JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

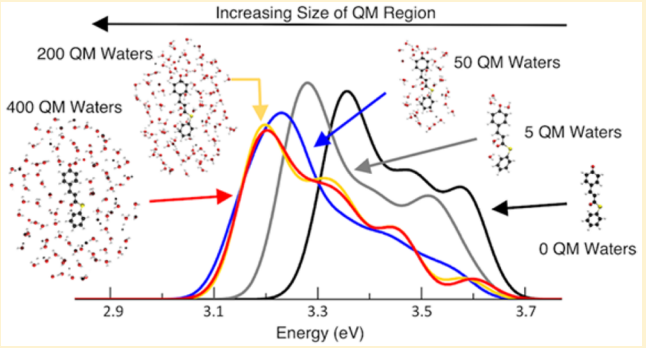
Convergence of Computed Aqueous Absorption Spectra with Explicit Quantum Mechanical Solvent

Joel M. Milanese, Makenzie R. Provorse, Enrique Alameda, Jr., and Christine M. Isborn*

Chemistry and Chemical Biology, University of California at Merced, Merced, California 95343, United States

Supporting Information

ABSTRACT: For reliable condensed phase simulations, an accurate model that includes both short- and long-range interactions is required. Short- and long-range interactions can be particularly strong in aqueous solution, where hydrogen-bonding may play a large role at short range and polarization may play a large role at long range. Although short-range solute−solvent interactions such as charge transfer, hydrogen bonding, and solute−solvent polarization can be taken into account with a quantum mechanical (QM) treatment of the solvent, it is unclear how much QM solvent is necessary to accurately model interactions with different solutes. In this work, we investigate the effect of explicit QM solvent on absorption spectra computed for a series of solutes with decreasing polarity. By adjusting the boundary between QM and classical molecular mechanical solvent to include up to 400 QM water molecules, convergence of the calculated absorption spectra with respect to the size of the QM region is achieved. We find that the rate of convergence does not correlate with solute polarity when excitation energies are calculated using time-dependent density functional theory with a range-separated hybrid functional, but it does correlate with solute polarity when using configuration interaction singles. We also find that larger basis sets converge the computed spectrum with fewer QM solvent molecules. To optimize the computational cost with respect to convergence, we test a mixed basis set with more basis functions for atoms of the chromophore and the solvent molecules that are nearest to it and fewer basis functions for the atoms of the remaining solvent molecules in the QM region. Our results show that using a mixed basis set is a potentially effective way to significantly lower the computational cost while reproducing the results computed with larger basis sets.



INTRODUCTION

Many chemical processes occur in the condensed phase. When simulating condensed phase chemistry, it is vital to accurately model both the solute and its environment. For example, including the correct hydrogen bonding and polarization environment can be key to obtaining accurate energies for ground and excited states. The accuracy of a given model may vary for different solutes or solvents. Quantum mechanics (QM) can accurately take into account electronic processes, such as the polarization and charge transfer that can occur in solution, but computational limitations restrict the number of atoms that can be treated quantum mechanically. A common approach to extending the size of the model system is to combine a QM method for modeling short-range interactions with a classical molecular mechanics (MM) method for modeling the long-range electrostatic interactions.[1] Both ground and excited state properties can be accurately modeled using QM/MM approaches.[2−9]

Recent computational advances have enabled the QM treatment of systems with thousands of atoms. For example, graphic processing unit (GPU) parallel architecture can greatly increase the computational efficiency[10] of quantum chemical calculations.[11−15] The electronic structure program TeraChem utilizes GPUs to increase the performance of ground and excited state calculations by 140−400× compared to that with traditional CPU architecture.[14,16] The linear-scaling electronic structure code ONETEP is optimized for parallel computation on CPUs, allowing large-scale calculations of solvation effects.[17−21] Additional linear-scaling electronic structure codes include CONQUEST,[22,23] SIESTA,[24,25] and OPENMX.[26] See ref 27 for a review of linear-scaling methods. Within the QM/MM approach, GPU-optimized and linear-scaling QM methods allow more explicit QM solvent molecules to be included in calculations than was previously practical.[9,28,29]

A key concern when using any QM/MM method is determining the boundary between the QM and MM regions.[9,29−34] Because of our interest in modeling excited states in the condensed phase, in this article we focus on convergence of excitation energies and spectral properties with respect to the size of the QM solvation region, with a focus on solutes of different polarity in water. Here, we review several convergence studies of aqueous solutes. Zuehlsdorff et al. found that the solvatochromic shift of aqueous alizarin (26 atoms)

Received: February 14, 2017
Published: March 31, 2017

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171

relative to the gas phase requires up to 380 QM water molecules for convergence; the relatively large number of explicit QM water molecules that must be included is attributed to the delocalization of the transition density onto water molecules within 7 Å of the solute (about 150 QM water molecules) and to long-range electrostatic interactions.[29] Murugan et al. reported that for aqueous phenol blue (31 atoms) three solvation shells (156 water molecules) needed to be included in the QM region to achieve convergence of the average excitation energy.[35] Using a frozen-density embedding method, Neugebauer et al. found that for a single solute−solvent configuration of aminocoumarin C151 (23 atoms) in aqueous solution at least 150 frozen-density QM water molecules (corresponding to a solvent shell radius of 7.8 Å) were necessary for convergence of the electronic excitation energy.[36] Our recent work suggests that anionic *trans*-thiophenyl-*p*-coumarate, the fluorescein anion, and the neutral fluorescein molecule require about 100 QM water molecules for convergence of the excitation energy, whereas neutral *trans*-thiophenyl-*p*-coumarin requires about 200 QM water molecules.[37] Flaig et al. found that the NMR ^{1}H and ^{13}C isotropic shielding constants of aminopyazole peptide (89 atoms) in aqueous solution converged with a spherical QM cutoff distance of 7 and 4 Å (corresponding to 192 and 50 QM water molecules), respectively.[38] The NMR ^{1}H and ^{13}C isotropic shielding constants of a molecular clip (88 atoms) in aqueous solution converged with a spherical QM cutoff distance of 4 Å (corresponding to 37 QM water molecules).[38]

Although these results suggest that a significant number of solvent molecules must be treated quantum mechanically to converge properties of molecules in solution, other studies find that these properties are independent of the QM region size. For example, Pavone et al. determined that the QM/MM electronic excitation energies and NMR ^{17}O isotropic shielding constants of acetone (10 atoms) in water were negligibly affected by QM treatment of up to 5 of the nearest water molecules.[39] The QM/MM excitation energy of pyridine (11 atoms) converged with just 1 QM water molecule.[40] We recently reported that phenol (13 atoms) and phenolate (12 atoms) require only 5 QM water molecules to achieve convergence of the excitation energy.[37] Ma and Ma computed the solvatochromic shift of aqueous acetone and aqueous benzene (12 atoms) as a function of the size of the QM region;[41] both solutes showed small changes in the solvatochromic shift as the QM region size increased from 0 to 36 QM water molecules, although benzene showed a larger change compared to acetone.

These studies suggest that the size of the solute affects the number of QM water molecules necessary to achieve convergence, but size alone does not explain all of the observed trends. For example, the anionic and neutral pCT molecules are of similar size, but the former requires only the first solvation shell to be treated with QM, whereas the latter requires nearly two solvation shells.[37] Additional properties may affect the convergence of solute−solvent interactions, such as solute polarity, electronic structure of the solute, and degree of hydrogen bonding. One aim of this work is to investigate the relationship between solute polarity and the convergence of excitation energy with respect to QM region size.

The degree of polarization and charge transfer calculated using different electronic structure methods can vary considerably. Therefore, the choice of QM method will likely affect the rate of convergence with respect to the amount of QM solvent. For density functional theory (DFT) methods, approximate density functionals can vary widely in their accuracy.[42−46] Recent studies from our group show that density functionals with small amounts of exact exchange produce a size-dependent error in the ionization potential;[47,48] when increasing the size of the QM solvation region, the ionization potential is increasingly underestimated.[48] Including exact exchange at long range via a long-range corrected (LRC) hybrid functional[49−51] improves the size-dependent error, and with enough exact exchange, it leads to the correct polarization response of the QM solvent.[48]

These studies address the importance of including exact exchange in ground state DFT calculations, but exact exchange is also important for obtaining physically correct excited state properties. Time-dependent density functional theory (TDDFT) is often the method of choice for computing the excitation energies of large chemical systems (on the order of 40 to a few hundred atoms) due to its moderate computational cost.[52] However, TDDFT can lead to unphysical results when modeling charge-transfer transitions because typical local and semilocal exchange functionals do not include a term to account for the long-range Coulombic interaction between the excited electron and hole in the TDDFT equations.[53−57] Thus, calculations with typical local and semilocal exchange functionals incorrectly predict that long-range charge-transfer transitions are lower in energy than localized, optically bright valence transitions. For QM calculations that include solvent, the TDDFT charge-transfer problem is particularly egregious because dark transitions that correspond to transitions between distant solvent molecules and the unoccupied levels of the solute can appear at low energies.[28,58−61] We have previously shown that many of these low-energy transitions result from the water environment being exposed to vacuum, resulting in water molecules that are in a position for being a hydrogen bond donor, but not an acceptor, and therefore have a partial negative charge.[61] When a system is carved out of a larger simulation box and then exposed to vacuum, a large dipole moment may exist due to the unscreened solvent surface exposed to the vacuum.[62] A simple way to keep the solvent from being exposed to vacuum is to surround the QM solvation sphere with a classical solvent model, such as a polarizable continuum model (PCM) or MM point charges (i.e., QM/MM method).[60,61,63] Another option is to rigorously separate the solvent from the solute in the TDDFT equations by partitioning the orbital spaces of the solvent and the solute to remove the possibility of solvent → solute charge-transfer excitations but to still include polarization in a self-consistent manner.[29,64] Because the correct asymptotic exchange behavior of LRC functionals yields the correct electron−hole interaction in the TDDFT equations,[49−51,65,66] LRC functionals often reduce the charge-transfer error in solution considerably.[60,61] An alternative excited state method that inherently models the correct asymptotic exchange behavior is configuration interaction singles (CIS), but it is known to err in the opposite direction of TDDFT: it has a systematic bias against charge-transfer transitions, predicting them to be too high in energy compared to valence transitions.[67] Because the CIS method lacks electron correlation, it generally underestimates polarizabilities[68] and overestimates excitation energies[63,69,70] relative to experiment. The role of exact exchange in the convergence of excitation energies is investigated in this article for solutes of varying polarity.

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171

For any QM method, the size of the basis set must be chosen to balance the computational cost and the accuracy of computed properties. Previous studies suggest that improving the basis set may change the computed values of properties but does not significantly change their convergence rates. For example, Neugebauer et al. reported that the addition of polarization functions changed the excitation energy but did not change the convergence of the solvatochromatic shift of aqueous aminocoumarin C151.[36] Also, the DZP and TZP basis sets gave the same electronic excitation energy of this system for all QM region sizes. The convergence of QM/MM excitation energies of pyridine in water reported by Nielsen et al. had negligible basis set effects when comparing the cc-pVDZ and aug-cc-pVDZ basis sets.[40] The computation of NMR ^{1}H and ^{13}C isotropic shielding constants of aminopyazole pepetide in aqueous solution using a large QM only model system had minimal basis set effects.[38] Our previous work suggests that adding polarization functions does not significantly change the excitation energy for a single solute−solvent configuration of aqueous pCT−, whereas adding a diffuse function decreases the excitation energy by 0.05 eV; the effect of the basis set on the convergence of the excitation energy was insignificant.[37] In this article, we go beyond previous studies of basis set effects of individual solute−solvent configurations, and in this work, we test how the choice of basis set affects the convergence of spectra computed from a larger number of solute−solvent configurations.

An alternative to increasing the size of the basis set of all QM molecules (which quickly becomes computationally intractable) is to use different basis sets to model different parts of the QM system. For example, Georg et al. used a larger basis set to model the water molecules near the solute compared to water molecules at the edge of the solvation sphere in a QM only cluster model of aqueous acrolein.[71] The authors noted that test calculations using a larger basis set for the outer water molecules changed the excitation energy but had negligible effects on the solvatochromic shift. To the best of our knowledge, convergence with respect to a mixed basis set cutoff within a QM/MM model has not been investigated.

In this article, we examine the following aspects of calculating condensed phase molecular properties using a QM/MM method: the size of the QM region, the QM method, and the basis set. Using the GPU-accelerated TeraChem program, the excitation energies are computed with up to 400 QM water molecules in the excited state calculations at the same level of QM theory as the solute. To investigate the effect of solute polarity on the convergence of computed excitation energies and absorption spectra, we consider four solutes of differing polarity (Figure 1): Solute 1, *trans*-thiophenyl-*p*-coumarate (IUPAC name: *S*-phenyl (*E*)-3-(4-hydroxyphenyl)prop-2-enethioate), has an experimental aqueous absorption spectrum λ_{max} of 3.14 eV;[72] this solute is an analogue of the anionic chromophore of photoactive yellow protein (PYP). In addition to rhodopsin, PYP is one of the most experimentally and theoretically studied photosensitive biological systems. The PYP chromophore has been studied extensively in vacuum and protein environments where the electronic excitations have proven challenging to model with TDDFT methods;[9,73−77] this is partially due to the excited state of the anionic chromophore lying above or near the first ionization potential.[78−82] However, TDDFT has been more successful at modeling the excitation energy of this chromophore in water,[9,83] which is likely a result of the change in electronic structure in an aqueous environment

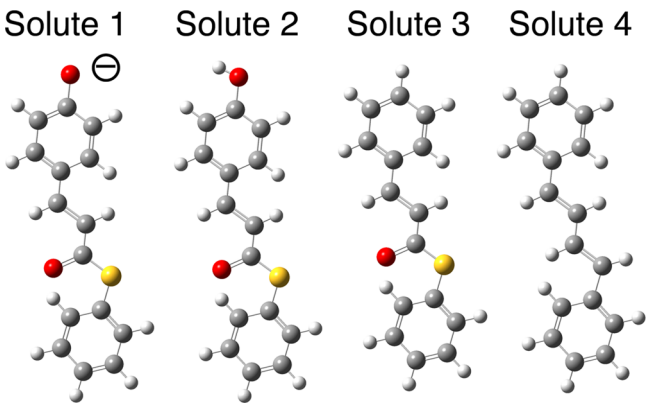


Figure 1. Ball and stick representations of the solutes examined in this study, which decrease in polarity going from Solute 1 to Solute 4. Carbon atoms are gray, oxygen atoms are red, and sulfur atoms are yellow.

where hydrogen bonds will stabilize the negative charge of the phenolate moiety. Solute 2 is the neutral version of Solute 1, *trans*-thiophenyl-*p*-coumarin (IUPAC name: (*E*)-4-(3-oxo-3-(phenylthio)prop-1-en-1-yl)phenol), and a chromophore similar to Solute 2 has an aqueous absorption λ_{max} at 3.68 eV.[72] Solute 3, *S*-phenyl (2*E*)-3-phenyl-2-propenethioate, is made less polar than Solute 2 by removing the −OH group from the phenolate moiety. We have designed Solute 3 as a model system for testing the effects of solute polarity, and to the best of our knowledge, an experimental absorption spectrum is not available for this solute. Solute 4, (1*E*,3*E*)-1,4-diphenylbuta-1,3-diene, is made less polar than Solute 3 by replacing all heteroatoms with saturated carbon atoms. The photoisomerization of Solute 4 has been studied by experiment[84−87] and theory.[88,89] Solute 4 has an absorption λ_{max} at 4.03 eV in vacuum,[85] 3.71−3.78 eV in nonpolar solutions,[84−86] and 3.75 eV in ethanol solution.[87] To the best of our knowledge, an aqueous experimental absorption spectrum is not available for Solute 4 since it is likely not soluble in water. We also compare TDDFT excitation energies computed with an LRC functional to those computed with CIS to investigate the role of exact exchange. We then turn to testing basis set effects on computed absorption spectra and present our findings using mixed basis sets to reduce computational cost.

COMPUTATIONAL DETAILS

MM Dynamics. Each of the four chromophores studied was solvated in a 32 Å radius sphere of water (approximately 4000 molecules). MM dynamics were run using AMBER 12[90] with the SANDER molecular dynamics module.[91] The generalized AMBER force field (GAFF)[92] was used to model the solute, and the TIP3P water model[93] was used for the solvent. After minimization, heating, and equilibration, snapshots from production dynamics were saved every 0.5 ps. Then, 19 randomly selected snapshots were used for the QM/MM calculations. Note that this number is not sufficient to achieve convergence of absorption spectra with respect to the number of snapshots (we find that this requires closer to 1000 snapshots).[9] However, 19 snapshots is sufficient to sample a variety of solute−solvent configurations; these were then used for computing the excitation energies and absorption spectra with different amounts of QM solvent.

QM/MM Calculations. To create the solvation shells, we systematically carved out between 0 and 400 water molecules

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171

closest to the solute to be included in the QM region (see examples with 5 QM solvent molecules, Figure 2a, and with 400 QM solvent molecules, Figure 2b). The QM region is always surrounded with the remaining water molecules, represented by TIP3P point charges that polarize the QM region via electrostatic embedding. The size of the QM region is determined by the number of solvent molecules included in the QM region (i.e., number of QM water molecules). We have also quantified the size of the QM region as the maximum distance between any solute atom and any QM solvent atom. We refer to this distance as the maximum solute−QM solvent distance. Average maximum solute−QM solvent distances across all snapshots for each solute with increasing number of QM water molecules are provided in the Supporting Information, Table S1. The completion of a solvation shell is identified as a relatively large change in average maximum solute−QM solvent distance for an incremental increase in the number of QM water molecules (Figure S1). For all solutes studied here, the first solvation shell consists of ∼50 QM water molecules and corresponds to an average maximum solute−QM solvent distance of ∼5.0 Å, whereas the second solvation shell consists of ∼200 QM water molecules and corresponds to an average maximum solute−QM solvent distance of ∼8.5 Å.

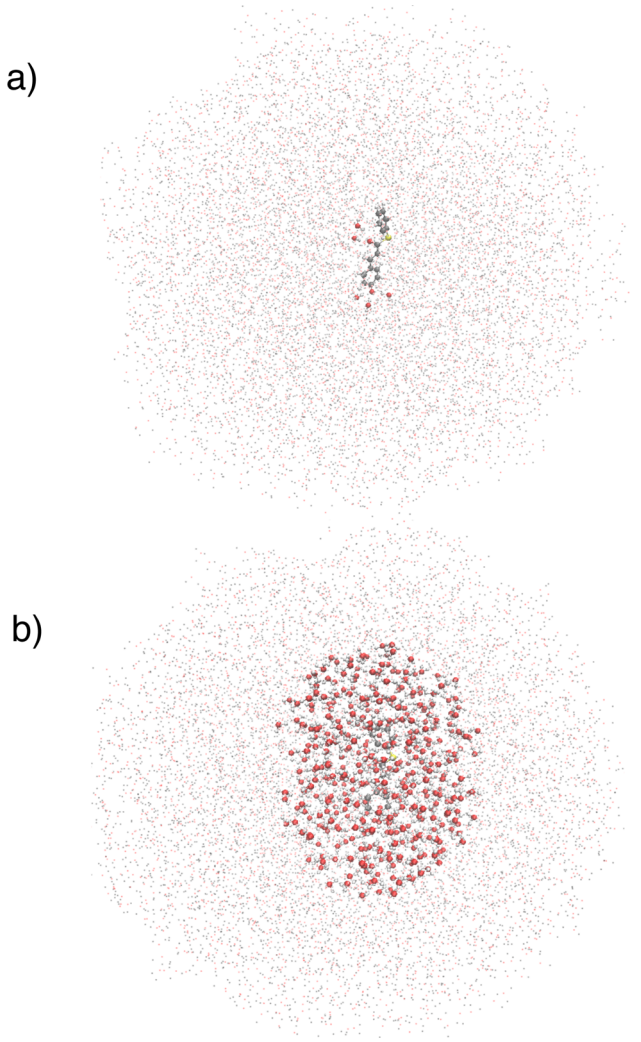


Figure 2. Schematic of the QM/MM solvation shells. The QM region is shown with a ball and stick model, and the MM region is shown with points. A single snapshot with the solute and (a) 5 or (b) 400 water molecules closest to it treated with QM; the remaining water molecules are represented by TIP3P point charges.

The QM/MM excited state calculations were carried out using the TeraChem electronic structure package.[14,16,94,95] For the TDDFT calculations, the long-range corrected LC-ωPBE functional[51] was employed with a range-separation parameter of $\omega = 0.2$ bohr^{-1}. This value of the range-separation parameter has been shown to provide good excitation energies across a variety of molecules.[66] The CIS calculations used the standard Hartree−Fock (HF) ground state as a reference. Calculations with the 3-21G, 6-31G, 6-31G*, and 6-31+G* basis sets were carried out to examine convergence with respect to basis set size. Because all TDDFT excitations up to 5.0 eV were computed, the number of excitations calculated varied significantly among the snapshots (minimum 5 states, maximum 85 states). In the analysis of excitation energy convergence, we chose the transition below 5.0 eV with the largest oscillator strength for the calculation with 400 QM solvent molecules and then traced this transition (based on excitation energy and oscillator strength) through all calculations with smaller QM regions. For the calculated absorption spectra, all TDDFT transitions were included for all snapshots. A Gaussian broadening parameter of 0.04 eV was used to mimic the broadening due to vibronic contributions to the spectra not accounted for in the TDDFT calculation.

RESULTS AND DISCUSSION

Comparison of Solutes and the Role of Exact Exchange. In this section, we compare the convergence of excitation energies computed by TDDFT with the LC-ωPBE functional and CIS for the four solutes studied. Solutes 1−4 decrease in polarity and in their ability to hydrogen bond with solvent, which we expect would lead to a decrease in the magnitude of solute−solvent interactions. The isotropic polarizability computed for Solutes 2−4 is quite similar, but it is larger for anionic Solute 1. Anionic Solute 1 has a single very bright transition (oscillator strength $f \sim 1.0$) that is usually to the lowest energy excited state and is predominantly highest occupied molecular orbital (HOMO) → lowest unoccupied molecular orbital (LUMO) in character. In contrast, for Solutes 2 and 3, there is generally no single bright state among the excitations below 5.0 eV; instead, there are multiple transitions with intermediate oscillator strengths (for example, $f \sim 0.3$). These transitions usually have many orbital contributions in addition to HOMO → LUMO. For 17 of the 19 configurations, Solute 4 has a single bright HOMO → LUMO transition below 5.0 eV, which is the transition we analyze for our bright state excitation energy. For the other two configurations, the bright transition is likely above 5.0 eV, and we instead analyze the convergence of transitions with a smaller oscillator strength. The TDDFT excitation energies and corresponding oscillator strengths for each solute are given in Figures S2 and S3, respectively.

To investigate how convergence varies across solutes, for each chromophore we calculated the deviation of the TDDFT excitation energy with various sizes of the QM region (ΔE) from the value computed with 400 QM water molecules (ΔE_{400}). These results are shown with black lines in Figure 3. For Solutes 1−4, the TD-LC-ωPBE excitation energies for all snapshots are converged within ±0.05 eV at 125, 175, 150, and 225 QM solvent molecules, respectively. The same energy difference data shown as a function of the average maximum

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171

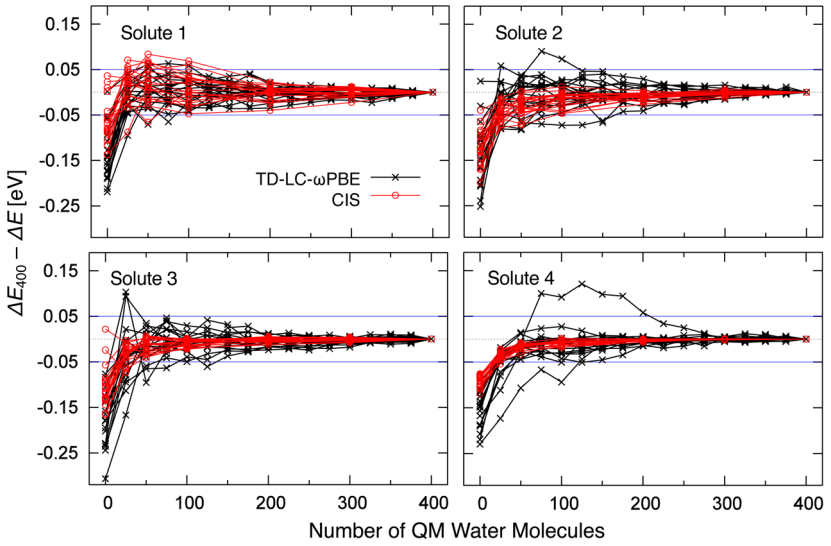


Figure 3. Difference between the calculated excitation energies (eV) with 400 water molecules in the QM region (ΔE_{400}) and the excitation energy with smaller numbers of water molecules in the QM region (ΔE). Black lines: TD-LC-ωPBE; red lines: CIS. The blue lines indicate a deviation of ±0.05 eV from ΔE_{400}.

solute−QM solvent distance is given in Figure S4; the corresponding solute−solvent distances for energy convergence are 7.3, 8.1, 7.7, and 9.0 Å, respectively. Surprisingly, the amount of QM solvent necessary to reach convergence of the TDDFT excitation energy within 0.05 eV does not correlate with solute polarity. Anionic Solute 1 is the most polar and can be expected to have the longest range electrostatic interactions, but it requires less QM solvent and a smaller solute−QM solvent distance for convergence of the excitation energy than neutral Solutes 2 and 3 that are both less polar. There are two configurations that are slow to converge for Solute 4; these are the two configurations with small oscillator strengths compared to the other configurations. If we neglect these configurations, nonpolar Solute 4 converges to within ±0.05 eV with 50 QM water molecules, which corresponds to a single solvation sphere. The smaller amount of QM solvent needed to converge the excitation energy is consistent with our rationale of Solute 4 having shorter-range solute−solvent interactions since it is the least polar of the solutes studied here.

To test whether this trend persists when using other theoretical methods, particularly ones that do not overstabilize charge-transfer states, we also examined the convergence of the CIS excitation energy differences, shown with red lines in Figure 3, with the excitation energies given in Figure S5. We see two notable differences between the TDDFT and CIS results. The first is that the CIS excitation energies with no QM solvent molecules are closer to the ΔE_{400} result than the TDDFT excitation energies, by 0.04−0.09 eV. This suggests that with DFT there is greater polarization of the solute in the presence of the solvent MM point charges compared to HF. This excess polarization could be due to the self-interaction error of approximate DFT methods that leads to overdelocalization of the electron density[51,96] or from DFT's inclusion of dynamic electron correlation that usually increases polarizability.[68] The second difference between TDDFT and CIS results is that, although the convergence trends for both methods are similar for Solute 1, for Solutes 2−4 the CIS excitation energies converge with fewer QM solvent molecules than the TDDFT excitation energies. The CIS excitation energies for Solutes 1−4 converge at 125, 100, 50, and 25 QM solvent molecules, respectively, which corresponds to average maximum solute−QM solvent distances of 7.3, 6.7, 4.9, and 4.1 Å, respectively. The CIS convergence trend is more closely aligned with our expectation that polar solutes in polar solvent should require more QM solvent and larger solute−QM solvent distances to converge than nonpolar solutes in polar solvent.

To better understand the discrepancy between the TDDFT and CIS results, we compared the electron density difference between the ground and excited states of each solute computed by TDDFT and CIS. An isosurface of the electron density difference for a representative snapshot of each solute computed with 100 QM water molecules is shown in Figure 4. The TDDFT density difference shows that Solute 2 has significant electron density originating from nearby water molecules, whereas the CIS density difference for Solute 2 shows no density contribution from the solvent. A similar result is observed for Solute 3. This difference shows that the TDDFT calculations, despite using an LRC hybrid functional, result in significant charge-transfer character between the solute and

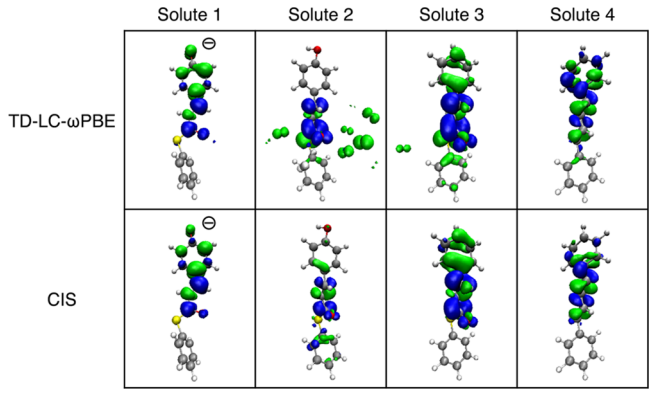


Figure 4. TD-ωPBE and CIS electron density difference between the ground state and the bright excited state computed with 100 QM water molecules (water molecules are not shown). The green isosurface represents an increase and the blue isosurface represents a decrease in the excited state electron density compared to the ground state (isovalues: $\pm 5.0 \times 10^{-5}$).

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171

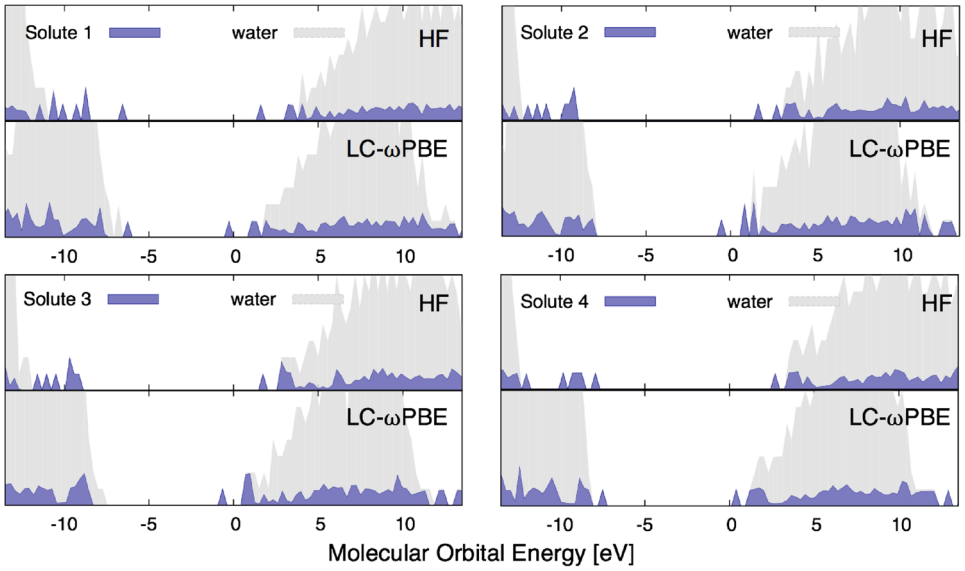


Figure 5. HF and LC-ωPBE density of states for one snapshot of the four solutes with 100 water molecules included in the QM region. The QM region is surrounded by solvent MM point charges.

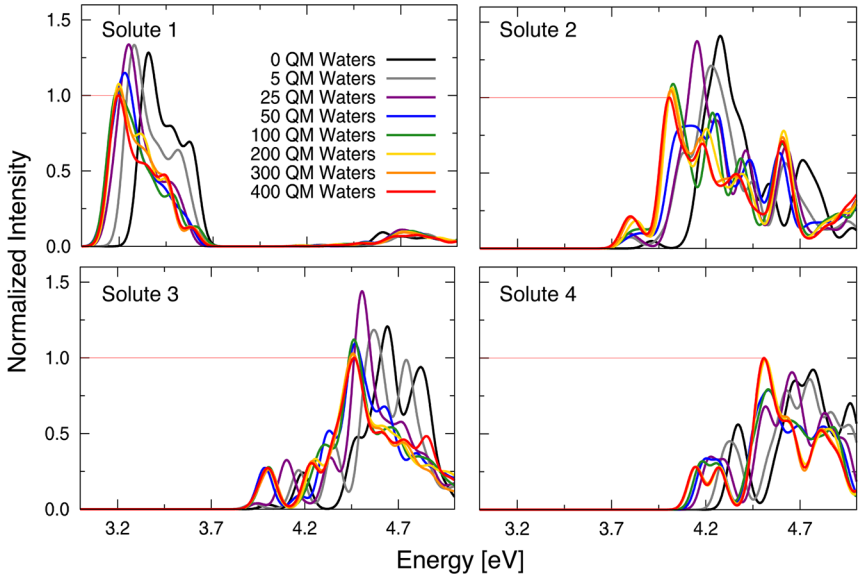


Figure 6. Calculated TD-LC-ωPBE/6-31G* absorption spectra for each solute. All transitions up to 5.0 eV are included. For each solute, spectra are normalized to the maximum intensity (arbitrary units) obtained with 400 water molecules included in the QM region, indicated by the horizontal red line.

solvent that is not seen with the CIS method. The large contribution from the solvent to the excited state in the TDDFT calculations is likely responsible for the slower convergence observed for Solutes 2 and 3 compared to CIS.

Several reasons may account for the increased contribution of water molecules to the electron density difference between the ground and excited states for TDDFT compared to CIS. First, the lack of the electron−hole interaction in the TDDFT excited state equations allows electrons in the excited state to delocalize onto solvent molecules more easily. Another, related, interpretation is that for CIS there is a greater energetic cost for delocalizing the transition density onto the nearby solvent. Given that both methods are known to have an error in their description of charge transfer, it is difficult to say which density picture is more accurate. Second, the inclusion of electron correlation via the DFT correlation functional could increase the polarizability of the QM solute and solvent, resulting in stronger solvent−solute interactions. Perhaps the best way to test if the TDDFT or CIS solute−solvent picture is more correct would be to compare calculated to experimental solvatochromic shifts when going from vacuum to solution. Unfortunately, previous TDDFT calculations for Solute 1 in vacuum are not accurate due to the excitation energies being close in energy to the ionization continuum[73−82] and experimental spectra are not available for Solutes 3 and 4. Experimental spectra are available for a chromophore similar to Solute 2,[72] so this chromophore may be an interesting candidate for further study.

An additional explanation for the difference between the TDDFT and CIS results is that DFT calculations may produce greater energy overlap in the ground state band structure of the solute and the solvent compared to HF. To explore this last

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171

idea, we consider the density of states (i.e., the number of molecular orbitals per unit energy) for each of the four solutes computed with HF and LC-ωPBE; these are shown in Figure 5. For all solutes, the occupied orbitals (valence band) are at energies of −6 eV and below, and the virtual orbitals (conduction band) begin near 0.0 eV. HF gives a larger band gap than DFT, as expected.[97,98] The most striking difference between the two methods is in their prediction of the relative energies of the occupied water orbitals compared to the solute. For LC-ωPBE, the water frontier occupied orbitals have similar energies as the solute HOMO, near −8 eV, whereas for HF these water frontier occupied orbitals are near −11 eV or below, about 3 eV lower in energy than the solute HOMO near −8 eV. For Solutes 1 and 4, the LC-ωPBE occupied water orbitals are lower in energy than the solute HOMO, whereas for Solutes 2 and 3 the frontier occupied water orbitals are similar to or higher in energy than the solute HOMO. With similar energies for the solute HOMO and for these solvent orbitals, the occupied orbitals of water are more likely to mix with the solute HOMO and contribute to the HOMO−LUMO transition. This difference in band structure may be the cause of the slower convergence of LC-ωPBE excitation energies compared to CIS.

The differences in DFT and HF band structure may be partially due to the band gap error of DFT, but the DFT band gap error should be significantly improved with the inclusion of a range-separated hybrid functional.[99] The DFT band gap shown for the water DOS in Figure 5 of ∼8 eV is in fact much closer to the experimental values of 6.8−9.0 eV[100−104] than the HF water band gap of ∼14 eV. Overall, it is difficult to conclude whether the excited state character observed with TDDFT using an LRC hybrid is more accurate than that computed with CIS. The density differences show charge-transfer character in the low energy transitions of Solutes 2 and 3; however, this may be a correct physical result rather than an error of DFT or TDDFT.

Comparison of Solutes and Spectral Convergence. We now go beyond a single excitation energy and examine the convergence of the TDDFT absorption spectra computed from all transitions below 5.0 eV with increasing numbers of solvent molecules in the QM region (Figure 6). The absorption spectra for each solute are normalized to the spectrum computed with 400 QM solvent molecules (arbitrary units). As can be surmised from the magnitude of oscillator strengths (Figure S3), the spectra for Solutes 1 and 4 have much larger intensities than the spectra for Solutes 2 and 3 before normalization. The spectra for all solutes show a shift to lower energies as the QM solvent region is enlarged, although individual snapshot excitation energies may shift to lower or higher energies (see Figure 3). For comparison, analogous absorption spectra calculated for CIS are shown in Figure S6. For Solutes 1 and 2, for which experimental aqueous λ_{max} values are available, the shift to lower energies brings the computed TDDFT λ_{max} closer to the experimental value.

To quantify the spectral convergence, we look at three metrics: (a) the maximum intensity of the normalized spectrum, (b) λ_{max}, and (c) the greatest intensity difference at any energy. In Figure 7, we plot these spectral properties as differences from the absorption spectrum computed with 400 QM water molecules for each solute. The maximum intensity converges to within ±0.10 (arbitrary units) for all solutes at 150−175 QM water molecules. For Solutes 1, 3, and 4, λ_{max} converges to within ±0.05 eV with only 50−75 QM water

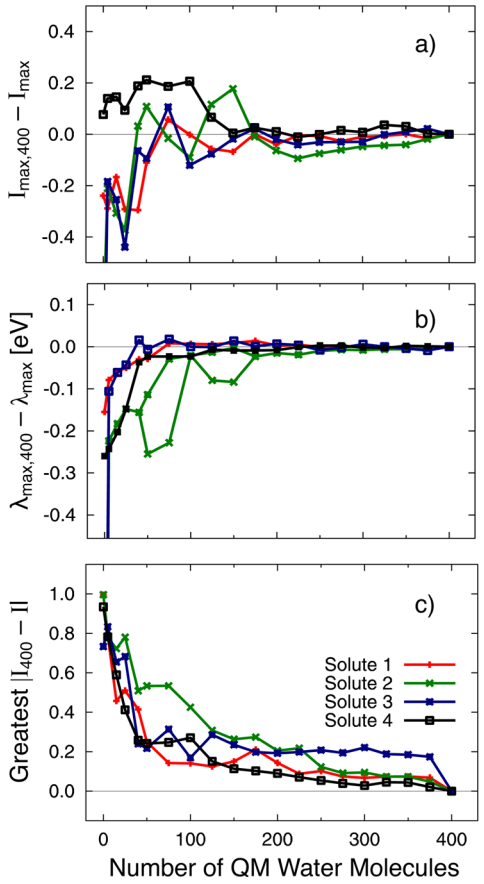


Figure 7. Three metrics of convergence for the TD-ωPBE/6-31G* absorption spectra of each solute in Figure 6: difference in (a) maximum intensity, (b) λ_{max}, and (c) greatest intensity at any energy compared to the spectrum with 400 QM solvent molecules.

molecules, whereas Solute 2 requires 175 QM water molecules. The additional QM water molecules required to converge λ_{max} of Solute 2 increases the maximum solute−QM solvent distance from 4.9 to 6.0 to 8.1 Å. The greatest intensity difference at any energy is converged to within +0.20 (arbitrary units) for Solutes 1 and 4 at 75 and 125 QM water molecules, respectively. Solutes 2 and 3 require 200 and 175 QM water molecules, respectively, to reach the same level of convergence. In part, the convergence of the spectrum with respect to the number of solvent molecules in the QM region is slow for Solute 2 because two peaks compete for the highest intensity. Therefore, in Figure 7b we have plotted the energy of both peaks for QM region sizes between 40 and 175 QM water molecules. For all solutes studied, the convergence trends for the calculated spectra consisting of multiple excitations are similar to the convergence trends for the excitation energies with the largest oscillator strengths for individual snapshots (see Figure 3). Overall, Solutes 1 and 4 have a dominant bright transition that is lower in energy than solvent → solute charge-transfer transitions; these solutes yield a spectrum that is nearly converged with a single layer of QM solvent (∼100 QM water molecules). Solutes 2 and 3 require going beyond the first solvation shell to achieve convergence of the spectral properties.

Basis Set Effects. We now investigate the effect of basis sets on the spectral convergence of Solute 1. In Figure 8, we show the absorption spectra computed from 19 snapshots of Solute 1 with 400 QM water molecules and different basis sets (3-21G,

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171

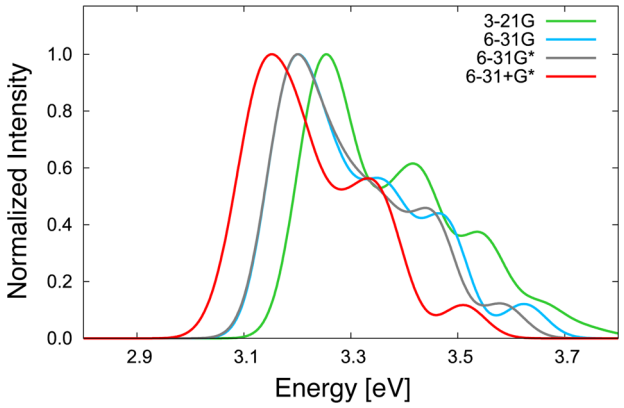


Figure 8. Absorption spectra calculated using TD-ωPBE and various basis sets for Solute 1 with 400 QM water molecules. Each spectrum is normalized.

6-31G, 6-31G*, 6-31+G*). As expected, using a larger basis set leads to a spectral red shift. By comparing the spectra calculated with the 6-31G and 6-31G* basis sets in Figure 8, we find that polarization functions on heavy (non-hydrogen) atoms do not have a significant effect on the absorption spectrum. We also examined one snapshot with 0, 25, 50, and 100 QM water molecules and find that the 6-31G** excitation energies and oscillator strengths are not significantly different from those computed with 6-31G or 6-31G*. Furthermore, excitation energies computed with 6-31++G are only red-shifted by ∼0.01 eV compared to the excitation energies computed with 6-31+G*. Therefore, for Solute 1, including diffuse functions on heavy atoms is more important than for hydrogen atoms, and we include only convergence studies of the spectra computed with the 3-21G, 6-31G, 6-31G*, and 6-31+G* basis sets.

To analyze basis sets effects on the convergence of absorption spectra, we show in Figure 9 the same three metrics as in the previous section: difference in (a) maximum intensity, (b) λ_{max}, and (c) the greatest intensity at any energy for spectra computed with different basis sets (see excitation energies in Figure S7, difference in excitation energies in Figure S8, and computed absorption spectra in Figure S9). We observe that increasing the size of the basis set decreases the number of solvent molecules that need to be treated quantum mechanically to achieve spectral convergence. For instance, the maximum intensity is converged to within ±0.10 (arbitrary units) at 75 QM water molecules for the 6-31G* and 6-31+G* basis sets, whereas the 3-21G and 6-31G basis sets require 125 and 100 QM water molecules, respectively. The 6-31G* and 6-31+G* basis sets converge λ_{max} to within ±0.05 eV with 40 QM water molecules, compared to 75+ QM water molecules for the smaller basis sets. The 6-31G* and 6-31+G* basis sets converge the greatest intensity difference at any energy to within ±0.20 (arbitrary units) with 75 QM water molecules, whereas the 3-21G and 6-31G basis set require 200 QM water molecules. Overall, convergence of all spectral properties is achieved with 75 QM water molecules for the 6-31+G* basis set compared to 200 QM water molecules for the 3-21G spectrum.

We now explore whether it is possible to reduce the computational cost of QM/MM calculations of absorption spectra without a significant loss of accuracy by treating only the solute and the QM water molecules nearest to it with the 6-31+G* basis set (23 basis functions per water molecule) and the remaining QM water molecules with the 3-21G basis set (13 basis functions per water molecule). We performed these

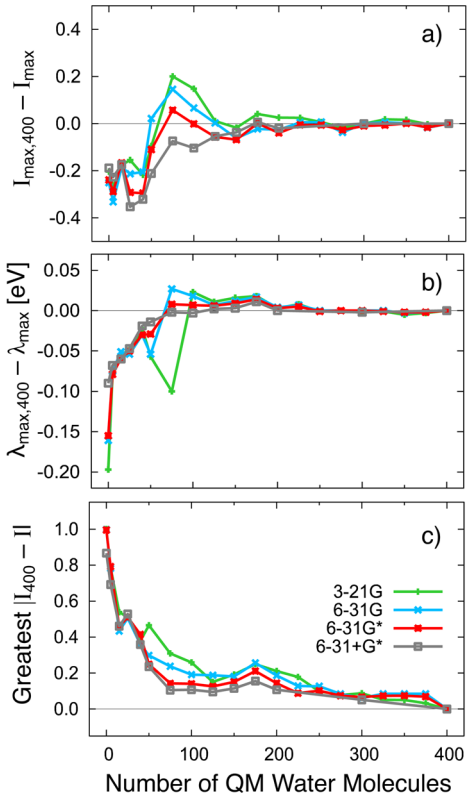


Figure 9. Three metrics of convergence for the TD-ωPBE absorption spectra of Solute 1 computed with various basis sets: difference in (a) maximum intensity, (b) λ_{max}, and (c) greatest intensity at any energy compared to the spectrum with 400 QM solvent molecules.

mixed basis set calculations for 19 snapshots of Solute 1 and aim to reproduce the results computed with 200 QM water molecules described by the 6-31+G* basis set that we will refer to as the "full 6-31+G* basis set" calculation. We performed the mixed basis set calculations with 200 QM water molecules and incrementally changed the ratio of 6-31+G* to 3-21G water molecules from 0:200 to 200:0. Figure 10 shows the excitation energies, oscillator strengths, and computational times for calculations performed using the mixed 6-31+G*/3-21G basis set relative to the full 6-31+G* basis set results as a function of the number of QM water molecules described with the 6-31+G* basis set for a single snapshot (average values are given in Figure S10). Describing all QM water molecules with the 3-21G basis set (i.e., zero 6-31+G* water molecules) converges the oscillator strengths to within ±0.10 (arbitrary units) of the full 6-31+G* basis set results, but it changes the excitation energy by a maximum of ∼0.075 eV. The mixed basis set calculations with 40 QM water molecules described by the 6-31+G* basis set reproduce TDDFT excitation energies to within ±0.05 eV of the values obtained with the full 6-31+G* basis set. With 40 closest QM water molecules described with 6-31+G*, these mixed basis set calculations reduce the computational time to ∼20% of the full 6-31+G* basis set calculations.

Figure 11 shows the TD-ωPBE spectra calculated using the mixed 6-31+G*/3-21G basis set with an increasing number of QM water molecules described by the 6-31+G* basis set (results with additional amounts of QM water molecules are shown in Figure S11). Although all spectra have similar λ_{max} values, the higher energy shoulder requires at least 80 6-31+G*

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171

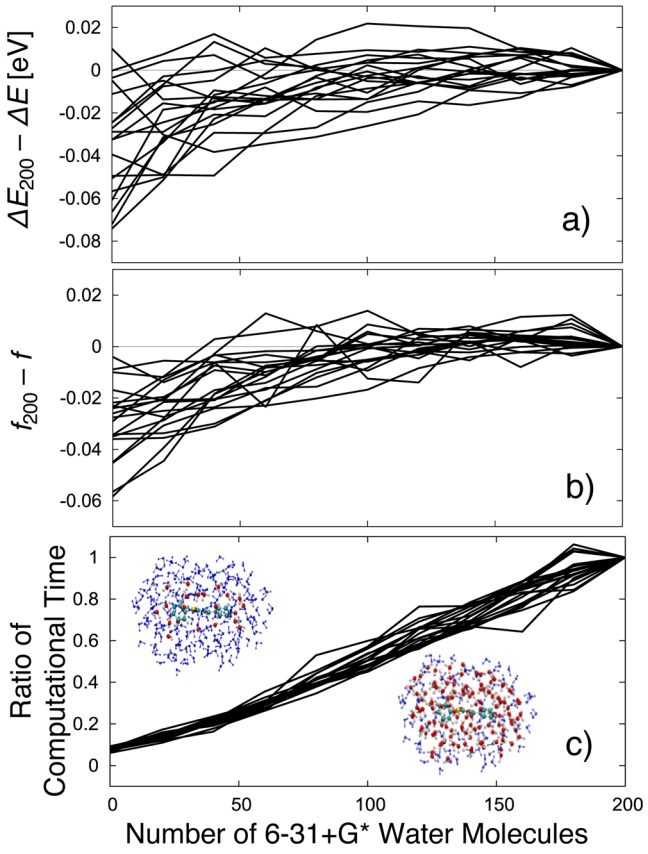


Figure 10. Comparison of the results and performance of TD-ωPBE calculations on Solute 1 with 200 QM solvent molecules performed using a mixed 6-31+G*/3-21G and the 6-31+G* basis sets: (a) difference in calculated excitation energies, (b) difference in calculated oscillator strengths, and (c) ratio of computational times. The inset in panel (c) shows QM water molecules described using the 3-21G basis set in blue for the system with 20 (left) and 180 (right) water molecules closest to the solute described using the 6-31+G* basis set.

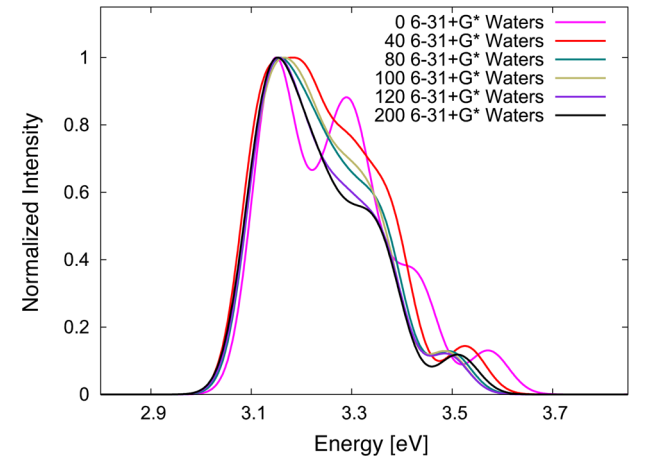


Figure 11. TD-ωPBE computed computed absorption spectra using a mixed 6-31+G*/3-21G and the 6-31+G* basis sets for Solute 1 with 200 QM water molecules. For the mixed basis set, the nearest, 0, 40, 80, 100, and 120 QM water molecules are described with the 6-31+G* basis set, and the remaining QM water molecules are described with the 3-21G basis set. Each spectrum is normalized.

water molecules to match the full 6-31+G* basis set spectrum. With the 80 closest QM water molecules (located up to ~6.0 Å from the solute) described with 6-31+G*, these mixed basis set calculations reduce the computational cost to ~40% of the full 6-31+G* basis set calculations. Overall, the mixed basis set spectra show that it is possible to achieve similar spectral results while using fewer basis functions on the QM solvent molecules. These results indicate that the QM solvent molecules are still able to adequately polarize, transfer charge, and hydrogen bond with the solute to give good spectral energies and shapes even when the outermost QM solvent molecules use a smaller basis set.

CONCLUSIONS

Utilizing large-scale QM/MM calculations with explicit QM solvent, we studied the convergence of excitation energies and absorption spectra with respect to the amount of QM solvent for four chromophores in aqueous solution. We investigated the effects of solute polarity, QM method, and basis set on the convergence of these properties, as well as the accuracy and computational cost of using a mixed basis set to model large amounts of QM solvent.

Overall, 50−200 QM water molecules, or 1−2 solvation shells with maximum solute−QM solvent distances of 5.0−8.6 Å, are required to converge the excitation energies and spectral shape of the four solutes studied. With TDDFT using a long-range corrected functional, the convergence trend of each solute was not found to depend on the polarity of the solute, as would be expected if the solute−solvent interaction is largely an electrostatic effect. Instead, it appears that the convergence trends depend more strongly on the nature of the excited state. For example, Solutes 1 and 4 have a bright transition that is well-separated in energy from solvent → solute transitions and require about one QM solvation shell (maximum solute−QM solvent distance of ~6.0 Å) to converge excitation energies and spectral properties, whereas Solutes 2 and 3 have weak transitions that potentially mix with solvent → solute transitions and require about two QM solvation shells (maximum solute−QM solvent distance of ~8.5 Å).

The CIS method requires a similar or smaller number of QM solvent molecules to converge QM/MM excitation energies relative to TDDFT with a LRC hybrid functional. Unlike the TD-LC-ωPBE convergence trend, the CIS convergence trend correlates with the polarity of the solute. In addition, the TD-LC-ωPBE ground to excited state electron density differences indicate that the solvent contributes substantially to the excited states of Solutes 2 and 3 (largest QM regions required for convergence), but the CIS electron density differences do not indicate solvent contribution for any solute. We identified two possible reasons for these differences. One is that TDDFT, even with an LRC functional such as LC-ωPBE, may still overstabilize charge transfer transitions, leading to increased solvent → solute character in the excited state. On the other hand, because CIS overestimates the energy of charge transfer transitions, this method may overestimate the energetic cost associated with delocalizing the transition to nearby water molecules. The second reason we identify for the different behavior may be that some of the DFT ground state energies of the occupied orbitals localized on water are near the solute HOMO energy, whereas HF occupied water orbitals, used as a reference for CIS, are all lower in energy than the solute HOMO. Although it is unclear whether the strong solvent → solute character of the excited states of Solutes 2 and 3 is of physical origin or due to errors in the QM method, it is likely responsible for the slow convergence of the TD-LC-ωPBE

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171

excitation energies and spectral properties with respect to the size of the QM region.

We found that using a larger basis set reduces the number of QM water molecules necessary to achieve convergence, presumably due to a more accurate description of the solvent polarizability. An approach that may provide the best trade-off between computational efficiency and accuracy is to use a mixed basis set with more basis functions for solvent molecules nearest to the solute and fewer basis functions for the remaining solvent molecules in the QM region. For our mixed 6-31+G*/3-21G basis set with 6-31+G* QM water molecules up to ∼6.0 Å from the solute and 3-21G QM water molecules from ∼6.0 Å to ∼8.5 Å, excitation energies are predicted to within ±0.05 eV of the full 6-31+G* basis set results, but with a ∼20% reduction in computational time compared to the full 6-31+G* basis set calculations.

This study demonstrates that the convergence of TDDFT QM/MM excitation energies and absorption spectra of aqueous chromophores with respect to the size of the QM region depends more strongly on the character of the electronic transition than the polarity of the solute. One of the most encouraging results of this study is the benefit of using a mixed basis set to model large amounts of QM solvent within a QM/MM model. We hope this work helps others to make more informed decisions about where to create the QM/MM boundary, which basis set to use, and which method/functional to use before setting out to create absorption spectra that can be compared to experimental spectra.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00159.

Excitation energies, oscillator strengths, CIS absorption spectra, excitation energies and absorption spectra of Solute 1 with four basis sets, mixed basis set results averaged over all snapshots of Solute 1, and mixed basis set absorption spectra (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: cisborn@ucmerced.edu. Phone: 209-228-4693.

ORCID

Makenzie R. Provorse: 0000-0003-4034-0248

Christine M. Isborn: 0000-0002-4905-9113

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Acknowledgement is made to the donors of the American Chemical Society Petroleum Research Fund for supporting the early stages of this research comparing convergence of the solutes, under Award Number 53674-DNI6. The later stages of this research, including comparison of basis set effects and theoretical methods, were supported by the Department of Energy, Office of Basic Energy Sciences CTC and CPIMS programs, under Award Number DE-SC0014437. We thank Drs. Aleksey Kocherzhenko and Tim Zuehlsdorff for thoughtful comments on this manuscript. M.R.P. gratefully acknowledges the financial support of the University of California, Merced Chancellor's Postdoctoral Fellowship.

■ REFERENCES

(1) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(2) Singh, U. C.; Kollman, P. A. A Combined Ab Initio Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the CH3Cl + Cl- Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.* **1986**, *7*, 718−730.

(3) Bakowies, D.; Thiel, W. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.* **1996**, *100*, 10580−10594.

(4) Monard, G.; Merz, K. M. Combined Quantum Mechanical/Molecular Mechanical Methodologies Applied to Biomolecular Systems. *Acc. Chem. Res.* **1999**, *32*, 904−911.

(5) Sherwood, P. Hybrid Quantum Mechanics/molecular Mechanics Approaches. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing, NIC Series: Jülich, Germany, 2000; pp 257−277.

(6) Gao, J.; Truhlar, D. G. Quantum Mechanical Methods for Enzyme Kinetics. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.

(7) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. Development of Effective Quantum Mechanical/Molecular Mechanical (QM/MM) Methods for Complex Biological Processes. *J. Phys. Chem. B* **2006**, *110*, 6458−6469.

(8) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theor. Chem. Acc.* **2007**, *117*, 185−199.

(9) Isborn, C. M.; Götz, A. W.; Clark, M. A.; Walker, R. C.; Martínez, T. J. Electronic Absorption Spectra from MM and Ab Initio QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein. *J. Chem. Theory Comput.* **2012**, *8*, 5092−5106.

(10) Kapasi, U. J.; Dally, W. J.; Rixner, S.; Mattson, P. R.; Owens, J. D.; Khailany, B. Efficient Conditional Operations for Data-Parallel Architectures. In *Proceedings of the 33rd Annual International Symposium on Microarchitecture*; ACM: Monterey, CA, 2000; pp 159−170.

(11) Anderson, A. G.; Goddard, W. A.; Schröder, P. Quantum Monte Carlo on Graphical Processing Units. *Comput. Phys. Commun.* **2007**, *177*, 298−306.

(12) Yasuda, K. Accelerating Density Functional Calculations with Graphics Processing Unit. *J. Chem. Theory Comput.* **2008**, *4*, 1230−1236.

(13) Vogt, L.; Olivares-Amaya, R.; Kermes, S.; Shao, Y.; Amador-Bedolla, C.; Aspuru-Guzik, A. Accelerating Resolution-of-the-Identity Second-Order Møller−Plesset Quantum Chemistry Calculations with Graphical Processing Units†. *J. Phys. Chem. A* **2008**, *112*, 2049−2057.

(14) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 2. Direct Self-Consistent-Field Implementation. *J. Chem. Theory Comput.* **2009**, *5*, 1004−1015.

(15) Asadchev, A.; Allada, V.; Felder, J.; Bode, B. M.; Gordon, M. S.; Windus, T. L. Uncontracted Rys Quadrature Implementation of up to G Functions on Graphical Processing Units. *J. Chem. Theory Comput.* **2010**, *6*, 696−704.

(16) Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martínez, T. J. Excited-State Electronic Structure with Configuration Interaction Singles and Tamm−Dancoff Time-Dependent Density Functional Theory on Graphical Processing Units. *J. Chem. Theory Comput.* **2011**, *7*, 1814−1823.

(17) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. Introducing ONETEP: Linear-Scaling Density Functional Simulations on Parallel Computers. *J. Chem. Phys.* **2005**, *122*, 084119.

(18) Haynes, P. D.; Skylaris, C.-K.; Mostofi, A. A.; Payne, M. C. ONETEP: Linear-Scaling Density-Functional Theory with Local Orbitals and Plane Waves. *Phys. Status Solidi B* **2006**, *243*, 2489−2499.

(19) Fox, S. J.; Pittock, C.; Fox, T.; Tautermann, C. S.; Malcolm, N.; Skylaris, C.-K. Electrostatic Embedding in Large-Scale First Principles

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171

Quantum Mechanical Calculations on Biomolecules. *J. Chem. Phys.* **2011**, *135*, 224107.
(20) Dziedzic, J.; Fox, S. J.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. Large-Scale DFT Calculations in Implicit Solvent-A Case Study on the T4 Lysozyme L99A/M102Q Protein. *Int. J. Quantum Chem.* **2013**, *113*, 771−785.
(21) Dziedzic, J.; Mao, Y.; Shao, Y.; Ponder, J.; Head-Gordon, T.; Head-Gordon, M.; Skylaris, C.-K. TINKTEP: A Fully Self-Consistent, Mutually Polarizable QM/MM Approach Based on the AMOEBA Force Field. *J. Chem. Phys.* **2016**, *145*, 124106.
(22) Bowler, D. R.; Choudhury, R.; Gillan, M. J.; Miyazaki, T. Recent Progress with Large-Scaleab Initio Calculations: The CONQUEST Code. *Phys. Status Solidi B* **2006**, *243*, 989−1000.
(23) Bowler, D. R.; Miyazaki, T. Calculations for Millions of Atoms with Density Functional Theory: Linear Scaling Shows Its Potential. *J. Phys.: Condens. Matter* **2010**, *22*, 074207.
(24) de Pablo, P. J.; Moreno-Herrero, F.; Colchero, J.; Gómez Herrero, J.; Herrero, P.; Baró, A. M.; Ordejón, P.; Soler, J. M.; Artacho, E. Absence of Dc-Conductivity in λ-DNA. *Phys. Rev. Lett.* **2000**, *85*, 4992−4995.
(25) Soler, J. M.; Artacho, E.; Gale, J. D.; García, A.; Junquera, J.; Ordejón, P.; Sánchez-Portal, D. The SIESTA Method for Ab Initio Order-N Materials Simulation. *J. Phys.: Condens. Matter* **2002**, *14*, 2745−2779.
(26) Ozaki, T. Variationally Optimized Atomic Orbitals for Large-Scale Electronic Structures. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2003**, *67*, 155108.
(27) Bowler, D. R.; Miyazaki, T. O(N) Methods in Electronic Structure Calculations. *Rep. Prog. Phys.* **2012**, *75*, 036503.
(28) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Charge Transfer and Polarization in Solvated Proteins from Ab Initio Molecular Dynamics. *J. Phys. Chem. Lett.* **2011**, *2*, 1789−1793.
(29) Zuehlsdorff, T. J.; Haynes, P. D.; Hanke, F.; Payne, M. C.; Hine, N. D. M. Solvent Effects on Electronic Excitations of an Organic Chromophore. *J. Chem. Theory Comput.* **2016**, *12*, 1853−1861.
(30) Filippi, C.; Buda, F.; Guidoni, L.; Sinicropi, A. Bathochromic Shift in Green Fluorescent Protein: A Puzzle for QM/MM Approaches. *J. Chem. Theory Comput.* **2012**, *8*, 112−124.
(31) Daday, C.; Curutchet, C.; Sinicropi, A.; Mennucci, B.; Filippi, C. Chromophore−Protein Coupling beyond Nonpolarizable Models: Understanding Absorption in Green Fluorescent Protein. *J. Chem. Theory Comput.* **2015**, *11*, 4825−4839.
(32) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J. How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol *O* -Methyltransferase. *J. Phys. Chem. B* **2016**, *120*, 11381−11394.
(33) Bose, S.; Chakrabarty, S.; Ghosh, D. Effect of Solvation on Electron Detachment and Excitation Energies of a Green Fluorescent Protein Chromophore Variant. *J. Phys. Chem. B* **2016**, *120*, 4410−4420.
(34) Karelina, M.; Kulik, H. J. Systematic Quantum Mechanical Region Determination in QM/MM Simulation. *J. Chem. Theory Comput.* **2017**, *13*, 563−576.
(35) Murugan, N. A.; Jha, P. C.; Rinkevicius, Z.; Ruud, K.; Ågren, H. Solvatochromic Shift of Phenol Blue in Water from a Combined Car−Parrinello Molecular Dynamics Hybrid Quantum Mechanics-Molecular Mechanics and ZINDO Approach. *J. Chem. Phys.* **2010**, *132*, 234508.
(36) Neugebauer, J.; Jacob, C. R.; Wesolowski, T. A.; Baerends, E. J. An Explicit Quantum Chemical Method for Modeling Large Solvation Shells Applied to Aminocoumarin C151. *J. Phys. Chem. A* **2005**, *109*, 7805−7814.
(37) Provorse, M. R.; Peev, T.; Xiong, C.; Isborn, C. M. Convergence of Excitation Energies in Mixed Quantum and Classical Solvent: Comparison of Continuum and Point Charge Models. *J. Phys. Chem. B* **2016**, *120*, 12148−12159.
(38) Flaig, D.; Beer, M.; Ochsenfeld, C. Convergence of Electronic Structure with the Size of the QM Region: Example of QM/MM NMR Shieldings. *J. Chem. Theory Comput.* **2012**, *8*, 2260−2271.
(39) Pavone, M.; Brancato, G.; Morelli, G.; Barone, V. Spectroscopic Properties in the Liquid Phase: Combining High-Level Ab Initio Calculations and Classical Molecular Dynamics. *ChemPhysChem* **2006**, *7*, 148−156.
(40) Nielsen, C. B. O.; Mikkelsen, K. V. Optical Properties of Pyridine and Methyl-Pyridinium in Water Using DFT/MM. *Mol. Phys.* **2015**, *113*, 3253−3263.
(41) Ma, H.; Ma, Y. Solvatochromic Shifts of Polar and Non-Polar Molecules in Ambient and Supercritical Water: A Sequential Quantum Mechanics/molecular Mechanics Study Including Solute-Solvent Electron Exchange-Correlation. *J. Chem. Phys.* **2012**, *137*, 214504.
(42) Handy, N. C.; Cohen, A. J. Left-Right Correlation Energy. *Mol. Phys.* **2001**, *99*, 403−412.
(43) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta−Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.
(44) Zhao, Y.; Truhlar, D. G. Hybrid Meta Density Functional Theory Methods for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions: The MPW1B95 and MPWB1K Models and Comparative Assessments for Hydrogen Bonding and van Der Waals Interactions. *J. Phys. Chem. A* **2004**, *108*, 6908−6918.
(45) Xu, X.; Goddard, W. A. From The Cover: The X3LYP Extended Density Functional for Accurate Descriptions of Nonbond Interactions, Spin States, and Thermochemical Properties. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2673−2677.
(46) Becke, A. D. Perspective: Fifty Years of Density-Functional Theory in Chemical Physics. *J. Chem. Phys.* **2014**, *140*, 18A301.
(47) Whittleton, S. R.; Sosa Vazquez, X. A.; Isborn, C. M.; Johnson, E. R. Density-Functional Errors in Ionization Potential with Increasing System Size. *J. Chem. Phys.* **2015**, *142*, 184106.
(48) Sosa Vazquez, X. A.; Isborn, C. M. Size-Dependent Error of the Density Functional Theory Ionization Potential in Vacuum and Solution. *J. Chem. Phys.* **2015**, *143*, 244105.
(49) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. A Long-Range-Corrected Time-Dependent Density Functional Theory. *J. Chem. Phys.* **2004**, *120*, 8425−8433.
(50) Yanai, T.; Tew, D. P.; Handy, N. C. A New Hybrid Exchange−correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51−57.
(51) Vydrov, O. A.; Scuseria, G. E. Assessment of a Long-Range Corrected Hybrid Functional. *J. Chem. Phys.* **2006**, *125*, 234109.
(52) Maitra, N. T. Perspective: Fundamental Aspects of Time-Dependent Density Functional Theory. *J. Chem. Phys.* **2016**, *144*, 220901.
(53) Dreuw, A.; Weisman, J. L.; Head-Gordon, M. Long-Range Charge-Transfer Excited States in Time-Dependent Density Functional Theory Require Non-Local Exchange. *J. Chem. Phys.* **2003**, *119*, 2943.
(54) Dreuw, A.; Head-Gordon, M. Single-Reference Ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chem. Rev.* **2005**, *105*, 4009−4037.
(55) Stein, T.; Kronik, L.; Baer, R. Reliable Prediction of Charge Transfer Excitations in Molecular Complexes Using Time-Dependent Density Functional Theory. *J. Am. Chem. Soc.* **2009**, *131*, 2818−2820.
(56) Autschbach, J. Charge-Transfer Excitations and Time-Dependent Density Functional Theory: Problems and Some Proposed Solutions. *ChemPhysChem* **2009**, *10*, 1757−1760.
(57) Moore, B.; Sun, H.; Govind, N.; Kowalski, K.; Autschbach, J. Charge-Transfer Versus Charge-Transfer-Like Excitations Revisited. *J. Chem. Theory Comput.* **2015**, *11*, 3305−3320.
(58) Bernasconi, L.; Sprik, M.; Hutter, J. Time Dependent Density Functional Theory Study of Charge-Transfer and Intramolecular Electronic Excitations in Acetone−water Systems. *J. Chem. Phys.* **2003**, *119*, 12417.
(59) Neugebauer, J.; Louwerse, M. J.; Baerends, E. J.; Wesolowski, T. A. The Merits of the Frozen-Density Embedding Scheme to Model Solvatochromic Shifts. *J. Chem. Phys.* **2005**, *122*, 094115.

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171

(60) Lange, A.; Herbert, J. M. Simple Methods To Reduce Charge-Transfer Contamination in Time-Dependent Density-Functional Calculations of Clusters and Liquids. *J. Chem. Theory Comput.* **2007**, *3*, 1680−1690.
(61) Isborn, C. M.; Mar, B. D.; Curchod, B. F. E.; Tavernelli, I.; Martínez, T. J. The Charge Transfer Problem in Density Functional Theory Calculations of Aqueously Solvated Molecules. *J. Phys. Chem. B* **2013**, *117*, 12189−12201.
(62) Lever, G.; Cole, D. J.; Hine, N. D. M.; Haynes, P. D.; Payne, M. C. Electrostatic Considerations Affecting the Calculated HOMO−LUMO Gap in Protein Molecules. *J. Phys.: Condens. Matter* **2013**, *25*, 152101.
(63) Eilmes, A. Solvatochromic Probe in Molecular Solvents: Implicit versus Explicit Solvent Model. *Theor. Chem. Acc.* **2014**, *133*, 1538.
(64) Liu, J.; Herbert, J. M. Local Excitation Approximations to Time-Dependent Density Functional Theory for Excitation Energies in Solution. *J. Chem. Theory Comput.* **2016**, *12*, 157−166.
(65) Chai, J.-D.; Head-Gordon, M. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J. Chem. Phys.* **2008**, *128*, 084106.
(66) Rohrdanz, M. A.; Herbert, J. M. Simultaneous Benchmarking of Ground- and Excited-State Properties with Long-Range-Corrected Density Functional Theory. *J. Chem. Phys.* **2008**, *129*, 034107.
(67) Subotnik, J. E. Communication: Configuration Interaction Singles Has a Large Systematic Bias against Charge-Transfer States. *J. Chem. Phys.* **2011**, *135*, 071104.
(68) Sim, F.; Chin, S.; Dupuis, M.; Rice, J. E. Electron Correlation Effects in Hyperpolarizabilities of P-Nitroaniline. *J. Phys. Chem.* **1993**, *97*, 1158−1163.
(69) Goerigk, L.; Grimme, S. Assessment of TD-DFT Methods and of Various Spin Scaled CIS(D) and CC2 Versions for the Treatment of Low-Lying Valence Excitations of Large Organic Dyes. *J. Chem. Phys.* **2010**, *132*, 184103.
(70) Gieseking, R. L.; Ratner, M. A.; Schatz, G. C. Implementation of INDO/SCI with COSMO Implicit Solvation and Benchmarking for Solvatochromic Shifts. *J. Phys. Chem. A* **2016**, *120*, 9878−9885.
(71) Georg, H. C.; Coutinho, K.; Canuto, S. A Sequential Monte Carlo Quantum Mechanics Study of the Hydrogen-Bond Interaction and the Solvatochromic Shift of the $n-\pi^*$ Transition of Acrolein in Water. *J. Chem. Phys.* **2005**, *123*, 124307.
(72) Nielsen, I. B.; Boyé-Péronne, S.; El Ghazaly, M. O. A.; Kristensen, M. B.; Brøndsted Nielsen, S.; Andersen, L. H. Absorption Spectra of Photoactive Yellow Protein Chromophores in Vacuum. *Biophys. J.* **2005**, *89*, 2597−2604.
(73) Sergi, A.; Grüning, M.; Ferrario, M.; Buda, F. Density Functional Study of the Photoactive Yellow Protein's Chromophore. *J. Phys. Chem. B* **2001**, *105*, 4386−4391.
(74) González, E. M.; Guidoni, L.; Molteni, C. Chemical and Protein Shifts in the Spectrum of the Photoactive Yellow Protein: A Time-Dependent Density Functional Theory/Molecular Mechanics Study. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4556−4563.
(75) Rocha-Rinza, T.; Christiansen, O.; Rajput, J.; Gopalan, A.; Rahbek, D. B.; Andersen, L. H.; Bochenkova, A. V.; Granovsky, A. A.; Bravaya, K. B.; Nemukhin, A. V.; Christiansen, K. L.; Nielsen, M. B. Gas Phase Absorption Studies of Photoactive Yellow Protein Chromophore Derivatives. *J. Phys. Chem. A* **2009**, *113*, 9442−9449.
(76) Rocha-Rinza, T.; Sneskov, K.; Christiansen, O.; Ryde, U.; Kongsted, J. Unraveling the Similarity of the Photoabsorption of Deprotonated P-Coumaric Acid in the Gas Phase and within the Photoactive Yellow Protein. *Phys. Chem. Chem. Phys.* **2011**, *13*, 1585−1589.
(77) Thellamurege, N. M.; Cui, F.; Li, H. Quantum Mechanical/Molecular Mechanical/Continuum Style Solvation Model: Time-Dependent Density Functional Theory. *J. Chem. Phys.* **2013**, *139*, 084106.
(78) Larsen, D. S.; Vengris, M.; van Stokkum, I. H. M.; van der Horst, M. A.; de Weerd, F. L.; Hellingwerf, K. J.; van Grondelle, R. Photoisomerization and Photoionization of the Photoactive Yellow Protein Chromophore in Solution. *Biophys. J.* **2004**, *86*, 2538−2550.
(79) Gromov, E. V.; Burghardt, I.; Köppel, H.; Cederbaum, L. S. Electronic Structure of the PYP Chromophore in Its Native Protein Environment. *J. Am. Chem. Soc.* **2007**, *129*, 6798−6806.
(80) Lammich, L.; Rajput, J.; Andersen, L. H. Photodissociation Pathways of Gas-Phase Photoactive Yellow Protein Chromophores. *Phys. Rev. E* **2008**, *78*, 051916.
(81) Zuev, D.; Bravaya, K. B.; Crawford, T. D.; Lindh, R.; Krylov, A. I. Electronic Structure of the Two Isomers of the Anionic Form of *P*-Coumaric Acid Chromophore. *J. Chem. Phys.* **2011**, *134*, 034310.
(82) Mooney, C. R. S.; Parkes, M. A.; Iskra, A.; Fielding, H. H. Controlling Radical Formation in the Photoactive Yellow Protein Chromophore. *Angew. Chem., Int. Ed.* **2015**, *54*, 5646−5649.
(83) Wang, Y.; Li, H. Excited State Geometry of Photoactive Yellow Protein Chromophore: A Combined Conductorlike Polarizable Continuum Model and Time-Dependent Density Functional Study. *J. Chem. Phys.* **2010**, *133*, 034108.
(84) Yee, W. A.; Hug, S. J.; Kliger, D. S. Direct and Sensitized Photoisomerization of 1,4-Diphenylbutadienes. *J. Am. Chem. Soc.* **1988**, *110*, 2164−2169.
(85) Dahl, K.; Biswas, R.; Maroncelli, M. The Photophysics and Dynamics of Diphenylbutadiene in Alkane and Perfluoroalkane Solvents. *J. Phys. Chem. B* **2003**, *107*, 7838−7853.
(86) Saltiel, J.; Krishna, T. R. S.; Laohhasurayotin, K.; Ren, Y.; Phipps, K.; Davis, P. H.; Yee, W. A. Medium Effects on the Direct Cis−Trans Photoisomerization of 1,4-Diphenyl-1,3-Butadiene in Solution. *J. Phys. Chem. A* **2011**, *115*, 2120−2129.
(87) Saltiel, J.; Redwood, C. E. Photochemistry of the 1,4-Diphenyl-1,3-Butadienes in Ethanol. Trapping Conical Intersections. *J. Phys. Chem. A* **2016**, *120*, 2832−2840.
(88) Saltiel, J.; Dmitrenko, O.; Pillai, Z. S.; Klima, R.; Wang, S.; Wharton, T.; Huang, Z.-N.; van de Burgt, L. J.; Arranz, J. Triplet and Ground State Potential Energy Surfaces of 1,4-Diphenyl-1,3-Butadiene: Theory and Experiment. *Photochem. Photobiol. Sci.* **2008**, *7*, 566−577.
(89) Zheng, X.; Zhai, G.; Gao, W.; Lei, Y.; Yu, L.; Zhu, C. Photoisomerization Mechanisms from Trans, Trans-1,4-Diphenyl-1,3-Butadiene: CASSCF on-the-Fly Trajectory Surface Hopping Dynamic Simulations. *Phys. Chem. Chem. Phys.* **2016**, *18*, 8971−8979.
(90) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; University of California San Francisco: San Francisco, CA, 2012.
(91) *AMBER Home Page*. http://ambermd.org (accessed Aug 13, 2015).
(92) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General AMBER Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(93) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.
(94) Luehr, N.; Ufimtsev, I. S.; Martínez, T. J. Dynamic Precision for Electron Repulsion Integral Evaluation on Graphical Processing Units (GPUs). *J. Chem. Theory Comput.* **2011**, *7*, 949−954.
(95) *PetaChem*. http://www.petachem.com (accessed Aug 13, 2015).
(96) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792−794.
(97) Muscat, J.; Wander, A.; Harrison, N. M. On the Prediction of Band Gaps from Hybrid Functional Theory. *Chem. Phys. Lett.* **2001**, *342*, 397−401.
(98) Baerends, E. J.; Gritsenko, O. V.; van Meer, R. The Kohn−Sham Gap, the Fundamental Gap and the Optical Gap: The Physical Meaning of Occupied and Virtual Kohn−Sham Orbital Energies. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16408−16425.

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171

(99) Seidl, A.; Görling, A.; Vogl, P.; Majewski, J. A.; Levy, M. Generalized Kohn-Sham Schemes and the Band-Gap Problem. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *53*, 3764−3774.
(100) Painter, L. R.; Hamm, R. N.; Arakawa, E. T.; Birkhoff, R. D. Electronic Properties of Liquid Water in the Vacuum Ultraviolet. *Phys. Rev. Lett.* **1968**, *21*, 282−284.
(101) Goulet, T.; Bernas, A.; Ferradini, C.; Jay-Gerin, J.-P. On the Electronic Structure of Liquid Water: Conduction-Band Tail Revealed by Photoionization Data. *Chem. Phys. Lett.* **1990**, *170*, 492−496.
(102) Bernas, A.; Ferradini, C.; Jay-Gerin, J.-P. On the Electronic Structure of Liquid Water: Facts and Reflections. *Chem. Phys.* **1997**, *222*, 151−160.
(103) Coe, J. V.; Earhart, A. D.; Cohen, M. H.; Hoffman, G. J.; Sarkas, H. W.; Bowen, K. H. Using Cluster Studies to Approach the Electronic Structure of Bulk Water: Reassessing the Vacuum Level, Conduction Band Edge, and Band Gap of Water. *J. Chem. Phys.* **1997**, *107*, 6023.
(104) Coe, J. V. Fundamental Properties of Bulk Water from Cluster Ion Data. *Int. Rev. Phys. Chem.* **2001**, *20*, 33−58.

DOI: 10.1021/acs.jctc.7b00159
J. Chem. Theory Comput. 2017, 13, 2159−2171