Solvent Effects on Electronic Excitations of an Organic Chromophore

T. J. Zuehlsdorff,[*,†] P. D. Haynes,[‡,¶] F. Hanke,[§] M. C. Payne,[†] and N. D. M. Hine[‖]

[†]Cavendish Laboratory, University of Cambridge, J. J. Thomson Avenue, Cambridge CB3 0HE, U.K.

[‡]Department of Physics and [¶]Department of Materials, Imperial College London, Exhibition Road, London SW7 2AZ, U.K.

[§]Dassault Systèmes BIOVIA, 334 Science Park, Cambridge CB4 0WN, U.K.

[‖]Department of Physics, University of Warwick, Coventry CV4 7AL, U.K.

S *Supporting Information*

ABSTRACT: In this work we study the solvatochromic shift of a selected low-energy excited state of alizarin in water by using a linear-scaling implementation of large-scale time-dependent density functional theory (TDDFT). While alizarin, a small organic dye, is chosen as a simple example of solute−solvent interactions, the findings presented here have wider ramifications for the realistic modeling of dyes, paints, and pigment−protein complexes. We find that about 380 molecules of explicit water need to be considered in order to yield an accurate representation of the solute−solvent interaction and a reliable solvatochromic shift. By using a novel method of constraining the TDDFT excitation vector, we confirm that the origin of the slow convergence of the solvatochromic shift with system size is due to two different effects. The first factor is a strong redshift of the excitation due to an explicit delocalization of a small fraction of the electron and the hole from the alizarin onto the water, which is mainly confined to within a distance of 7 Å from the alizarin molecule. The second factor can be identified as long-range electrostatic influences of water molecules beyond the 7 Å region on the ground-state properties of alizarin. We also show that these electrostatic influences are not well reproduced by a QM/MM model, suggesting that full QM studies of relatively large systems may be necessary in order to obtain reliable results.
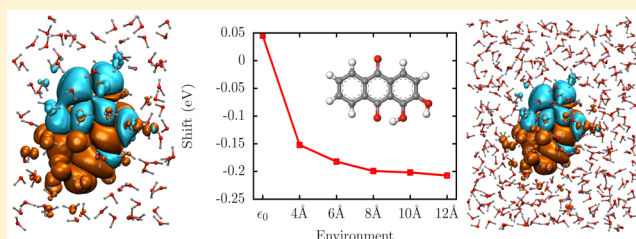
1. INTRODUCTION

Understanding the interactions between chromophores and their environment is one of the important objectives of the computational study of dyes in solvents[1−7] or pigment−protein complexes in biological systems.[8−15] These systems are generally characterized by a complex interplay between the chromophore and its surroundings, leading to spectral properties that deviate in nontrivial ways from gas-phase results. Most prominently, many chromophores experience a strong solvatochromic shift in solution, which is defined as the energy difference between an excitation in solution and the same excitation in the gas phase. Accounting for environmental effects in a correct and consistent way is crucial when trying to determine the color of dyes in different solutions.[2,4] On the other hand, the site energies of pigments in pigment−protein complexes such as the Fenna-Matthews-Olson complex[11,12] are found to be finely tuned by their environment in order to achieve very high quantum efficiencies in funneling excitons through the system. Any theoretical study of these types of systems necessarily relies on the correct treatment of environmental effects on excitation energies of pigments.

In this work, we consider an explicit example of a pigment-solvent system, namely alizarin in water, which has been the focus of a number of previous studies.[3,7] Alizarin is chosen because it is a small organic dye that does not deprotonate and has a single electronic ground state but nevertheless forms hydrogen bonds such that environmental effects are important. It can therefore be seen as an ideal simple model system for studying environmental effects on electronic excitations, while the insights gained can then be applied to more complicated examples of environmental effects, such as pigment−protein complexes.

Most recent works targeting the properties of pigments in solution use time-dependent density functional theory (TDDFT)[16−18] as the method of choice, due to its relatively low computational cost compared to higher order quantum mechanical methods. Computational approaches to represent the environment of a pigment of interest however show a considerable variety, ranging from classical continuum descriptions of solvent molecules through polarizable continuum models (PCMs)[4,7,19] to mixed quantum mechanical-classical (QM/MM) models, both with standard[9,15,20] and polarizable force fields[10,21−23] or a fully quantum mechanical treatment of the pigment and the environment.[2,3,24] QM/MM methods are often taken to provide a good balance between an accurate treatment of the environment and computational costs, but results obtained can be very sensitive to the size of the QM region.[9,10,15] Meanwhile, with traditional approaches,

quantum mechanical calculations are very computationally expensive and thus often rely on relatively small system sizes.[2,3]

Here, we focus on studying the sensitivity of calculated solvatochromic shifts with respect to the volume of surrounding environment explicitly included in the TDDFT calculation. While it has been discovered in previous studies that properties such as solvatochromic shifts can show a slow convergence with the size of the QM region,[6,15] to the best of our knowledge, no study has been performed in a simple test system that allows for a detailed breakdown and analysis of both the origin of the shift and the need for large system sizes. In this work, such an analysis is provided by studying the lowest dipole-allowed transition of alizarin in water, a transition of mainly HOMO→ LUMO character. The large-scale TDDFT calculations at the center of this work are enabled by an efficient approach capable of computing excited states in systems containing thousands of atoms,[25,26] while the analysis of solvatochromic shifts is facilitated by an innovative method that allows for the computation of excited states constrained to be confined within certain regions of larger systems. Furthermore, it will be demonstrated that this approach solves certain problems reported in previous studies regarding the computation of excited states in extended systems using (semi)-local exchange-correlation functionals that relate to the presence of unphysical charge-transfer excitations in the low-energy spectrum.[27]

2. COMPUTATIONAL METHODS

To analyze the solvent effects, we need to create a realistic model of alizarin and its local water environment. We are mainly interested in getting correct solvatochromic shifts, which have previously been shown to be sensitive to long-range interactions requiring large system-sizes, and it is therefore impractical to generate structures using a fully *ab initio* MD scheme. We therefore perform a classical molecular dynamics simulation of alizarin in water using AMBER.[28] While the structures obtained this way are expected to differ from those that would be obtained with a full QM treatment of the system, they can nevertheless be considered to form a representative model to study the long-range effects of the solvent on the excitation energies of the solute. Alizarin in its ground-state DFT structure is solvated in a truncated octahedral box containing approximately 6000 water molecules described by the TIP3P model.[29] The resulting system is heated up from 0 to 300 K for 40 ps using the Langevin thermostat with a collision frequency of 1 ps^{-1}, followed by a 200 ps volume equilibration at 1 atm of pressure in the NPT ensemble. We then perform a 200 ps production run at constant temperature and pressure. A time step of 2 fs is used throughout. Five frames are taken from the last 100 ps of the MD trajectory with a spacing of 25 ps, and these are used as input geometries for TDDFT calculations. We note that the purpose of this work is to determine the sensitivity of the solvatochromic shift to the treatment of the environment and not to provide a thermally averaged result for the energy of the low-energy dipole-allowed state of interest. Therefore, five evenly spaced MD frames can be considered to form a sufficient sample to span the range of magnitudes of shifts.

The dimensions of the alizarin molecule allow us to define a rectangular box via the more rigid carbon atoms of the structure, yielding a size of approximately 3.5 × 8.6 Å with a very small thickness, which contains all the carbon atoms of alizarin. To create different representations of the solvent environment, we increase the size of the box in all directions by

4 Å, 6 Å, 8 Å, 10 Å, and 12 Å, including all water molecules fully within the extended rectangular boxes into the ground-state DFT and excited-state TDDFT calculations. We thus generate five models of different amounts of explicit solvent environment for each of the five MD frames. The five solvent models vary slightly in the number of atoms for each of the five frames, but the 4 Å systems contain roughly 120 atoms (30 water molecules), while the 12 Å systems contain on the order of 1900 atoms (630 water molecules). For a detailed breakdown of how many water molecules are explicitly included in the different models for each frame, see Table 1. Note that for each model, the water molecules specified in Table 1 are treated fully quantum mechanically, both in the ground-state DFT and the TDDFT calculations.

Table 1. Number of Water Molecules Explicitly Included in Each Model for Each Frame[a]

	frame 1	frame 2	frame 3	frame 4	frame 5
4 Å model	33	28	30	28	36
6 Å model	92	91	95	88	103
8 Å model	193	208	204	202	218
10 Å model	337	371	390	383	387
12 Å model	636	630	626	628	633

[a]These water molecules are treated fully quantum mechanically both in the ground-state DFT calculations and in the TDDFT calculations.

The DFT and TDDFT calculations in this work are performed with the linear-scaling code ONETEP,[30,31] which is capable of computing optical excitations in systems containing thousands of atoms,[25] using the PBE functional.[32] ONETEP makes use of localized atom-centered functions to represent the single-particle density matrix. These atom-centered support functions are expanded in an underlying, systematically improvable basis equivalent to plane waves.[33] The support functions are optimized *in situ* during the calculation, which means that only a minimal number is needed to provide an ideal representation of the density matrix. For the calculations performed in this work, we choose an energy cutoff of 800 eV and a localization radius of 5.29 Å for all support functions (see the Supporting Information for further comments on the ONETEP calculation parameters and a demonstration of the insensitivity of the TDDFT results in this work with respect to an increase of the support function localization radius). We have demonstrated in previous works[25,26] that the optimized localized representation used in ONETEP yields TDDFT results of a quality comparable to that of large diffuse Gaussian basis sets like aug-cc-pVTZ.[34]

For the TDDFT calculations we first take the atomic positions of alizarin from each of the snapshots and calculate the energy of the first dipole-allowed state, a transition of mainly HOMO→ LUMO character, in vacuum. Depending on the chosen MD frame, this state is found to be one of the lowest three excitations of the system in vacuum. We then calculate the excitation energies for the same state in each of the explicit solvent models for all snapshots and compare these energies to the vacuum energies to determine the effective solvatochromic shifts. All calculations containing explicit solvent molecules are performed by embedding the system in an implicit solvent model,[35] which uses a static dielectric constant of $\epsilon_0 = 80$ in the region outside the water cluster to represent the continuum. The solvation cavity for the implicit solvent calculations is defined as an isosurface of the ground-

state density of the system in vacuum, and the dielectric constant becomes a function of the density that varies smoothly from $\epsilon_0 = 1$ for regions with high density to $\epsilon_0 = 80$ for regions low density. All parameters of the implicit solvent model are taken from Scherlis et al.[36] The use of an implicit solvent model has two desirable effects: First, the implicit solvent model prevents all electrostatic interactions between periodic images of the system that would otherwise be present in periodic boundary conditions.[37] Furthermore, it screens any unphysical dipoles on the surface of the solvent region explicitly included in the calculation, preventing a spurious closure of the band gap[38] and greatly reducing spurious charge-transfer states between "edge waters" on the surface of the solvent region and the pigment as observed in the literature.[27]

We also perform two further calculations on each snapshot: one where no explicit solvent is considered but in which the ground state of the alizarin molecule is computed within an implicit solvent with $\epsilon_0 = 80$; and another where we additionally include dynamic solvent effects into the excited-state calculations using a dynamic dielectric function of $\epsilon_\infty = 1.78$. In the first case, the solvent effects due to nuclear rearrangements of solvent molecules on the ground state of alizarin are fully considered, but any solvent reaction to the excitation itself is ignored. In the second case, the excitation is treated to occur on a time scale too fast for nuclear solvent degrees of freedom to react to it, but solvent polarization effects as a reaction to the excitation are modeled via the dynamic dielectric constant. It should be stressed that all calculations containing explicit water molecules are modeled using a static dielectric constant $\epsilon_0 = 80$ in the region beyond the QM region only, as dynamic polarization effects are taken to enter into the calculation through the explicit representation of water molecules rather than the averaged dielectric medium. All TDDFT calculations are performed in the Tamm-Dancoff approximation (TDA).[39] However, using a recent implementation of full TDDFT in ONETEP,[26] we have confirmed that the discrepancies between shifts predicted by the TDA and full TDDFT for these models is on the order of 10 meV or less.

3. TRUNCATION OF THE RESPONSE DENSITY MATRIX

In the ONETEP code, the TDDFT eigenvector associated with a given excitation is expressed via a transition density matrix $\mathbf{P}^{\{1\}}$[25,26] that is represented by two sets of *in situ* optimized localized atom-centered orbitals $\{\chi_\alpha\}$ and $\{\phi_\beta\}$, where $\{\chi_\alpha\}$ ideally spans the low-energy conduction manifold[40] and $\{\phi_\alpha\}$ ideally spans the valence manifold. This representation comes with a number of advantages, one being its compact size that allows for TDDFT calculations of very large system sizes,[25,26,41] while another is outlined in more detail in this section.

It can be demonstrated that atom-centered representation of TDDFT eigenvectors through $\mathbf{P}^{\{1\}}$ allows one to exclude certain characters of transitions from a TDDFT calculation. This is achieved by setting the corresponding density matrix elements of $\mathbf{P}^{\{1\}}$ to zero.[25,41] As an example, consider the system of alizarin in explicit water as described above. Matrix elements $P^{\{1\}\alpha\beta}$, where χ_α or ϕ_β are centered on an atom belonging to a water molecule can be set to zero. This would force the excitation to be strictly localized on the alizarin (see Figure 1a). Alternatively if nonzero matrix elements are allowed in case that χ_α and ϕ_β are both centered on the alizarin or both centered on the water, the local excitations of the alizarin are allowed to couple to local excitations in the water, but any
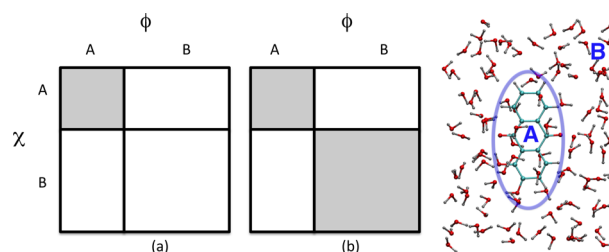

Figure 1. Schematic representation of two different truncations of the response density matrix $\mathbf{P}^{\{1\}}$. Gray areas denote dense matrix blocks, while white areas denote block matrix elements set to zero. Truncation scheme (a) only allows local excitations in region A of the system. Truncation scheme (b) allows local excitations in region B to couple with local excitations in region A, but all charge transfer excitations between the regions are suppressed.

charge-transfer excitations between the two regions are suppressed (see Figure 1b).

Applying this type of density matrix truncation directly to $\mathbf{P}^{\{1\}}$ would yield invalid results as every response density matrix has to follow an implicit invariance constraint of the form[25,41]

$$\mathbf{P}^{\{1\}'} = \mathbf{P}^{\{c\}}\mathbf{S}^\chi \mathbf{P}^{\{1\}}\mathbf{S}^\phi \mathbf{P}^{\{v\}} = \mathbf{P}^{\{1\}} \qquad (1)$$

where $\mathbf{P}^{\{v\}}$ and $\mathbf{P}^{\{c\}}$ are the ground-state valence and conduction space density matrices, and $\mathbf{S}^\phi$ and $\mathbf{S}^\chi$ are the overlap matrices associated with $\{\phi_\alpha\}$ and $\{\chi_\beta\}$. The invariance constraint is a direct consequence of the fact that $\{\chi_\beta\}$ spanning the conduction manifold is not orthogonal to $\{\phi_\alpha\}$ spanning the valence manifold. In order to ensure that $\mathbf{P}^{\{1\}}$ obeys the invariance constraint, we introduce the auxiliary density matrix $\mathbf{L}^{\{1\}}$, where $\mathbf{L}^{\{1\}}$ can have the sparsity pattern which appropriately constrains the excitation. $\mathbf{L}^{\{1\}}$ is then used to define the correct response density matrix via

$$\mathbf{P}^{\{1\}} = \mathbf{P}^{\{c\}}\mathbf{S}^\chi \mathbf{L}^{\{1\}}\mathbf{S}^\phi \mathbf{P}^{\{v\}} \qquad (2)$$

Since $\mathbf{P}^{\{c\}}$ and $\mathbf{P}^{\{v\}}$ are idempotent, $\mathbf{P}^{\{1\}}$ now obeys the constraint of eq 1 by construction. We stress that in defining $\mathbf{P}^{\{1\}}$ through the auxiliary density matrix $\mathbf{L}^{\{1\}}$, the lowest excitations of the system for a given truncation scheme can be obtained fully variationally.[41]

We demonstrate in the next section how different constraints placed on $\mathbf{L}^{\{1\}}$ allow us to analyze the origins of solvatochromic shifts of alizarin placed in explicit water.

4. RESULTS AND DISCUSSION

In this section we perform a detailed analysis of the origins of solvatochromic shifts for alizarin in water. The section is organized in three parts. In the first part, the convergence of the solvatochromic shift with the amount of explicit solvent included in the calculation is discussed. The second part contains a detailed breakdown of the origin of the solvatochromic shift by making use of the techniques introduced in Section 3. Finally, we discuss the performance of QM/MM calculations, where some parts of the explicit solvent are replaced by classical charges.

4.1. Convergence of Solvatochromic Shifts with Explicit Solvent. For illustrative purposes, Figure 2 shows plots of the electron−hole difference density of the lowest dipole-allowed transition of alizarin for one MD frame for three different sizes of solvent representation. As can be clearly seen, the electron−hole densities in the implicit solvent, the 6 Å and the 10 Å explicit models all describe the same excited state
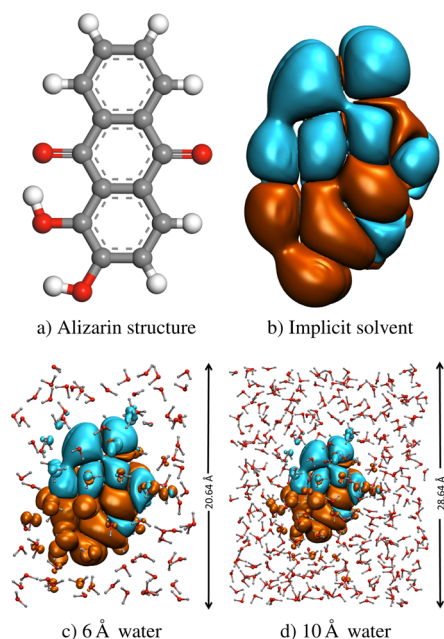
a) Alizarin structure   b) Implicit solvent

c) 6 Å water   d) 10 Å water

Figure 2. Structure of alizarin (a), as well the electron–hole difference density plot of the lowest dipole-allowed transition of a single MD frame (frame 5) of alizarin, as calculated in implicit solvent (ϵ_0 only, b), in 6 Å of explicit water (c), and in 10 Å of explicit water (d). The hole density is displayed in orange, while the electron density is shown in cyan. The state retains a localized character, with small amounts of electron and hole delocalization onto the water. Plots of the electron–hole difference density were created using VMD.[42]

localized on the alizarin. However, it is also evident that for the models containing an explicit representation of the solvent molecules, fractions of the electron and the hole delocalize onto the solvent molecules, with most of the delocalization confined to a relatively small region around the alizarin. Characterizing and quantifying this delocalization is the main focus of the present work, and a detailed analysis will be provided in the following section. Here we just note that there is a clear qualitative difference between the excited state of the system within an implicit solvent model and the same state within an explicit representation of solvent molecules.

The influence of different volumes of explicit solvent environment on the excitation can be seen in Figure 3, where the solvatochromic shift of the lowest dipole-allowed transition with respect to to the vacuum value is plotted for all five MD frames and different solvent models. While the implicit solvent model with static dielectric constant $\epsilon_0 = 80$ shows a positive solvatochromic shift of less than 0.05 eV for all snapshots, the magnitude of the shift is decreased once the dynamic screening ϵ_∞ of the excitation is considered. However, in explicit solvent models, the picture completely changes, with some MD frames (frame 3 and frame 5) showing a large solvatochromic shift of more than −0.2 eV, while frame 1 produces almost no solvatochromic shift at all. Furthermore, note that the shifts for some of the MD frames (frames 3 and 5) show a very slow convergence with the volume of explicit water included in the calculation, with a good level of convergence only reached for the 10 Å system of explicit water, at which point the TDDFT calculations contain around 380 water molecules. The only outlier in the convergence behavior is frame 4, where the solvatochromic shift increases significantly when going from 10 Å to 12 Å. It is worth pointing out that the energy scale of
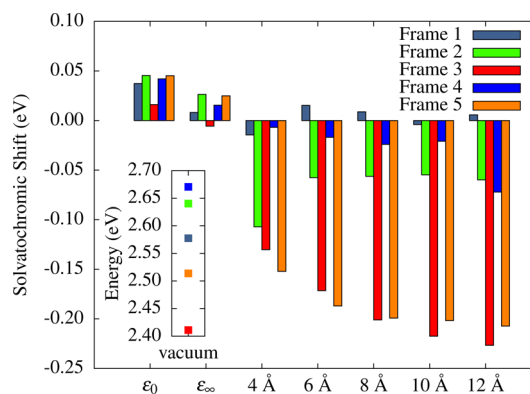

Figure 3. Solvatochromic shifts of the lowest dipole-allowed transition for 5 different MD frames of alizarin in different solvent representation models. The inserted figure shows the energy of the selected state in vacuum. With the exception of frame 4, which shows an anomalous behavior moving from the 10 Å to the 12 Å water model, all other shifts are converged at system sizes of 10 Å.

Figure 3 is larger than one color division on the rainbow for visible light, such that the observed sensitivity of the solvatochromic shift to the volume of explicit water matters greatly when considering the prediction of colors of dyes in solution.

The wide range of solvatochromic shifts for different snapshots clearly shows the importance of accounting explicitly for the local solvent environment. Furthermore, the implicit solvent models predict a small blueshift for almost all frames, while most of the converged transitions in explicit solvent are strongly red-shifted. The results obtained suggest that a potential origin of the red-shift can be ascribed to a partial delocalization of the excitation onto the water (see Figure 2) that helps screen the dipole moment of the excited state and thus lowers the excitation energy in a way that cannot be easily accounted for in an implicit solvation approach. We note that a similar analysis to the solvatochromic shift in Figure 3 can also be performed for the oscillator strength of the excitation and is provided in the Supporting Information. It is generally found that accounting for polarization effects through ϵ_∞ leads to an increase in oscillator strengths, while accounting for the environment quantum mechanically and letting the excitation delocalize causes the oscillator strengths to decrease from the ϵ_∞ values. This finding again reveals that the implicit solvent model accounts for polarization effects in a fundamentally different way from the explicit water models, as an external dielectric screening of the response density allows for a higher dipole moment to be supported on the alizarin itself, which tends to increase the oscillator strength of the transition. On the other hand, if the excitation is allowed to delocalize onto neighboring atoms, this delocalization helps to reduce the net dipole moment, which tends to lower the oscillator strengths of the excitation.

All calculations to this point have been performed using the PBE semilocal exchange-correlation functional.[32] In order to ensure that the strong solvatochromic shifts seen in explicit solvent are not a result of a spurious delocalization of the excitation due to the local nature of the functional, we also perform calculations using the CAM-B3LYP range-separated hybrid functional[43] implemented in NWChem[44] on the vacuum, implicit solvent (ϵ_0),[45] and 4 Å system. To account for the long-range solvent effects, the 4 Å system is also

simulated within an implicit solvent model. All NWChem calculations are performed using a 6-311G*[46] Gaussian basis set. The CAM-B3LYP functional is known to yield a much better description of long-range and charge-transfer excitations, as the long-range Hartree−Fock exchange interaction of the functional does not vanish if the electron and the hole of the TDDFT excitation show little overlap. A further reason why the CAM-B3LYP functional might be expected to provide a more realistic picture of the screening of the solvent environment is that the band gap of water is found to be considerably closer to the experimental value than that obtained with PBE, suggesting that charge-transfer delocalizations of the excitation onto the water have the correct energy cost associated with them. The results of the calculations in comparison to the PBE results are detailed in Figure 4. While we do not expect the absolute values
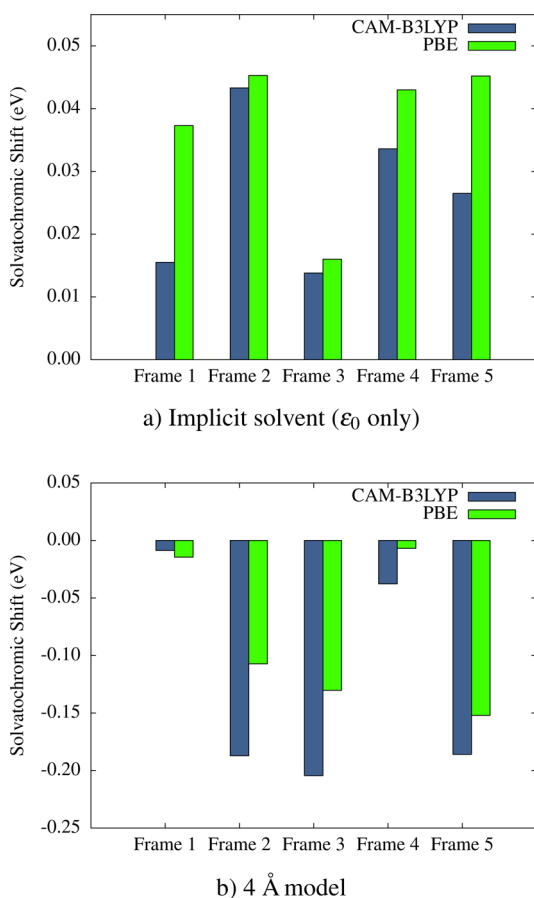

a) Implicit solvent (ε_0 only)

b) 4 Å model

Figure 4. Solvatochromic shifts of the lowest dipole-allowed transition in implicit solvent and the 4 Å model for 5 different MD frames, as calculated using the PBE functional and optimized support functions in ONETEP and the CAM-B3LYP functional and a 6-311G* basis set in NWChem. Both sets of calculations are performed in the Tamm-Dancoff approximation. A good qualitative agreement between the CAM-B3LYP and the PBE results can be observed.

of solvatochromic shifts to be in close agreement due to the above-mentioned differences in functionals, we nevertheless find that Figure 4 shows a correlation between CAM-B3LYP and PBE results. It is noteworthy that all frames produce a negative solvatochromic shift for the 4 Å system with the CAM-B3LYP functional and a positive shift for the implicit solvent model, just as the PBE functional does. Furthermore, the range-separated hybrid functional produces a strong redshift in the 4

Å model for the same frames (frames 2, 3, and 5) for which PBE also produces a strong redshift, while a small shift is produced for the other two frames. While discrepancies between the two methods are sizable (with RMS discrepancies of 53 meV in the 4 Å model) the correlation of trends regarding the observed solvatochromic redshift suggests that a similar delocalization effect as in the PBE functional occurs with the CAM-B3LYP, as the electrostatic effects of the environment are treated in a similar way in both functionals. The CAM-B3LYP results thus suggest that a partial delocalization of the excitation, as observed with the PBE functional, is physical and not a spurious overdelocalization caused by an underestimation of the energy of charge-transfer states.

4.2. Truncated TDDFT. We now shift the focus on to MD frames 1, 3, and 5, for a more detailed analysis of the origin of the solvatochromic shifts using the techniques introduced in Section 3. Frames 1, 3, and 5 are chosen due to the very different behavior of the solvatochromic shift in those frames, with frame 1 showing almost no shift and frames 3 and 5 showing a strong redshift of the excitation energy with increasing amounts of explicit solvent. While frames 3 and 5 show a similar behavior, the shift in frame 3 seems to converge somewhat slower with system size, making these frames interesting to study in comparison to each other.

In order to analyze the origin of the solvatochromic shifts in detail for these three frames, we perform a calculation where the excitation is localized on the alizarin and can couple with excited states localized on the surrounding water, but charge-transfer character transitions between the water and the alizarin are suppressed. We also perform two further calculations where the excitation is allowed to spread into the 4 Å and 7 Å box of explicit water, but charge-transfer character excitations between the 4 Å and 7 Å box and the water beyond are suppressed. The truncation schemes used here correspond to the scenario (b) in Figure 1. However, we have repeated the same calculations using truncation scenario (a) corresponding to excitations fully localized in the region around the alizarin and have obtained results within a few meV to the results obtained when local excitations in the water are considered. This demonstrates that the delocalization of excitations of alizarin is mainly charge-transfer in character and not due to a coupling of localized excitations on the alizarin to excitations within the water. The results of the three different truncation regions for frames 1, 3, and 5 can be found in Figure 5.

As can be seen, the three snapshots show very different behavior when constraining the excitation to be localized entirely on the alizarin, with frame 1 showing a strong blueshift of 0.167 eV and frames 3 and 5 a moderate redshift of −0.076 eV and −0.077 eV, respectively. However, while the final shift for the constrained excitation is similar between frames 3 and 5, the convergence of the shift with the volume of explicit water is considerably slower for frame 3. Since the excitation stays fully localized on the alizarin in these calculations, the obtained shift can be considered to be due to purely electrostatic effects of the surrounding solvent environment on the Kohn−Sham states associated with alizarin. The effect of allowing the excitation to delocalize onto the surrounding water is a strong red-shift in all three frames, confirming the origin of the solvatochromic shift suggested in Section 4.1. Note that while for frame 5, results to within 10 meV of the unconstrained results can be obtained when constraining the excitation to within 4 Å for the other frames it is necessary to let the excitation delocalize within the 7 Å box to recover all of the solvatochromic shift predicted from
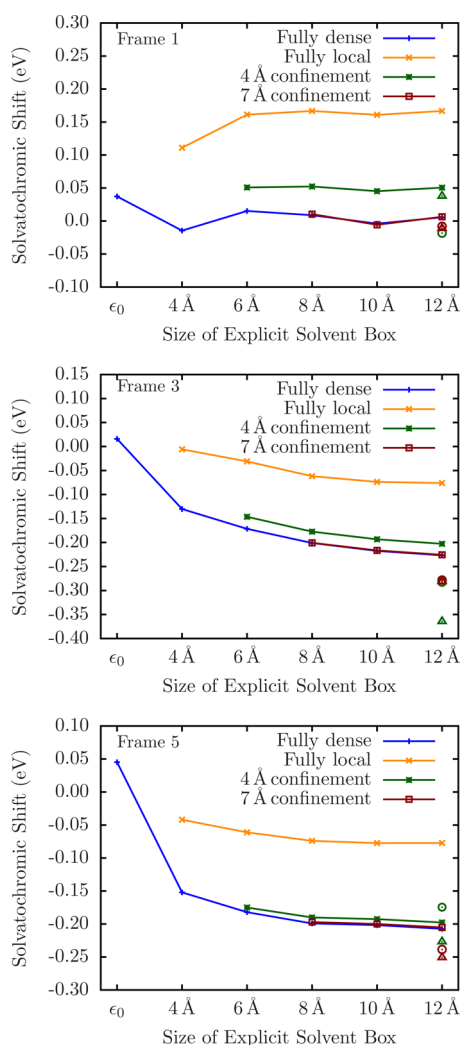
Figure 5. Solvatochromic shifts of the lowest dipole-allowed state for frames 1 (upper figure), 3 (middle figure), and 5 (lower figure). The colors refer to different degrees of truncation of the response density kernel describing the excitation. A truncation of the response density matrix to 7 Å is found to fully recover the solvatochromic shift obtained from fully dense response density matrices in all frames. The open circle denotes QM/MM results where atoms beyond 4 Å (green) and 7 Å (red) are replaced by classical charges from the TIP3P model. The open triangle denotes QM/MM results where the charges used in the MM region are derived from a Mulliken analysis of the DFT ground-state calculation.

the completely delocalized calculation. However, in all three frames the excitation retains a relatively localized character. It can be summarized that the origin of the solvatochromic redshift is a partial delocalization of the excitation that is charge-transfer in character but is confined to within a region of about 7 Å from the alizarin.

The second component to the solvatochromic shift is the purely electrostatic component due to the specific configuration of the surrounding water. This electrostatic contribution can cause an additional redshift (frames 3 and 5) or a blueshift counteracting the redshift of the delocalization of the excitation (frame 1). Furthermore, this electrostatic contribution to the solvatochromic shift can show a very slow convergence with system size (frame 3), requiring considerably larger volumes of explicit water than the 7 Å volume into which the excitation delocalizes. While the origin of the slow convergence of the

shift in frame 3 as compared to other frames is difficult to determine, we point out that frame 3 is the only frame in which a water molecule forms a hydrogen bond with one of the oxygens of the alizarin itself (see the Supporting Information). It is plausible that a larger amount of explicit solvent environment is needed to correctly converge the electrostatic influences on the water molecule and its effect on the electronic structure of the alizarin molecule.

In summary, we note that the explicit charge-transfer character delocalization of the excitation is a large contributor to the total solvatochromic shift in all frames studied and is a purely quantum mechanical effect that is likely difficult to capture with any polarizable continuum or classical force-field model in that region. It can therefore be concluded that the minimum size of a QM region for alizarin in water has to be chosen at around 7 Å ($\approx$150 molecules of water) in order to obtain consistent results for the solvatochromic redshift due to delocalization. This result has wider ramifications for the modeling of bright transitions in other pigment-solvent and pigment–protein complexes, where it is likely that similar redshifts can occur if the system response and thus the excitation energy can be lowered by a partial delocalization of the excited state into its surrounding environment.

While the truncation of the response density matrix is only introduced in this work to analyze the origin of the solvatochromic shift, it comes with a number of desirable consequences from a computational point of view. First, a sparsity in $\mathbf{P}^{\{1\}}$ allows for fully linear-scaling LR-TDDFT calculations[25] with a potential to significantly speed up the largest TDDFT calculations performed in this work. Furthermore, the truncation removes all spurious charge-transfer states between the alizarin and "edge waters" from the subspace of allowed solutions, causing physical excitations localized on the alizarin to rigorously become the lowest excitations of the system for all models and frames considered. The truncation is, thus, an effective technique to avoid the convergence problems observed in the literature[27] even for local exchange-correlation functionals, greatly reducing the computational cost by reducing the number of states that need to be converged.

The localized character of the excitation can be quantified by integrating over the part of the electron and the hole density that is described by support functions in the basis set localized on atoms belonging to alizarin. The resulting value can be interpreted as the percentage of the electron and the hole that is localized on the alizarin. It is found that only about 2% of the electron and 5% of the hole delocalize onto the water (see Figure 6). Interestingly, the 12 Å system actually shows *less* delocalization than the 6 Å system. This suggests that if the volume of solvent environment is chosen large enough to obtain a converged solvatochromic shift, the redshift due to explicit delocalization of the excitation can be achieved with a minimal amount of electron–hole delocalization. When the volume of solvent environment is chosen too small, the constraint on the excitation is compensated by a slight overdelocalization to nearby waters. Also, it is worth noting that the converged degree of electron–hole delocalization is very similar for both snapshots, even though the lowest dipole-allowed state in frame 3 shows a considerably larger redshift than the corresponding state in frame 4. This again suggests that the amount of solvatochromic redshift due to explicit delocalization is similar for each frame, while the large differences in solvatochromic shifts between different frames
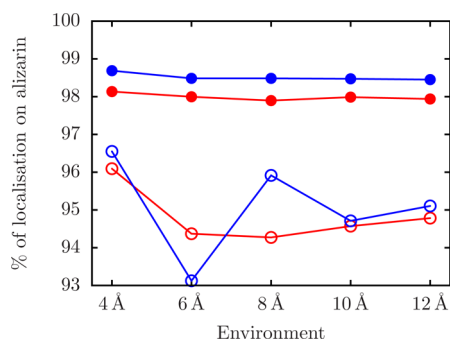
Figure 6. Plot comparing the percentage of electron (filled circle) and hole (empty circle) densities for MD frames 3 (red) and 4 (blue) that is confined to the alizarin only.

are due to electrostatic effects originating from the specific configurations of solvent molecules.

4.3. QM/MM. The relative localization of the excitation within 7 Å raises the question of whether it is possible to replace the water molecules beyond that region by classical charges in a QM/MM approach to reduce computational costs. To test for this we take the 12 Å model of frames 1, 3, and 5 and perform two sets of different calculations. In the first set we replace all atoms beyond 4 Å and 7 Å, respectively, with classical charges taken from the TIP3P model. In the second set of calculations the atoms beyond 4 Å and 7 Å, respectively, are replaced by classical charges obtained from a Mulliken analysis of the full ground-state DFT calculation in `ONETEP`. Both sets of results are compared to the results where the full 12 Å region is treated quantum mechanically but the excitation is constrained to within 4 Å and 7 Å. We note that the approach of simply replacing the waters beyond a certain region of the system by their classical TIP3P charge equivalents is a simple approach to a QM/MM calculation, as only the long-range Coulombic part of the potential is considered and the effective charges on each water molecule are fixed. More sophisticated approaches to QM/MM involve polarizable force fields[22,23] which are likely to yield a better description of the classical region. We therefore consider a second set of QM/MM results, where the representation of the classical region is based on Mulliken charges that are computed fully quantum mechanically. It should be noted there are more sophisticated methods available in obtaining net atomic charges from electron densities that overcome some of the well-known limitations of Mulliken charges[47] and might yield a better representation of the full DFT-derived electrostatic potential of the large 12 Å systems. However, since standard QM/MM approaches do not have access to the full DFT density of the system, the Mulliken approach chosen here can be treated as a good representation of an ideal QM/MM model obtainable with sophisticated polarizable force fields. Both sets of calculations taken together then reveal to which degree the long-range electrostatic influences on the excitation can be reproduced by classical charges when considering a relatively large QM region.

The results of the QM/MM analysis for three frames, using both the TIP3P charges and classical Mulliken charges for the classical regions, can be found in Figure 5. A detailed analysis is provided in the Supporting Information. Here we just note that the performance of both classical models is found to vary significantly for different frames. While the Mulliken charge model agrees well with the fully quantum mechanical representation for frame 1, it produces significant errors for

frame 3 and shows a mixed performance for frame 5. The TIP3P model generally performs slightly worse in predicting the correct shifts when going from a 4 Å QM region to a 7 Å QM region, and both models have the tendency to overestimate the total solvatochromic redshift, sometimes by a large amount (up to 0.16 eV for the 4 Å QM region and Mulliken charge model in frame 3).

Thus, while the water molecules beyond a 7 Å region evidently do not take part directly in the excitation, their quantum mechanical treatment is still important to obtain reliable results, and hence a QM/MM approach could produce significant errors. This is particularly evident when considering the mixed performance of the QM/MM model when using classical charges derived from a Mulliken analysis of the ground-state DFT calculation, as even sophisticated polarizable force fields[48] are unlikely to outperform DFT-derived charges. This suggests that the differences between the fully quantum mechanical and the classical treatment of the environment beyond the 7 Å region, either through classical charges or through continuum models, are due to a ground-state effect that is quantum mechanical in nature. It implies that the Kohn−Sham states associated with the water that are involved in the delocalization of the excitation are influenced by the long-range treatment of the solvent environment.

The slow convergence of the solvatochromic shift with respect to system size is thus due to an interplay of two different effects: a red-shift due to a charge-transfer delocalization of the excitation, mainly within a region of about 7 Å from the alizarin containing about 470 atoms, and electrostatic effects due to the specific configurations of the water molecules of a given snapshot, which can induce a red- or blueshift and are relatively long-ranged. We also note that the influence of solvent molecules beyond the 7 Å region cannot be reproduced to very high accuracy by a QM/MM approach for a number of frames, although the performance of a Mulliken classical charge model is found to be relatively good for one of the frames considered.

5. CONCLUSION

We have performed a detailed study of the origins of solvatochromic shifts of alizarin in water, with a special emphasis being placed on the effects of including large volumes of explicit solvent in the calculation. We have found that large system sizes, of up to 380 molecules of explicit water, can become necessary to obtain accurate results, providing a challenge for the study of organic dyes in solution. However, we have also demonstrated that the excitation itself retains a relatively localized character and can be well-represented by a truncated response density matrix, thus opening up the possibility of significantly speeding up TDDFT calculations in these systems by exploiting linear-scaling techniques. These techniques have the additional advantage of removing any spurious charge-transfer states from the solute to the edge of the solvent box from the calculation. Therefore, when using appropriate truncation techniques, (semi)-local functionals are found to be well-suited for the study of pigment-solvent systems, thus removing problems found in previous studies.[27]

The degree of delocalization of the excited states observed in the alizarin suggests that any approach which includes waters within a range of less than 7 Å from the solute in the TDDFT calculation must introduce significant errors by forcing the excitation to artificially localize. We showed that this delocalization is not simply an artifact of semilocal functionals

by repeating a number of tractable calculations with a range-separated hybrid functional. We stress that our test system forms a very simple example of pigment-solvent interactions and that larger pigments as well as more complicated pigment–protein complexes will most likely need considerably larger TDDFT calculations to allow for the correct delocalization of excited states. Thus, our findings are in agreement with previous studies suggesting the necessity of the use of large QM regions in these systems.[6,15] The issues raised in the present work, especially regarding the contribution to solvatochromic shifts due to delocalization effects that are quantum mechanical in nature, are therefore likely to be of relevance in the computational modeling of a wide variety of different systems, ranging from the prediction of colors of dyes in solutions to the study of large pigment–protein complexes in computational biology.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b01014.

> Pictures of the arrangement and hydrogen bonds in nearest neighbor water molecules for all frames, as well as an analysis regarding the behavior of the oscillator strength of the excitation with the treatment of the environment; convergence study of the solvatochromic shift with respect to the localization radius of the ONETEP support functions; detailed analysis on the performance of the QM/MM models with classical charges (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: tjz21@cam.ac.uk.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Marini, A.; Muñoz-Losa, A.; Biancardi, A.; Mennucci, B. J. Phys. Chem. B 2010, 114, 17128−17135.

(2) Malcioğlu, O. B.; Calzolari, A.; Gebauer, R.; Varsano, D.; Baroni, S. J. Am. Chem. Soc. 2011, 133, 15425−15433.

(3) Douma, D. H.; M'Passi-Mabiala, B.; Gebauer, R. J. Chem. Phys. 2012, 137, 154314.

(4) Manzhos, S.; Komatsu, M.; Nakazaki, J.; Segawa, H.; Yamashita, K. J. Photonics Energy 2012, 2, 028001.

(5) Wada, T.; Nakano, N.; Sato, H. J. Chem. Theory Comput. 2014, 10, 4535−4547.

(6) Eilmes, A. Theor. Chem. Acc. 2014, 133, 1538.

(7) Amat, A.; Miliani, C.; Romani, A.; Fantacci, S. Phys. Chem. Chem. Phys. 2015, 17, 6374.

(8) Filippi, C.; Zaccheddu, M.; Buda, F. J. Chem. Theory Comput. 2009, 5, 2074−2087.

(9) Filippi, C.; Buda, F.; Guidoni, L.; Sinicropi, A. J. Chem. Theory Comput. 2012, 8, 112−124.

(10) Daday, C.; Curutchet, C.; Sinicropi, A.; Mennucci, B.; Filippi, C. J. Chem. Theory Comput. 2015, 11, 4825−4839.

(11) König, C.; Neugebauer, J. J. Chem. Theory Comput. 2013, 9, 1808−1820.

(12) Cole, D. J.; Chin, A. W.; Hine, N. D. M.; Haynes, P. D.; Payne, M. C. J. Phys. Chem. Lett. 2013, 4, 4206.

(13) González, E. M.; Guidoni, L.; Molteni, C. Phys. Chem. Chem. Phys. 2009, 11, 4556−4563.

(14) Wang, Y.; Li, H. J. Chem. Phys. 2010, 133, 034108.

(15) Isborn, C. M.; Götz, A. W.; Clark, M. A.; Walker, R. C.; Martínez, T. J. J. Chem. Theory Comput. 2012, 8, 5092−5106.

(16) Runge, E.; Gross, E. K. U. Phys. Rev. Lett. 1984, 52, 997.

(17) Petersilka, M.; Gossmann, U. J.; Gross, E. K. U. Phys. Rev. Lett. 1996, 76, 1212.

(18) Casida, M. E. J. Mol. Struct.: THEOCHEM 2009, 914, 3−18.

(19) Tomasi, J.; Mennucci, B.; Cammi, R. Chem. Rev. 2005, 105, 2999−3094.

(20) Warshel, A.; Levitt, M. J. Mol. Biol. 1976, 103, 227−249.

(21) Aidas, K.; Møgelhøj, A.; Nilsson, E. J. K.; Johnson, M. S.; Mikkelsen, K. V.; Christiansen, O.; Söderhjelm, P.; Kongsted, J. J. Chem. Phys. 2008, 128, 194503.

(22) Olsen, J. M.; Aidas, K.; Mikkelsen, K. V.; Kongsted, J. J. Chem. Theory Comput. 2010, 6, 249−256.

(23) Olsen, J. M.; Aidas, K.; Kongsted, J. J. Chem. Theory Comput. 2010, 6, 3721−3734.

(24) Car, R.; Parrinello, M. Phys. Rev. Lett. 1985, 55, 2471−2474.

(25) Zuehlsdorff, T. J.; Hine, N. D. M.; Spencer, J. S.; Harrison, N. M.; Riley, D. J.; Haynes, P. D. J. Chem. Phys. 2013, 139, 064104.

(26) Zuehlsdorff, T. J.; Hine, N. D. M.; Payne, M. C.; Haynes, P. D. J. Chem. Phys. 2015, 143, 204107.

(27) Isborn, C. M.; Mar, B. D.; Curchod, B. F. E.; Tavernelli, I.; Martínez, T. J. J. Phys. Chem. B 2013, 117, 12189−12201.

(28) Case, D. A.; Berryman, T. J.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B. L.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. AMBER 2015; University of California: San Francisco, 2015.

(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J. Chem. Phys. 1983, 79, 926.

(30) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. J. Chem. Phys. 2005, 122, 084119.

(31) Haynes, P. D.; Skylaris, C.-K.; Mostofi, A. A.; Payne, M. C. Psi-k Newsletter 2005, 72, 78−91.

(32) Perdew, J. P.; Burke, K.; Ernzerhof, M. Phys. Rev. Lett. 1996, 77, 3865.

(33) Mostofi, A. A.; Skylaris, C.-K.; Haynes, P. D.; Payne, M. C. Comput. Phys. Commun. 2002, 147, 788−802.

(34) Dunning, T. H., Jr. J. Chem. Phys. 1989, 90, 1007.

(35) Dziedzic, J.; Helal, H. H.; Skylaris, C.-K.; Mostofi, A. A.; Payne, M. C. Europhys. Lett. 2011, 95, 43001.

(36) Scherlis, D. A.; Fattebert, J.-L.; Gygi, F.; Cococcioni, M.; Marzari, N. J. Chem. Phys. 2006, 124, 074103.

(37) Hine, N. D. M.; Dziedzic, J.; Haynes, P. D.; Skylaris, C.-K. J. Chem. Phys. 2011, 135, 204103.

(38) Lever, G.; Cole, D. J.; Hine, N. D. M.; Haynes, P. D.; Payne, M. J. Phys.: Condens. Matter 2013, 25, 152101.

(39) Hirata, S.; Head-Gordon, M. Chem. Phys. Lett. 1999, 314, 291−299.

(40) Ratcliff, L. E.; Hine, N. D. M.; Haynes, P. D. Phys. Rev. B: Condens. Matter Mater. Phys. 2011, 84, 165131.

(41) Zuehlsdorff, T. J. Computing the Optical Properties of Large Systems; Springer International Publishing: Switzerland, 2015; pp 97−165.

(42) Humphrey, W.; Dalke, A.; Schulten, D. *J. Mol. Graphics* **1996**, *14*, 33−38.

(43) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51−57.

(44) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; et al. *Comput. Phys. Commun.* **2010**, *181*, 1477.

(45) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799−805.

(46) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.

(47) Manz, T. A.; Sholl, D. S. *J. Chem. Theory Comput.* **2010**, *6*, 2455−2468.

(48) Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236−242.