) Rinkevicius, Z.; Li, X.; Sandberg, J. A. R.; Mikkelsen, K. V.; Ågren, H. A Hybrid Density Functional Theory/Molecular Mechanics Approach for Linear Response Properties in Heterogeneous Environments. *Journal of Chemical Theory and Computation* **2014**, *10*, 989–1003.
- (7) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology* **1976**, *103*, 227–249.
- (8) QM/MM Methods for Biomolecular Systems - Senn - 2009 - Angewandte Chemie International Edition - Wiley Online Library. <http://onlinelibrary.wiley.com/doi/10.1002/anie.200802019/pdf>.

- (9) Yoo, S.; Zahariev, F.; Sok, S.; Gordon, M. S. Solvent effects on optical properties of molecules: A combined time-dependent density functional theory/effective fragment potential approach. *The Journal of Chemical Physics* **2008**, *129*, 144112.
- (10) Sneskov, K.; Schwabe, T.; Kongsted, J.; Christiansen, O. The polarizable embedding coupled cluster method. *The Journal of Chemical Physics* **2011**, *134*, 104108.
- (11) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited States in Solution through Polarizable Embedding. *Journal of Chemical Theory and Computation* **2010**, *6*, 3721–3734.
- (12) Schwabe, T.; Sneskov, K.; Haugaard Olsen, J. M.; Kongsted, J.; Christiansen, O.; Hättig, C. PERI-CC2: A Polarizable Embedded RI-CC2 Method. *Journal of Chemical Theory and Computation* **2012**, *8*, 3274–3283.
- (13) Hedegård, E. D.; List, N. H.; Jensen, H. J. A.; Kongsted, J. The multi-configuration self-consistent field method within a polarizable embedded framework. *The Journal of Chemical Physics* **2013**, *139*, 044101.
- (14) Gagliardi, L.; Lindh, R.; Karlström, G. Local properties of quantum chemical systems: the LoProp approach. *J. Chem. Phys.* **2004**, *121*, 4494–500.
- (15) Pedersen, M. N.; Hedegård, E. D.; Olsen, J. M. H.; Kauczor, J.; Norman, P.; Kongsted, J. Damped Response Theory in Combination with Polarizable Environments: The Polarizable Embedding Complex Polarization Propagator Method. *Journal of Chemical Theory and Computation* **2014**, *10*, 1164–1171.
- (16) Murugan, N. A.; Kongsted, J.; Rinkevicius, Z.; Ågren, H. Color modeling of protein optical probes. *Physical Chemistry Chemical Physics* **2012**, *14*, 1107.
- (17) Aidas, K.; Møgelhøj, A.; Nielsen, C. B.; Mikkelsen, K. V.; Ruud, K.; Christiansen, O.; Kongsted, J. Solvent Effects on NMR Isotropic Shielding Constants. A Comparison between Ex-

- PLICIT Polarizable Discrete and Continuum Approaches. *The Journal of Physical Chemistry A* **2007**, *111*, 4199–4210.
- (18) Nielsen, C. B.; Christiansen, O.; Mikkelsen, K. V.; Kongsted, J. Density functional self-consistent quantum mechanics/molecular mechanics theory for linear and nonlinear molecular properties: Applications to solvated water and formaldehyde. *The Journal of Chemical Physics* **2007**, *126*, 154112.
- (19) Eriksen, J. J.; Olsen, J. M. H.; Aidas, K.; Ågren, H.; Mikkelsen, K. V.; Kongsted, J. Computational protocols for prediction of solute NMR relative chemical shifts. A case study of L-tryptophan in aqueous solution. *Journal of Computational Chemistry* **2011**, *32*, 2853–2864.
- (20) Harczuk, I.; Murugan, N. A.; Vahtras, O.; Ågren, H. Studies of pH-Sensitive Optical Properties of the deGFP1 Green Fluorescent Protein Using a Unique Polarizable Force Field. *Journal of Chemical Theory and Computation* **2014**, *10*, 3492–3502.
- (21) Olsen, J.; Jørgensen, P. Linear and Nonlinear Response Functions for An Exact State and for An MCSCF State. *J. Chem. Phys.* **1985**, *82*, 3235–3264.
- (22) Sałek, P.; Vahtras, O.; Helgaker, T.; Ågren, H. Density-Functional Theory of Linear and Nonlinear Time-Dependent Molecular Properties. *J. Chem. Phys.* **2002**, *117*, 9630–9645.
- (23) Applequist, J. A multipole interaction theory of electric polarization of atomic and molecular assemblies. *The Journal of Chemical Physics* **1985**, *83*, 809–826.
- (24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926.
- (25) The Dalton quantum chemistry program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4*, 269–284.

- (26) McLachlan, A. D.; Ball, M. A. Time-Dependent Hartree—Fock Theory for Molecules. *Reviews of Modern Physics* **1964**, *36*, 844.
- (27) Vahtras, O. LoProp for Dalton. 2014; <http://dx.doi.org/10.5281/zenodo.13276>.
- (28) Hansen, T.; Jensen, L.; Åstrand, P.-O.; Mikkelsen, K. V. Frequency-Dependent Polarizabilities of Amino Acids as Calculated by an Electrostatic Interaction Model. *Journal of Chemical Theory and Computation* **2005**, *1*, 626–633.
- (29) Case, D. et al. **2012**, AMBER 12, University of California, San Francisco.
- (30) van der Spoel, D.; Lindahl, E.; Hess, B.; the GROMACS development team, GROMACS User Manual version 4.6.5.
- (31) Kurumbail, R. G.; Stevens, A. M.; Gierse, J. K.; McDonald, J. J.; Stegeman, R. A.; Pak, J. Y.; Gildehaus, D.; Miyashiro, J. M.; Penning, T. D.; Seibert, K.; others, Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature* **1996**, *384*, 644–648.
- (32) Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy. *The Journal of Chemical Physics* **2003**, *119*, 3599.
- (33) Söderhjelm, P.; Ryde, U. How Accurate Can a Force Field Become? A Polarizable Multipole Model Combined with Fragment-wise Quantum-Mechanical Calculations. *The Journal of Physical Chemistry A* **2009**, *113*, 617–627.
- (34) Wildmark, P. O.; Malmqvist, P.; Roos, B. O. Density Matrix Averaged Atomic Natural Orbital (ANO) Basis Sets for Correlated Molecular Wave Functions. *Theor. Chim. Acta.* **1990**, *77*, 291–306.
- (35) Aidas, K.; Møgelhøj, A.; Nilsson, E. J. K.; Johnson, M. S.; Mikkelsen, K. V.; Christiansen, O.; Söderhjelm, P.; Kongsted, J. On the performance of quantum chemical methods

- to predict solvatochromic effects: The case of acrolein in aqueous solution. *The Journal of Chemical Physics* **2008**, *128*, 194503.
- (36) Jensen, L.; van Duijnen, P. T.; Snijders, J. G. A discrete solvent reaction field model within density functional theory. *The Journal of Chemical Physics* **2003**, *118*, 514.
- (37) Sylvester-Hvid, K. O.; Mikkelsen, K. V.; Nymand, T. M.; Åstrand, P.-O. Refractive Index of Liquid Water in Different Solvent Models. *The Journal of Physical Chemistry A* **2005**, *109*, 905–914.
- (38) Yu, H.; van Gunsteren, W. F. Accounting for polarization in molecular simulation. *Computer Physics Communications* **2005**, *172*, 69–85.
- (39) Rinkevicius, Z.; Murugan, N. A.; Kongsted, J.; Aidas, K.; Steindal, A. H.; Ågren, H. Density Functional Theory/Molecular Mechanics Approach for Electronic g -Tensors of Solvated Molecules. *The Journal of Physical Chemistry B* **2011**, *115*, 4350–4358.