Simple Methods To Reduce Charge-Transfer Contamination in Time-Dependent Density-Functional Calculations of Clusters and Liquids

Adrian Lange and John M. Herbert*
Department of Chemistry, The Ohio State University, Columbus, Ohio 43210
Received May 23, 2007

Abstract: Using as benchmarks a series of increasingly large hydrated uracil clusters, we examine the nature and extent of charge-transfer (CT) contamination in condensed-phase, time-dependent density-functional theory. These calculations are plagued by a large number of spurious CT excitations at energies comparable to (and sometimes below) the valence excitation energies, even when hybrid density functionals are used. Spurious states below the first nπ* and ππ* states of uracil are observed in clusters as small as uracil–(H2O)4. Reasonable electronic absorption spectra can still be obtained, upon configurational averaging, despite pervasive CT contamination, but the spurious states add significantly to the cost of the calculations and severely complicate attempts to locate optically dark nπ* states. The extent of CT contamination is reduced substantially by introducing an electrostatic (point charge) description of an extended solvent network, even in cases where the region of solvent described by density functional theory is large (>120 atoms). Alternatively, CT contamination may be reduced by eliminating certain excitation amplitudes from the linear response equations, with minimal loss of accuracy (<0.1 eV) in the valence excitation energies.

I. Introduction

Time-dependent density functional theory (TD-DFT) is currently the most popular method for calculating excited electronic states of gas-phase molecules with ~10–200 atoms, owing to its favorable computational scaling (cubic or better with respect to system size)1,2 and reasonable accuracy (0.2–0.3 eV for the lowest few valence excitations).2–4 Condensed-phase TD-DFT calculations, on the other hand, are beset by serious contamination from spurious, low-energy charge-transfer (CT) excited states,5–11 the proximate cause of which is TD-DFT’s tendency to underestimate long-range CT excitation energies.2,3,12–15 Although this problem is present already in the gas phase (and will manifest itself in TD-DFT calculations of well-separated molecules14 or even sufficiently large single molecules),15–17 it is much more pervasive in liquids and clusters.

Underestimation of long-range CT energetics is a consequence of incorrect asymptotic behavior on the part of the exchange-correlation potential,14 and several long-range correction schemes have been developed recently in an attempt to alleviate this problem.18–23 These corrections appear to mitigate CT problems for well-separated molecules in the gas phase, though only one of them has been tested in a cluster environment.11 Furthermore, these corrections do not rectify all of the problems associated with the long-range behavior of existing density functionals,24 and moreover the improved asymptotic behavior sometimes comes at the expense of diminished accuracy for ground-state properties.25 In the present work, we explore some alternative methods for reducing CT contamination that are different from (though fully compatible with) these long-range correction schemes.

Several previous assessments of the performance of TD-DFT in liquids and clusters have focused exclusively on weakly allowed nπ→π* excitations in systems such as aqueous acetone5,6,10,11 and aqueous formamide.9 In the case of acetone in liquid water5,6 or water clusters,10 spurious CT bands overlap the lowest nπ→π* band at 4.5 eV when nonhybrid
ChARGE-TRANSFER CONTAMINATION IN DENSITY-FUNCTIONAL THEORY

(but gradient-corrected) density functionals are employed. Hartree–Fock exchange does have the correct long-range behavior for CT states, and hybrid functionals with 20–25% Hartree–Fock exchange are found to remove CT contamination from the lowest valence band, by pushing the offending CT states to higher excitation energies.6

In the present work, we use uracil as a typical example of a molecule possessing both bright states (\(1\pi \pi^*\)) and dark states (\(1\pi^*\)). Our results for uracil–water clusters demonstrate that hybrid functionals alone do not guarantee that the lowest valence band will be free of CT contamination; clusters as small as uracil–(H\(_2\)O)\(_4\) exhibit spurious CT states at energies comparable to or below the lowest \(n \pi^*\) and \(\pi \pi^*\) excitation energies. These extra states significantly increase the cost of the calculations, in both time and memory, and for a large cluster like uracil–(H\(_2\)O)\(_4\), the memory bottleneck precludes us from calculating any states at all above 6 eV.

Two simple procedures to reduce CT contamination are examined here. First, we demonstrate that a mixed quantum mechanics/molecular mechanics (QM/MM) formalism significantly reduces the number of spurious CT states, as compared to calculations performed on the gas-phase QM region. This is true even for large QM regions and allows us to calculate a full electronic absorption spectrum for a QM region consisting of uracil–(H\(_2\)O)\(_4\). In conjunction with liquid-phase QM/MM calculations, or on its own in the gas phase, spurious CT states can also be removed by omitting TD-DFT excitation amplitudes that correspond to long-range CT. For the present systems, this typically increases the valence excitation energies by \(\approx 0.1\) eV.

II. Computational Details

As the only long-range component of contemporary density functionals, Hartree–Fock exchange is known to reduce contamination from long-range CT excited states by pushing these states to higher excitation energies.6,13,14,17 As such, our study will focus primarily on the hybrid functionals B3LYP27,31 and PBE0,28–30 though for comparison we present a few results obtained with the nonhybrid functional BLYP.27,31 The PBE0 functional (also known as PBE1PBE)29 consists of PBE correlation in conjunction with 25% Hartree–Fock exchange and 75% PBE exchange and has been specifically recommended for excited-state calculations.32,33 While a larger fraction of Hartree–Fock exchange—for example, Becke’s “half and half” mixture of Hartree–Fock and Slater exchange,34 in conjunction with LYP correlation—can reduce the overall number of CT states even further, this functional is less accurate for valence excitation energies as well as for ground-state thermochemistry.35 Newer, highly parametrized functionals that include full Hartree–Fock exchange may be superior in these respects,36 but such functionals are not yet widely available, nor have they been widely tested. We shall restrict our attention to the popular hybrids B3LYP and PBE0.

All TD-DFT calculations reported here employ the Tamm–Dancoff approximation and were performed using Q-Chem.38 Only singlet excitations are considered. Density plots were rendered with the Visual Molecular Dynamics program using a contour value of 0.001 au in all cases.

The basis-set dependence of the lowest \(n \pi^*\) and \(\pi \pi^*\) excitation energies in uracil–water clusters appears to be very mild, as demonstrated by benchmark calculations for uracil–(H\(_2\)O)\(_4\) that are listed in Table 1. For both B3LYP and PBE0, excitation energies obtained with the 6-31+G* basis set differ by no more than 0.1 eV from those obtained with much larger basis sets. As such, all TD-DFT calculations will employ 6-31+G*, along with the SG-0 quadrature grid.

Our interest lies in liquid-phase environments, and thus we wish to employ uracil–water geometries representative of aqueous uracil rather than a gas-phase cluster. We obtain such geometries from a molecular dynamics (MD) simulations of aqueous uracil at constant temperature (298 K) and density (0.9989 g/cm\(^3\)). Uracil was added to a pre-equilibrated, 25 Å × 25 Å × 25 Å periodic box of flexible water molecules, which was then re-equilibrated using 300 ps of MD. The AMBER9941 and TIP3P42 force fields (as implemented in the Tinker software package) were used for uracil and for water, respectively. Following equilibration, uracil–water clusters were extracted from the simulation based on distance criteria that are described in section III. Water molecules near the uracil (according to these criteria) are included explicitly in the TD-DFT calculations, while additional water molecules up to 20.0 Å away (about 2300 molecules) are incorporated, in some cases, as TIP3P point charges.

III. Results and Discussion

A. CT Contamination in Uracil–Water Clusters. In an effort to understand just how “long range” the long-range CT problem in TD-DFT really is, we performed TD-DFT calculations on a sequence of increasingly large uracil–water clusters extracted from the MD simulation described in section II, by selecting all water molecules having at least one atom within a specified distance \(d\) of any uracil atom. All other water molecules were discarded. All clusters were generated from the same MD snapshot, so that each successively larger cluster contains the smaller clusters as its core, and these clusters range in size from bare uracil (when \(d = 1.5\) Å) to uracil–(H\(_2\)O)\(_{37}\) (when \(d = 4.5\) Å).

For each cluster in this sequence, we calculated the first 40 TD-PBE0/6-31+G* excited states. Table 2 summarizes...
the results, including two simple measures of the extent of CT contamination: the excitation energy \( \omega_{40} \) of the 40th state above the ground state and the number of excited states within 6 eV of the ground state. (In these clusters, the second electronic absorption band typically consists of a few states in the 6.0–6.5 eV range, so 6 eV provides a lower bound to the number of TD-DFT excited states that must be calculated in order to reach this second band.)

At the TD-PBE0/6-31+G* level, bare uracil possesses five excited states below 6 eV, the lowest two of which are an \( \pi\pi^* \) dark state (at 4.56 eV) and a \( \pi\pi^* \) bright state (at 5.33 eV). There are also two more dark states of mixed \( \pi\pi^* \)/Rydberg character, plus one optically allowed \( n\pi^* \) state whose oscillator strength is 30\% of that associated with the \( \pi\pi^* \) state. (Uracil is slightly nonplanar in the geometries extracted from the MD simulation, so we use “bright” and “dark” as qualitative descriptions of transition intensities. The “optically allowed” \( n\pi^* \) state, for example, correlates in a planar chromophore to an excitation out of an \( \alpha^* \) lone pair orbital.)

The 40 excitations calculated for bare uracil reach 9 eV above the ground state, but due to the appearance of spurious CT states, \( \omega_{40} \) drops as cluster size increases, while at the same time the number of states below 6 eV increases. By the time the cluster size reaches \( d = 4.5 \) Å [uracil–(H2O)\( _3 \)], the first 40 excited states reach only 5.65 eV, well below the energy of the second absorption band. At these energies, the density of excited states is \( \approx 60 \) states/eV, and using Q-Chem on a machine with 4 Gb of memory, we are unable to calculate enough states to reach 6 eV. Excluding core orbitals from the TD-DFT excitation space (which changes the excitation energies by \( < 10^{-4} \) eV) reduces the required memory for the Davidson iterations\(^4\) by a factor of \( N_{core}/N_{occupied} \approx 0.21 \) and (just barely) allows us to calculate the 59 states that are required to reach 6 eV, by which point the density of states has reached \( \approx 80 \) states/eV. (For comparison, multireference calculations of gas-phase uracil find a total of eight \( n\pi^* \) and \( \pi\pi^* \) states in the 5.0–7.0 eV range.)\(^4\)

The results in Table 2 are for PBE0, but B3LYP paints a similar picture (with even a slightly larger number of spurious CT states, consistent with its slightly smaller fraction of Hartree–Fock exchange). We conclude that, despite their success for acetone in liquid water,\(^5\) in certain systems the popular hybrid functionals B3LYP and PBE0 may still suffer from considerable CT contamination at or below the lowest valence excitation energies. Whereas a \( \pi\pi^* \) bright state ought to be either the first or second excited state (depending on the order of the \( n\pi^* \) and \( \pi\pi^* \) states, which changes as a function of cluster size and geometry), we see from Table 2 that clusters as small as uracil–(H2O)\( _3 \) exhibit spurious states below the first bright state. Apparently, the “long range” CT problem in TD-DFT can manifest even at hydrogen-bond distances, and even when using hybrid functionals with up to 25\% Hartree–Fock exchange. That said, it should be emphasized that the problem is dramatically worse for nonhybrids—a BLYP calculation on the \( d = 2.5 \) Å cluster, for example, yields more than 40 states below 6 eV, even though there are only seven water molecules, while at \( d = 3.0 \) Å, CT states appear starting at 2.85 eV and the first \( \pi\pi^* \) state is not even among the first 40 excited states!

In further contrast to the case of acetone in water, where no significant hybridization is observed between the water molecules and the acetone lone pairs,\(^5\) we do observe hybridization between water and the carbonyl lone pairs of uracil. Consequently, the real \( n\pi^* \) states (and sometimes even \( \pi\pi^* \) states with some \( n\pi^* \) character) are sometimes difficult to discern from the spurious CT states simply on the basis of the TD-DFT excitation amplitudes and Kohn–Sham molecular orbitals (MOs). Such ambiguity is avoided by instead examining electron attachment and detachment densities obtained from the eigenvectors of the difference density matrix between the ground and excited states.\(^6\) The detachment density represents the part of the density that is removed from the ground state and rearranged in the excited-state to form the attachment density.\(^2\) We make exclusive use of these densities in identifying the qualitative character of the excited states.

Typical examples of low-energy CT states appearing in small uracil–water clusters are illustrated in Figure 1, while Figure 2 depicts some typical CT states in a larger cluster. In small clusters, the CT states below about 5.5 eV are almost exclusively water-to-uracil CT states of the type depicted in Figure 1(a), where the detachment density is dominated by the out-of-plane lone pair on a single water molecule. Such states appear in larger clusters as well [Figure 2(a)], where the water molecule in question tends to be located at the surface of the cluster.

---

**Table 2. Summary of TD-PBE0/6-31+G* Calculations on Uracil–Water Clusters Extracted from a Single Snapshot of an Aqueous-Phase MD Simulation**

<table>
<thead>
<tr>
<th>( \Delta d ) Å</th>
<th>no. water molecules</th>
<th>no. states below 6 eV</th>
<th>( \omega_{40} ) eV</th>
<th>state no.</th>
<th>( \omega ) eV</th>
<th>oscillator strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>0</td>
<td>5</td>
<td>9.05</td>
<td>2</td>
<td>5.33</td>
<td>0.1234</td>
</tr>
<tr>
<td>2.0</td>
<td>4</td>
<td>6</td>
<td>8.06</td>
<td>3</td>
<td>5.22</td>
<td>0.1394</td>
</tr>
<tr>
<td>2.5</td>
<td>7</td>
<td>11</td>
<td>7.46</td>
<td>5</td>
<td>5.28</td>
<td>0.1280</td>
</tr>
<tr>
<td>3.0</td>
<td>15</td>
<td>19</td>
<td>6.60</td>
<td>3</td>
<td>5.08</td>
<td>0.0807</td>
</tr>
<tr>
<td>3.5</td>
<td>18</td>
<td>20</td>
<td>6.50</td>
<td>4</td>
<td>5.18</td>
<td>0.0709</td>
</tr>
<tr>
<td>4.0</td>
<td>25</td>
<td>29</td>
<td>6.22</td>
<td>9</td>
<td>5.08</td>
<td>0.0190</td>
</tr>
<tr>
<td>4.5</td>
<td>37</td>
<td>59</td>
<td>5.65</td>
<td>18</td>
<td>5.06</td>
<td>0.1353</td>
</tr>
</tbody>
</table>

---

\( ^a \) Distance threshold for selecting water molecules. \( ^b \) Excitation energy of the 40th state above the ground state. \( ^c \) Indicates where the state appears in the TD-DFT excitation manifold.

---

**Figure 1.** Typical examples of spurious CT excitations in small uracil–water clusters: (a) water-to-uracil CT and (b) uracil-to-water CT. Each excitation may be conceptualized as a rearrangement of the electron detachment density on the left into an attachment density on the right.
The appearance of these states is easy to understand. First note that the out-of-plane lone pairs on the water molecules (except possibly those at the center of a large cluster) are the highest occupied MOs (HOMOs) in the system, while the lowest unoccupied MO (LUMO) is always a uracil \( \pi^* \) orbital. In the limit of large separation between an occupied and a virtual MO, and absent any component of Hartree–Fock exchange, TD-DFT will predict CT between these orbitals at an excitation energy equal to the difference in their Kohn–Sham eigenvalues.\(^{14,21} \) Thus, if any frontier occupied MOs are spatially separated from low-lying virtual MOs, then one will obtain spurious, low-energy CT excitations, unless a large component of Hartree–Fock exchange (greater than 25%, evidently) is employed. Such states should be anticipated in most condensed-phase systems.

In addition to the uracil \( \pi^* \) LUMO, larger uracil–water clusters also possess low-lying virtual MOs localized on the solvent that are not present in small clusters. This opens up another avenue to spurious CT in large clusters and affords water-to-water CT excitations such as that depicted in Figure 2(b). Occasionally these states have some uracil-to-water CT character as well, but mostly the uracil-to-water CT states [e.g., Figure 1(b)] appear at energies above 5.5 \( \text{eV} \), though they proliferate rapidly at higher excitation energies.

The small-cluster CT states are intriguing, because a cluster like uracil–(\( \text{H}_2\text{O} \))\(_7\) might not immediately come to mind upon mention of “long-range” CT in TD-DFT. To emphasize that small clusters are indeed susceptible to CT contamination, we examine in detail the excited states of the \( d = 2.5 \) \( \text{Å} \) cluster, of which there are six within 5.5 \( \text{eV} \) of the ground state. Detachment densities for these six states are depicted in Figure 3, along with excitation energies (\( \omega_i \)) and oscillator strengths (\( f_i \)). Attachment densities are not shown, as each one is dominated by the LUMO and resembles the attachment density shown in Figure 1(a). The detachment densities identify states 3 and 5 as the first \( n\pi^* \) and \( \pi\pi^* \) states, respectively, whereas the remaining states below 5.5 \( \text{eV} \) involve water-to-uracil CT of the type discussed above.

With regard to the oscillator strengths, we note that the \( n\pi^* \) state borrows sufficient intensity to achieve an oscillator strength 25% as large as that of the nominal bright state, whereas the CT excitations are mostly dark, consistent with nearly nonoverlapping attachment and detachment densities. In system configurations where \( n\pi^* /\pi\pi^* \) intensity borrowing is less significant, however, oscillator strengths for the low-energy CT states sometimes exceed that of the \( n\pi^* \) state. Thus the real dark states cannot be identified simply from a list of excitation energies and oscillator strengths, but only by careful analysis of the MOs or (better yet) attachment/detachment densities.

In larger clusters, however, CT states can undergo a type of ersonal intensity borrowing that greatly complicates interpretation of the vertical excitation spectrum. A hint as to this behavior is the overall decrease in the oscillator strength of the first \( \pi\pi^* \) excited state as a function of cluster size (see Table 2), though the trend is not monotonic—the \( \pi\pi^* \) intensity recovers at \( d = 4.5 \) \( \text{Å} \), at least for this one particular cluster geometry. The reason for this diminished intensity is that, as the density of spurious CT states increases, there appear CT states with energies comparable to that of the \( \pi\pi^* \) state, and these spurious excitations borrow intensity from the real bright state. Since oscillator strengths out of the ground state are positive and sum to a constant (the Thomas–Reiche–Kuhn sum rule),\(^{47} \) this decreases the oscillator strength of the real bright state. (This explanation is only qualitative, since the sum rule is not exactly fulfilled within the Tamm–Dancoff approximation that we employ here.)\(^2 \)

In larger clusters, this form of intensity borrowing actually makes it difficult to determine which excitation is the real bright state. The \( d = 4.0 \) \( \text{Å} \) cluster, for example, exhibits five excited states between 5.05 and 5.20 \( \text{eV} \) that have significant intensity (states 7–11 in the excitation manifold), which are depicted in Figure 4. With the exception of state 11 (which has the smallest oscillator strength of the five), each of the detachment densities has a significant uracil \( \pi \) component, but in all cases there is a significant contribution from a water lone pair as well. All five of the attachment densities are dominated by the uracil \( \pi^* \) LUMO. State 9 is selected as the \( \pi\pi^* \) state in Table 2 because its TD-DFT eigenvector contains a larger component of the uracil \( \pi \rightarrow \pi^* \) excitation than any of the other four states, but note that this is not the strongest transition of the five, as in this case the spurious CT states have borrowed the majority of the oscillator strength of the \( \pi\pi^* \) bright state. In reporting a vertical excitation spectrum, then, it is not appropriate simply

**Figure 2.** Typical examples of spurious CT excitations in a uracil–(\( \text{H}_2\text{O} \))\(_7\) cluster: (a) water-to-uracil CT and (b) water-to-water CT.

**Figure 3.** Excitation energies, oscillator strengths, and detachment densities for the lowest six TD-PBE0/6-31+G* excited states of a uracil–(\( \text{H}_2\text{O} \))\(_7\) cluster.
to report the transition with the largest oscillator strength as “the” bright state.

B. QM/MM Simulations of Aqueous Uracil. In a recent TD-PBE0 study of hydrated uracil,\textsuperscript{48} it was found that a uracil–(H\textsubscript{2}O)\textsubscript{4} complex embedded in a polarizable continuum induces a redshift of only 0.1 eV in the first uracil $\pi - \pi^*$ excitation, whereas the experimentally measured solvatochromatic shift is about 0.5 eV. In fact, the polarizable continuum accounts for the entirety of the calculated shift; the four explicit water molecules do not modify the gas-phase excitation energy at all.\textsuperscript{48} (A recent TD-BLYP study of $s$-tetrazine in aqueous solution also found that those water molecules that are directly hydrogen-bonded to the chromophore do not suffice to explain the observed solvatochromatic shift.)\textsuperscript{8} As a next step, it seems natural to consider QM/MM simulations of aqueous uracil, using a QM region substantially larger than uracil–(H\textsubscript{2}O)\textsubscript{4}. Such calculations are discussed in the present section.

To make comparison with results in the previous section, we first consider a sequence of calculations whose QM regions are precisely the same series of increasingly large uracil–water clusters described in Table 2 of section IIIA. The MM region in these new calculations consists of all additional water molecules extracted from our MD simulation, out to a distance of 20.0 Å away from uracil. These MM water molecules (about 2300 in all) are incorporated as TIP3P point charges. Table 3, which is analogous to Table 2 in the previous section, summarizes the results of TD-PBE0/6-31+G* calculations on these QM/MM systems.

Addition of the MM solvent region has a very small effect on the excitation energy for the first $\pi - \pi^*$ state, inducing a shift of no more than 0.07 eV, even in cases where the QM region consists only of uracil, or of uracil plus only a few water molecules. There is also no clear trend in the direction of this shift.

The MM solvent region does have one tremendously important effect, however: it dramatically reduces the number of spurious CT states at all values of $d$, the distance threshold for selecting QM water molecules. With the addition of point charges, even a large QM region like $d = 4.5$ Å (which is 13–14 Å across, and contains 123 atoms) affords only 10 excited states within 6 eV of the ground state. Absent the TIP3P charges, the same QM region affords an estimated 60 states below 6 eV. Figure 5(a) plots the growth in the number of low-energy excited states as a function of $d$, for TD-PBE0 calculations with and without MM point charges. For gas-phase clusters the number of states rises rapidly with cluster size, but this growth is very sluggish in the presence of MM point charges.

Although MM point charges eliminate many low-energy CT states (for reasons explained below), Figure 5(b) reveals a steady decrease in $\omega_{40}$ as a function of $d$, even for the QM/MM calculations, though the falloff is sharper in the absence of point charges. The decrease in $\omega_{40}$ indicates that the MM charges do not remove all spurious CT states, especially at higher excitation energies. While the QM/MM calculation at $d = 4.5$ Å yields only 10 states below 6 eV, there are another 30 states (mostly spurious) between 6.00 and 6.91 eV. On the other hand, a far greater number of spurious states appear in this energy regime when the MM charges are removed, and this is an important practical consideration, given that the number of excited states requested in a TD-DFT calculation determines the memory required for the Davidson iterations.\textsuperscript{44} In fact, we are unable to locate the second electronic absorption band in gas-phase uracil–(H\textsubscript{2}O)\textsubscript{37} ($d = 4.5$ Å) due to the large number of states required. The total memory requirement for such a calcula-

\begin{table}[h]
\centering
\begin{tabular}{cccc}
\hline
$d$/Å & no. QM water molecules & no. states below 6 eV & $\omega_{40}$/eV & state no. & $\omega$/eV & oscillator strength \\
\hline
1.5 & 0 & 3 & 9.73 & 2 & 5.31 & 0.1396 \\
2.0 & 4 & 5 & 8.66 & 2 & 5.25 & 0.1168 \\
2.5 & 7 & 5 & 8.16 & 2 & 5.22 & 0.1707 \\
3.0 & 15 & 8 & 7.59 & 2 & 5.15 & 0.1616 \\
3.5 & 18 & 7 & 7.44 & 1 & 5.13 & 0.1041 \\
4.0 & 25 & 8 & 7.15 & 1 & 5.09 & 0.1422 \\
4.5 & 37 & 10 & 6.91 & 1 & 5.10 & 0.1624 \\
\hline
\end{tabular}
\caption{Summary of TD-PBE0/6-31+G* QM/MM Calculations on Aqueous Uracil, as a Function of the Size of the QM Region}
\end{table}

\textsuperscript{a} Distance criterion for selecting the QM region.
molecules, lowering their Kohn–Sham eigenvalues and thereby increasing the excitation energy associated with water-to-uracil CT. Importantly, this stabilization is sufficient to remove low-energy CT states only in conjunction with a hybrid functional; TD-BLYP calculations are still beset by numerous CT states at low energies, even within a QM/MM framework.

To emphasize this point, Table 4 summarizes TD-BLYP calculations on our sequence of uracil–water clusters, both with and without point charges. (These data are plotted alongside TD-PBE0 results in Figure 5.) Although the MM solvent does reduce the number of states below 6 eV, the number of such states remains large, even in the QM/MM calculations. Using BLYP, attempts to locate the second absorption band quickly become intractable as cluster size increases.

These observations clarify the results of Bernasconi, Sprik, and Hutter,5,6 who simulated electronic absorption spectra of aqueous acetone using plane-wave Carr–Parrinello MD. Using BLYP, these authors find that the lowest valence \( n\pi^* \) band is buried beneath a much broader and more intense CT band, comprised of several spurious CT states. In contrast, the B3LYP and PBE0 functionals shift the CT band upward by \( \sim 1 \) eV, well above the valence \( n\pi^* \) band.6 Our results show that this is only partially attributable to the use of hybrid functionals. Equally important is the fact that Bernasconi et al. use plane-wave DFT (hence periodic boundary conditions), which means that there were no “surface” water molecules present that might contribute low-energy CT excitations. On the other hand, Hartree–Fock exchange is incredibly expensive to evaluate in a plane-wave basis. Gaussian-orbital-based electronic structure theory, in conjunction with MM point charges to model an extended solvent network, thus represents a useful, affordable alternative.

C. Electronic Absorption Spectra. To this point, all calculations have used geometries taken from the same MD snapshot, which allows us to discuss trends with respect to cluster size. Solvent and chromophore geometry, however, play important roles in modulating the excitation energies, modifying the order and relative intensities of the valence excitations at least, and possibly the CT excitations as well. In order to take these effects into account, we next discuss electronic absorption spectra simulated as averages over a

<table>
<thead>
<tr>
<th>( d/\text{Å} ) ( \Delta )</th>
<th>( \omega_{\text{states}} ) ( \Delta )</th>
<th>( \omega_{40}/\text{eV} ) ( \Delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>no</td>
<td>12</td>
</tr>
<tr>
<td>2.0</td>
<td>no</td>
<td>28</td>
</tr>
<tr>
<td>2.5</td>
<td>no</td>
<td>&gt;40</td>
</tr>
<tr>
<td>3.0</td>
<td>no</td>
<td>&gt;40</td>
</tr>
<tr>
<td>1.5</td>
<td>yes</td>
<td>7</td>
</tr>
<tr>
<td>2.0</td>
<td>yes</td>
<td>17</td>
</tr>
<tr>
<td>2.5</td>
<td>yes</td>
<td>26</td>
</tr>
<tr>
<td>3.0</td>
<td>yes</td>
<td>&gt;40</td>
</tr>
</tbody>
</table>

\( \Delta \) Denotes the size of the cluster or the size of the QM region.  
\( \Delta \) Point charges were used in some calculations to represent additional water molecules out to \( d = 20.0 \) Å.
total of 100 configurations extracted from an MD simulation at intervals of 1.0 ps. (A 2-fold reduction in the number of configurations has a modest effect on the spectra, but the rough spectral envelopes appear to be converged with respect to configuration sampling.) As with the QM/MM calculations discussed above, water molecules near the uracil (according to criteria described below) are included in the QM region, while additional water molecules out to 20.0 Å are incorporated as TIP3P point charges.

Absorption spectra are obtained by constructing histograms of the TD-DFT excitation energies (using a bin width of 0.1 eV), wherein the excitations are summed according to their oscillator strengths; these spectra are plotted, at both the TD-B3LYP and TD-PBE0 levels, in Figure 6. Electronic densities of states (also plotted in the figure) are obtained in similar fashion, by assigning equal weight to each excited state. Once again PBE0 predicts slightly fewer low-energy CT states than B3LYP, which is also consistent with the aforementioned multireference calculations, which find a total of six nπ* and ππ* states in the 5.5–7.0 eV range.45

Examining next the results for the microhydrated QM region, Figure 6(c),(d), we see that the density of states within the first absorption band is largely unchanged and, importantly, decays nearly to zero around 5.5 eV, in between the first and second absorption bands. (The experimental spectrum also decays nearly to zero around 5.5 eV,40 Unlike the case of a uracil-only QM region, however, the density of states shows no sign of dropping in the tail of the second absorption band and, in the B3LYP case at least, appears to be increasing above 6.5 eV, even as the spectral intensity decays. Embedding the QM region in an MM solvent pushes the CT threshold up to about 6.0 eV, with many more spurious states above 6.5 eV.

Finally there is the QM region consisting of a full solvation shell, Figure 6(e),(f). Here, the threshold for observing a substantial number of CT states creeps down somewhat from the 6.0 eV observed above, and consequently the density of states no longer decays to zero at 5.5 eV. In addition, a small number of system configurations exhibit CT states around 4.5 eV, below the first absorption band.

Regarding the solvatochromatic shifts, we note that gas-phase TD-B3LYP/6-31+G* and TD-PBE0/6-31+G* calcu-
that summing the oscillator strengths (as opposed, say, to finding the configurationally averaged excitation energy of the most intense transition) gathers up the intensity of any CT states that may borrow intensity from the $\pi\pi^*$ state.

D. Truncation of the TD-DFT Excitation Space. Although an MM embedding provides a simple and affordable means to reduce CT contamination (and for large systems is a method of choice in its own right), in some cases one might be interested in a gas-phase cluster rather than a proper liquid. In this section, we examine a separate (though compatible) procedure, whereby CT states are eliminated by ansatz, by removing from the linear-response eigenvalue equation those occupied-to-virtual $(i|\rightarrow|\alpha)$ excitation amplitudes that correspond to long-range CT. Automated criteria for performing this truncation of the excitation space have been developed by Besley,9 whose procedure we adopt here. Truncation of the excitation space has also been explored, within the context of plane-wave DFT, by Odelius et al.8

According to Besley’s procedure,9 one first identifies a subset of the atoms as belonging to the chromophore, for which we choose all of the uracil atoms. Excitation amplitudes $x_\alpha$ are then removed unless the occupied Kohn–Sham orbital $|\alpha\rangle$ contains a significant contribution from basis functions centered on chromophore atoms, as measured by the contribution that these basis functions make to the Mulliken population of $|\alpha\rangle$. We denote the threshold contribution as $\kappa_{occ}$; if the chromophore-centered basis functions do not contribute at least $\kappa_{occ}$ electrons to $|\alpha\rangle$, then the $x_\alpha$ are omitted, for all $\alpha$. (Values given for $\kappa_{occ}$ in this work are total populations, including both spins.)

Besley’s9 suggests additional truncation based on a second threshold $\kappa_{virt}$, according to which the sum of squares of the MO coefficients, $\sum_i|\mu_i|^2$, is used to measure the contribution that the set of chromophore-centered basis functions $\{|\mu_i\rangle\}$ makes to the virtual orbital $|\alpha\rangle$. As a result of the diffuse functions present in our basis set, however, we find that this sum is quite similar for each of the low-lying virtual MOs, whether or not they are localized around the uracil molecule. One way to circumvent this problem is to employ a mixed basis set, eliminating diffuse functions on the chromophore so that uracil-centered basis functions no longer contribute significantly to virtual MOs localized on the solvent. In practice, we find that useful results can be obtained without any truncation of the virtual space, so we retain 6-31+G* for all atoms and truncate the excitation space based solely on the occupied orbital criterion.

Table 5 lists TD-PBE0 excitation energies for the first and second $\pi\pi^*$ states of two different uracil–water clusters, using several different values of $\kappa_{occ}$ ranging from $\kappa_{occ} = 0$ (a full excitation space) to $\kappa_{occ} = 1$ (the value used in Besley’s benchmark calculations).9 Results are presented both with and without MM point charges. In the smaller of these two clusters, uracil–$\text{(H}_2\text{O})_7$ ($d = 2.5$ Å), we find that the accuracy of the excitation energies degrades rather slowly as a function of $\kappa_{occ}$. Already at $\kappa_{occ} = 0.2$, all CT states below the second $\pi\pi^*$ state are eliminated, meanwhile no significant error is incurred in the $\pi\pi^*$ excitation energies. Errors of $\pm 0.04$ eV were also reported by Besley9 using $\kappa_{occ}$.
= 1.0 and $\kappa_{\text{virt}} = 0.8$ for a formamide–(H$_2$O)$_4$ cluster; larger clusters were not considered in that study.

The larger of the two clusters is uracil–(H$_2$O)$_{37}$ ($d = 4.5$ Å), and in this case the valence excitation energies are more sensitive to the value of $\kappa_{\text{occ}}$. Even $\kappa_{\text{occ}} = 0.2$ engenders errors of 0.1 and 0.2 eV, respectively, in the first and second $\pi\pi^*$ excitation energies. As before, $\kappa_{\text{occ}} = 0.2$ is sufficient to remove the water-to-uracil CT states; therefore, further increase of $\kappa_{\text{occ}}$ is of no benefit. The spurious states that remain involve water-to-water and uracil-to-water CT, and elimination of these states would require truncation based on the virtual orbitals.

To assess the accuracy of truncation over a range of geometries, we recalculate the optical spectrum for the large (full solvation shell) QM/MM calculations, using a truncation threshold of $\kappa_{\text{occ}} = 0.2$, and in Figure 8(a) we compare this spectrum to that obtained using a full excitation space.

Truncation produces virtually no change in the overall absorption envelope, except that it shifts the entire spectrum to higher energy. The magnitude of this overall shift is something less than the bin width of the histogram, 0.1 eV. (Test calculations on smaller systems indicate that additional configurations are required in order to achieve better than 0.1 eV resolution.)

We conclude that truncation affords a consistent level of accuracy across many system configurations.

The density of states for this calculation, Figure 8(b), shows that truncation does not remove any CT states within the first absorption band—these were removed already by the introduction of MM point charges. As the excitation energy increases, however, calculations in the full excitation space predict an increasingly large number of spurious states, relative to results obtained with $\kappa_{\text{occ}} = 0.2$. The practical upshot is that the latter calculations consistently require only 15 excited states to reach 6.8 eV (the energy cutoff in Figure 8), whereas 30–40 states are required when a full excitation space is employed. This represents an approximately 2-fold reduction in the memory required for the Davidson iterations, which is roughly proportional to the number of excited states requested.

Truncation of the excitation space is similarly accurate for gas-phase clusters, as demonstrated when we remove the MM point charges from the QM/MM calculations discussed above. Figure 9(a) compares spectra obtained with $\kappa_{\text{occ}} = 0.2$ to those calculated with a full excitation space, while Figure 9(b) compares the corresponding densities of states.

In the absence of truncation, we obtain excited states all the way down to 3 eV, and 40 excited states are required just to reach $\omega = 5.4$ eV. Thus the spectra in Figure 9(a) include only the first absorption band. When $\kappa_{\text{occ}} = 0.2$, this energy cutoff is reached consistently with only the first 25 excited states. As before, the spectrum calculated with the truncated excitation space is shifted to higher energy by $\leq 0.1$ eV, with little change in the overall absorption envelope.

Finally, we note that the success of orbital truncation as a means to reduce CT contamination is contingent upon use of a hybrid density functional. Even for a fairly small, $d = 3.0$ Å cluster, with MM point charges included and using a truncation threshold of $\kappa_{\text{occ}} = 0.2$, a TD-BLYP calculation yields 17 excited states below the first $\pi\pi^*$ state. Keeping $\kappa_{\text{occ}} = 0.2$ but omitting the point charges, CT becomes so prevalent that it is impossible to discern the identity of the $\pi\pi^*$ state simply on the basis of oscillator strengths, as so much of this intensity has bled into the spurious CT states.
embedding removes most low-energy CT states, even when the QM region is rather large (> 120 atoms), but works only in conjunction with hybrid functionals. Although the same effect can be achieved without MM point charges by using plane-wave DFT with periodic boundary conditions, the plane-wave calculations are prohibitively expensive because hybrid functionals are still required.

As an alternative to, or in conjunction with, a QM/MM embedding, CT states can also be eliminated by removing certain excitation amplitudes from the TD-DFT linear response equations, according to an automated procedure. This procedure must be used with extreme caution, as it eliminates CT states (real or spurious) by ansatz. In cases where no real CT is expected, however, this technique significantly reduces the number of spurious states while introducing errors in the valence excitation energies that are typically smaller than 0.1 eV, at least for the examples considered here. Once again, the success of this technique is contingent upon use of a hybrid functional. Using BLYP, serious CT contamination persists, despite either of the aforementioned measures.

Finally, we note that the aforementioned procedures are intended only to remove CT “contamination”, that is, the appearance of spurious CT states at low energies. Where real CT states are present (whose energies will of course be grossly underestimated by standard TD-DFT), the prescribed truncation of the excitation manifold will eliminate these as well. Electrostatic embedding, on the other hand, will modulate the energetics of real CT states, but it cannot be expected to compensate for the fundamentally incorrect way in which these states are described by contemporary TD-DFT. The simple procedures described here are therefore most applicable to studies of optically bright, valence excitations in large molecular systems.

Acknowledgment. This work was supported by grant no. PAS0291 from the Ohio Supercomputer Center and by start-up funds from The Ohio State University.

References
