Assessment of the Ab Initio Bethe–Salpeter Equation Approach for the Low-Lying Excitation Energies of Bacteriochlorophylls and Chlorophylls

Zohreh Hashemi and Linn Leppert*

Cite This: J. Phys. Chem. A 2021, 125, 2163–2172

ABSTRACT: Bacteriochlorophyll and chlorophyll molecules are crucial building blocks of the photosynthetic apparatus in bacteria, algae, and plants. Embedded in transmembrane protein complexes, they are responsible for the primary processes of photosynthesis: excitation energy and charge transfer. Here, we use ab initio many-body perturbation theory within the GW approximation and Bethe–Salpeter equation (BSE) approach to calculate the electronic structure and optical excitations of bacteriochlorophylls and chlorophylls. We systematically study the effects of the structure, basis set size, partial self-consistency in GW, and the underlying exchange–correlation approximation and compare our calculations with results from time-dependent density functional theory, multireference RASPT2, and experimental literature results. We find that optical excitations calculated with GW+BSE are in excellent agreement with experimental data, with an average deviation of less than 100 meV for the first three bright excitations of the entire family of (bacterio)chlorophylls. Contrary to state-of-the-art time-dependent density functional theory (TDDFT) with an optimally tuned range-separated hybrid functional, this accuracy is achieved in a parameter-free approach. Moreover, GW+BSE predicts the energy differences between the low-energy excitations correctly and eliminates spurious charge transfer states that TDDFT with (semi)local approximations is known to produce. Our study provides accurate reference results and highlights the potential of the GW+BSE approach for the simulation of larger pigment complexes.

INTRODUCTION

Electronic excitations form the foundation of some of the most fundamental natural processes. In photosynthesis, plants, algae, and bacteria convert solar energy into chemical energy, utilizing a cascade of coupled energy and charge transfer excitations that are performed by pigment–protein complexes with high quantum efficiency. Bacteriochlorophyll (BCL) and chlorophyll (CL) molecules are among the most important building blocks of these pigment–protein complexes. They are responsible for the absorption and transfer of excitation energy and for the charge separation necessary for establishing a proton gradient that eventually drives the synthesis of chemical energy in plants and bacteria. Accurately calculating the electronic structure and excitations of these molecules from first principles is the prerequisite for understanding their interactions with each other and with the surrounding proteins and, consequently, energy and charge transfer in natural photosynthesis. BCL and CL molecules constitute a family of substituted tetrapyrroles with varying absorption properties depending on conjugation and the number and nature of substitutions. CL a and b are present in plants and green algae, whereas green bacteria mostly rely on BCL c, d, and e for excitation energy transfer and BCL a for concentrating excitations close to the reaction center of the photosynthetic unit. BCL a is also the main pigment in purple bacteria, whose light harvesting apparatus and reaction center are among the most thoroughly studied natural light-harvesting systems. The optical excitation spectrum of these pigments possesses two characteristic absorption bands: (1) the Q band in the visible part of the spectrum, composed of excitations Qa and Qb with high- and low-oscillator strength, respectively and (2) the B (or Soret) band in the near ultraviolet.

In the field of finite organic and biological molecular systems, neutral excitations and optical spectra are predominantly calculated using time-dependent density functional theory (TDDFT). In conjunction with model Hamiltonian...
approaches, TDDFT has been employed for the simulation of large photosynthetic pigment–protein complexes. The accuracy of its approximations and implementations has been tested for a variety of biochromophores. However, TDDFT’s standard approximations are inadequate for describing long-range charge transfer excitations and high-energy Rydberg states due to self-interaction errors and an incorrect asymptotic behavior. Exchange–correlation (xc) functionals that contain long-range exact exchange, such as optimally tuned range-separated hybrid functionals (OT-RSH), can be employed as a remedy in such cases, but the use of such functionals requires a tedious per-system tuning procedure.

Multireference wavefunction-based methods have scarcely been used for molecules as large as BCL and CL. Vertical excitation energies of CL, based on ADC(2) and different coupled cluster approaches show a spread of ~0.4 eV, strongly depending on the method, basis set, and structural model used in these calculations. In 2016 and 2019, Anda et al. reported multistate RASPT2/RASSCF excitation energies of several BCL units within the light-harvesting system 2 of a purple bacterium. A RASPT2 approach was also combined with electrostatic embedding of fixed-point charges to simulate the effect of the protein environment on excitation energies of the same system by Segatta et al. While these reports constitute important advances in the use of wavefunction-based methods for complex biological molecules, they were performed with relatively small basis sets and show a dependence on the choice of the restricted active space (RAS).

The ab initio Bethe–Salpeter equation (BSE) approach, when rigorously based on many-body Green’s function theory, is an alternative method for describing neutral excitations of correlated many-electron systems. It is based on a framework of charged excitation energies that correspond to electron addition and removal energies and that are most frequently calculated within the GW approximation. The GW+BSE approach has been shown to be successful in predicting the optical spectra of bulk solids and low-dimensional materials. In recent years, it has also begun to be applied to finite systems, such as small molecules, and larger molecular complexes, for which its accuracy has been shown to be comparable to single-reference wavefunction methods for both localized and charge transfer excitations at a substantially reduced computational cost.

In this article, we assess the accuracy of the ab initio GW+BSE approach for the Qex, Qex, and the first bright B excitation of several members of the BCL and CL family and the chemically closely related bacteriochlorin (BC) molecule. We compare two different approaches for approximating the electronic self-energy Σ = iGW: (1) G0W0, a one-shot method, in which the zeroth-order Green’s function G0 and screened Coulomb interaction W0 are constructed from a DFT eigensystem and directly used to correct DFT eigenvalues perturbatively and (2) partially self-consistent GW (evGW), in which the corrected eigenvalues are used to iteratively recalculate G and/or W until self-consistency is reached. We compare our results to TDDFT calculations with the local density approximation (LDA), two global hybrids and an OT-RSH functional, with RASPT2 literature results, and with experimental data.

We find that the GW+BSE approach used in a partially self-consistent manner results in excitation energies in the visible and near-ultraviolet within less than 100 meV from experiment for the entire family of pigments studied here. Our results are almost completely independent of the DFT eigensystem used as an input for the GW+BSE calculations. In fact, even a simple and computationally inexpensive LDA starting point leads to excellent agreement with experiment and eliminates spurious charge transfer excitations between Qex and Qex that TDDFT with (semi)local functionals produces. Contrary to TDDFT, GW+BSE also correctly predicts the energy difference between the two Q-band excitations, a crucial prerequisite for understanding the coupling of excitations in systems consisting of more than one pigment. Finally, we show that differences between evGWQex+BSE and state-of-the-art TDDFT calculations using an OT-RSH functional can be explained almost entirely based on differences in how electron–hole interactions are described by the xc kernel of TDDFT and the BSE kernel, respectively. Eigenvalue differences as computed with evGWQex and DFT with an OT-RSH functional are almost identical.

The remainder of this article is structured as follows: we start by briefly reviewing the GW+BSE approach and report computational details and numerical convergence. We then show the effect of different DFT starting points and partial self-consistency on the excitation energies of the BC molecule. After this, we discuss our results for BCL a, b, c, d, and e and CL a and b, followed by a comparison with literature results and an in-depth discussion of differences between our GW+BSE and TDDFT results for BCL a.

### METHODS

#### GW+BSE Approach

In Green’s function-based many-body perturbation theory, the calculation of charged excitations, corresponding to electron removal and addition energies, is based on knowledge of the exact interacting single-particle Green’s function G, which can, in principle, be computed from a set of self-consistent integro-differential equations—Hedin’s equations—linking G to the electronic self-energy Σ, the screened Coulomb interaction W, the irreducible polarizability γ, and the vertex function Γ. The lowest-order expansion of Σ with respect to W leads to the GW approximation, in which the electronic self-energy Σ = iGW. Quasiparticle (QP) eigenvalues can be obtained by solving

\[ -\frac{\hbar^2}{2m}V^2 + V_{\text{hom}}(r) + V_{\text{H}}(r)\phi_{\text{QP}}(r) + \int d^3r'\Sigma(r, r'; \epsilon_n^{\text{QP}})\phi_{\text{QP}}(r') = \epsilon_n^{\text{QP}}\phi_{\text{QP}}(r) \]

where V_{\text{hom}} is the ionic potential, V_{\text{H}} is the Hartree potential, and \( \epsilon_n^{\text{QP}} \) and \( \phi_{\text{QP}} \) are QP energies and wavefunctions, respectively.

To avoid the high computational cost of a self-consistent solution of eq 1, the GW approach is commonly used within a one-shot scheme, in which \( G_0 \) and \( W_0 \) are constructed from a (generalized) Kohn–Sham (gKS) eigensystem obtained from a preceding DFT calculation. We use the notation \( G_0W_0\text{gKS} \) to refer to \( G_0W_0 \) based on the gKS eigensystem \( (\epsilon_n^{\text{gKS}}, \phi_n^{\text{gKS}}) \) computed with the xc functional \( E_{\text{xc}}^{\text{gKS}} \). In this approach, QP corrections are calculated to first order in \( \Sigma \) as

\[ \epsilon_n^{\text{QP}} = \epsilon_n^{\text{gKS}} + \langle\phi_n^{\text{gKS}}|\Sigma|\phi_n^{\text{QP}}\rangle + V_{\text{xc}}\rho_n^{\text{gKS}} \]

where \( \rho_n \) is the xc potential, and it is assumed that \( \rho_n^{\text{QP}} \approx \rho_n^{\text{gKS}} \).

While the \( G_0W_0\text{gKS} \) approach has been used with much success, in particular, for the calculation of band gaps and band structures of solids, a well-known and well-documented dependence on the gKS eigensystem used to construct \( G_0 \).
and $W_0$ limits its predictive power.\textsuperscript{34–36} Partial self-consistency in the QP eigenvalues can often mitigate this problem. In eigenvalue self-consistent $GW$, the $g$KS eigenvalues used to construct $G$ and/or $W$ are replaced with those from the output of a prior $GW$ step; the self-energy corrections are then iterated until the QP eigenvalues converge. This approach, which we call ev$G_0W_n$ in the following ($n$ refers to the number of iterations), has been shown to remove much of the starting point dependence for a range of different systems.\textsuperscript{37,38}

The BSE is an equation for the two-particle electron–hole Green’s function and allows for the calculation of the polarizability including electron–hole interactions through the screened Coulomb interaction $W$. In practice, the BSE is usually solved by neglecting the frequency dependence of $W$. Within this static approximation, it can be written in a form equivalent to Casida’s equations of TDDFT

$$\begin{pmatrix} A & B \\ -B & -A \end{pmatrix} \begin{pmatrix} X' \\ Y' \end{pmatrix} = \Omega \begin{pmatrix} X' \\ Y' \end{pmatrix}$$

\hspace{1cm} (3)

where $\Omega$ represents neutral excitation energies and $(X', Y')$ the corresponding eigenvectors.\textsuperscript{19} $A$ and $-A$ represent resonant and antiresonant transitions that can be expressed as

$$A_{ij}^{ab} = \left( \epsilon_i^{QP} - \epsilon_j^{QP} \right) \delta_{ij} \delta_{ab} - 2 (a|b) + W_{ij}^{ab}(\omega = 0)$$

\hspace{1cm} (4)

and that are coupled through $B$ and $-B$, defined as

$$B_{ij}^{ab} = -2 (a|b) + W_{ij}^{ab}(\omega = 0)$$

\hspace{1cm} (5)

for singlet excitations. In these expressions, $i$ and $j$ are occupied and $a$ and $b$ are unoccupied states and $(i|a|j|b)$ stands for

$$\int d\mathbf{r} d\mathbf{r}' \phi_i^{QP}(\mathbf{r}) \phi_j^{QP}(\mathbf{r}') \frac{1}{|\mathbf{r} - \mathbf{r}'|} \phi_a^{QP}(\mathbf{r}) \phi_b^{QP}(\mathbf{r}')$$

\hspace{1cm} (6)

Note that $q_i^{QP} = q_i^{gKS}$, whenever the $G_0W_0$ or ev$G_0W_n$ approaches are used to construct $A$ and $B$.

**Computational Details.** Our calculations of charged and neutral excitations were performed using the $GW$+BSE and TDDFT implementation in the open-source MOLGW software package (version 2B), which relies on Gaussian basis functions.\textsuperscript{39} We used the frozen-core approximation throughout, which changes excitation energies by less than 1 meV. We also employed the resolution-of-the-identity (RI) method, in order to reduce the calculation of four-center integrals to two- and three-center integrals. For BCL $a$, the RI changes the QP highest occupied molecular orbital (HOMO)–lowest unoccupied molecular orbital (LUMO) gap by less than 50 meV using a 6-31G basis set and BHLYP as a starting point, but we expect the effect of the RI to be even smaller for the larger basis sets used in the remainder of this article.\textsuperscript{40} To further reduce the computational cost of the evaluation of the $GW$ polarizability, we use the single-pole approximation (see the Supporting Information for details). The Tamm–Dancoff approach, which corresponds to neglecting the $B$ matrix elements in eq 3 is not used as it consistently increases both $GW$+BSE and TDDFT results by $\sim0.3$ eV, in agreement with previous findings.\textsuperscript{6,27}

We tested the influence of the Gaussian basis set size on HOMO–LUMO gaps and $Q_x$ and $Q_y$ excitations of BCL $a$ (using a structure from ref 16) with $G_0W_0$@BHLYP+BSE, considering seven different basis sets, namely, the Pople basis sets 6-31G, 6-311G, 6-311++G*, and 6-311++G(2d,2p), combined with the DeMon auxiliary basis set,\textsuperscript{31} and the

$$1 - \frac{1}{r} = \frac{1 - \text{erf}(\omega r)}{r} + \frac{\text{erf}(\omega r)}{r}$$

\hspace{1cm} (7)

Karlsruhe basis sets def2-SVP, def2-TZVP, and def2-TZVPP and their corresponding auxiliary basis sets.\textsuperscript{42}

Figures 1 and S2 show the convergence of the HOMO–LUMO gap and $Q_x$ and $Q_y$ excitation energies as a function of

![Figure 1](https://dx.doi.org/10.1021/acs.jpca.1c01240)

**Figure 1.** Convergence as a function of the number of basis functions $1/N_{basis}$ for (a) HOMO–LUMO gap and (b) $Q_x$ and $Q_y$ excitation energies, calculated with $G_0W_0$@BHLYP+BSE. Dashed lines represent a linear fit.
In turn, \( Q_y \) by up to \(-0.4\) eV, depending on the xc functional.\(^{27}\) TDDFT with the OT-RSH \( \text{oPBE} \) is in best agreement with experiment, overestimating it by \(-0.2\) eV for both excitations. We further find that \( G_0W_0@\text{LDA}+\text{BSE} \) underestimates \( Q_y \) by \(0.4\) eV and \( Q_x \) by \(0.6\) eV, whereas the use of a BHLYP and \( \text{oPBE} \) starting point results in excitations within \(0.1\) eV of the experimental results. In accordance with prior studies, we observe that most of the starting point dependence of the \( G_0W_0+\text{BSE} \) results is inherited from the starting point dependence of the HOMO–LUMO gaps.\(^{25}\)

In order to investigate the effect of eigenvalue self-consistency in the GW+BSE approach, we tested the effect of updating the eigenvalues in the construction of \( G_0 \) only \((\text{ev} G_0W_0)\) and of both \( G \) and \( W \) \((\text{ev} G_0W_n)\). Eigenvalue self-consistency in \( G_0 \) alone only slightly changes the results as compared to \( G_0W_0 \). In contrast, full eigenvalue self-consistency largely eliminates the starting point dependence. In particular, using an LDA starting point results in excitation energies within \(0.1\) eV from experiment—similar to the \( \text{oPBE} \) starting point, but at considerably reduced computational cost. In Table S4, we report similar results for the more complex pigment BCL \( a \). In the remainder of this article, we therefore focus primarily on eigenvalue self-consistent results based on LDA and \( \text{oPBE} \) starting points.

Excitation Energies of Bacteriochlorophylls and Chlorophylls. Next, we turn to reporting the vertical excitation energies of several members of the BCL and CL family of pigments. All structures were obtained from ref \(^{45}\) and geometry-optimized using DFT as implemented in the Turbomole code with a def2-TZVP basis set and the B3LYP xc functional.\(^{46}\) Atomic coordinates of all relaxed structures can be found in the Supporting Information. We used both LDA and \( \text{oPBE} \) starting points for our \( \text{ev} G_0W_n+\text{BSE} \) and \( \text{oPBE} \) for our TDDFT calculations. Unlike the \( Q_y \) excitation of BCL \( a \) and \( b \), which has significant oscillator strength, the \( Q_y \) excitation of BCL \( c-c' \) is dark. Following ref \(^8\), we therefore also compare our calculations with experimental results for the higher-energy B band.\(^{31}\) We report the vertical excitation energies and corresponding oscillator strengths of the first six excitations of all pigments in Tables S6 and S7. In these calculations, we included a total of 20 excitations, in order to ensure that the higher lying excitations are well converged.

Table 2 demonstrates that \( \text{ev} G_0W_n+\text{BSE} \) is in excellent agreement with experiment for the entire family of BCL and CL molecules. The MAE is about \(50\) meV for the \( Q_x \) and \( Q_y \) and between \(100\) and \(200\) meV for the \( B \) excitation. Our \( \text{ev} G_0W_n+\text{BSE} \) results also accurately reflect the spectral shifts of the \( Q_y \) excitation when comparing different BCL pigments with each other. For example, the BCL \( b \) molecule differs from BCL \( a \) through an ethylenide side group, which shifts the \( Q_y \) excitation by \(40\) meV to the red. This red shift is perfectly reproduced in our \( GW+BSE \) calculations. This is the first main result of this study. The second one is that our results are essentially independent of the DFT eigensystem used as an input for the \( GW+BSE \) approach: a computationally inexpensive LDA starting point results in the same level of agreement with experiment as the more tedious \( \text{oPBE} \) calculation that involves a system-dependent tuning procedure for the range separation parameter \( \omega \). This is in stark contrast to TDDFT. TD-LDA leads to spurious excitations with charge transfer character in between \( Q_x \) and \( Q_y \) as well as slightly above \( Q_y \), depending on the structure, as discussed below and in the literature.\(^{37}\) TDDFT with the optimally tuned \( \text{oPBE} \)
molecules, TDDFT tends to overestimate all three excitations of a molecule and to other benchmark studies of complex organic systems. Similar to our results for the BC molecule, TDDFT slightly overestimates the average, underestimates experiment by 100 meV, whereas the relative energy difference is close to zero, because TDDFT slightly overestimates the \( Q_x \) and \( Q_y \) excitations but underestimates the B excitation of these pigments. For all other BCL molecules and the two CL molecules studied here, we consistently find that the average deviation of \( evG_W @\text{LDA+BSE} \) is significantly smaller than that of TDDFT. Similar to our results for the BC molecule and to other benchmark studies of complex organic molecules,\(^7\) TDDFT tends to overestimate all three excitations by between 200 and 300 meV. \( evG_W @\text{LDA+BSE} \) is in much closer agreement with experiment for these pigments, on average, overestimating their excitation energies by less than 100 meV. We stress again that these results are independent of the DFT starting point, whereas our TDDFT results rely on a per-system tuning procedure.

Our results are in excellent agreement with correlated excited-state methods for those systems for which such studies have been reported, primarily CL \( a \) and BCL \( a \). ADC(2) excitation energies of the first three excitations of CL \( a \) reported by Suomivuori et al. are 1.97, 2.11, and 2.95 eV, within ~0.1 eV of our \( evG_W @\text{LDA+BSE} \) results.\(^8\) In another study by the same authors, the ADC(2) \( Q_x \) excitation energy of histidin-ligated BCL \( a \) was reported to be 1.46 eV, again within ~0.1 eV of our results, although it should be noted that the structures of ligated and free-standing BCL \( a \) slightly differ, leading to excitation energy differences of 10–30 meV at the ADC(2) level.\(^9\) Furthermore, Sirohiwal et al. used a pair-natural orbital coupled cluster approach to study CL \( a \) and reported \( Q_x \) and \( Q_y \) excitation energies of 1.75 and 2.24 eV, respectively, for CL \( a \), also within ~0.1 eV of our \( GW+BSE \) results for these excitations.\(^9\)

Not only the absolute energies of \( Q_x \) and \( Q_y \) excitations are important for understanding and predicting excitation energy and charge transfer in photosynthetic systems but also their relative energy difference, \( \Delta Q_x - Q_y \), plays a role, in particular, in coupled systems of several pigment units. It is therefore reassuring that \( evG_W @\text{LDA+BSE} \) predicts \( \Delta Q_x - Q_y \) in very good agreement with experiment, with a deviation of only 10 meV for BCL \( a \), BCL \( b \) and CL \( a \) and 120 meV for CL \( b \) for the LDA starting point and a slightly larger deviation of, on average, 60 meV for the \( \omega \text{PBE} \) starting point. TDDFT based on \( \omega \text{PBE} \) tends to underestimate \( \Delta Q_x - Q_y \), by, on average, 110 meV for these four pigments. For BCL \( a \), we also show in Table 3 that \( \Delta Q_x - Q_y \) strongly depends on the xc functional used in the TDDFT calculations, primarily because of the strong dependence of the \( Q_x \) excitation on the amount of exact exchange, which can be seen by comparing the results based on the LDA (0% of exact exchange), B3LYP (≈23%), and BHLYP.

Table 2. \( Q_x \), \( Q_y \), and First B Band Excitation Energy of BCLs and CLs Calculated Using a 6-311++G(2d,2p) Basis Set\(^a\)

<table>
<thead>
<tr>
<th>molecule</th>
<th>( Q_x )</th>
<th>( Q_y )</th>
<th>( B )</th>
<th>( Q_x )</th>
<th>( Q_y )</th>
<th>( B )</th>
<th>( Q_x )</th>
<th>( Q_y )</th>
<th>( B )</th>
<th>( Q_x )</th>
<th>( Q_y )</th>
<th>( B )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCL ( a )</td>
<td>1.52</td>
<td>2.08</td>
<td>3.25</td>
<td>1.50</td>
<td>2.10</td>
<td>3.16</td>
<td>1.75</td>
<td>2.16</td>
<td>3.33</td>
<td>1.60</td>
<td>2.15</td>
<td>3.46</td>
</tr>
<tr>
<td>BCL ( b )</td>
<td>1.48</td>
<td>2.07</td>
<td>2.95</td>
<td>1.45</td>
<td>2.09</td>
<td>3.05</td>
<td>1.69</td>
<td>2.15</td>
<td>3.19</td>
<td>1.56</td>
<td>2.14</td>
<td>3.37</td>
</tr>
<tr>
<td>BCL ( c )</td>
<td>1.85</td>
<td>2.05</td>
<td>2.94</td>
<td>1.84</td>
<td>2.11</td>
<td>3.02</td>
<td>2.05</td>
<td>2.21</td>
<td>3.21</td>
<td>1.88</td>
<td>2.14</td>
<td>2.89</td>
</tr>
<tr>
<td>BCL ( d )</td>
<td>1.90</td>
<td>2.15</td>
<td>2.94</td>
<td>1.89</td>
<td>2.21</td>
<td>3.05</td>
<td>2.00</td>
<td>2.29</td>
<td>3.19</td>
<td>1.90</td>
<td>2.39</td>
<td>2.93</td>
</tr>
<tr>
<td>BCL ( e )</td>
<td>2.01</td>
<td>2.04</td>
<td>2.78</td>
<td>1.96</td>
<td>2.13</td>
<td>2.88</td>
<td>2.10</td>
<td>2.23</td>
<td>3.02</td>
<td>1.92</td>
<td>2.72</td>
<td>2.72</td>
</tr>
<tr>
<td>CL ( a )</td>
<td>1.85</td>
<td>2.13</td>
<td>2.91</td>
<td>1.86</td>
<td>2.19</td>
<td>3.02</td>
<td>2.06</td>
<td>2.29</td>
<td>3.16</td>
<td>1.87</td>
<td>2.14</td>
<td>2.88</td>
</tr>
<tr>
<td>CL ( b )</td>
<td>1.95</td>
<td>2.17</td>
<td>2.85</td>
<td>1.93</td>
<td>2.20</td>
<td>2.85</td>
<td>2.10</td>
<td>2.29</td>
<td>2.97</td>
<td>1.92</td>
<td>2.26</td>
<td>2.72</td>
</tr>
<tr>
<td>MAE</td>
<td>0.05</td>
<td>0.06</td>
<td>0.12</td>
<td>0.04</td>
<td>0.05</td>
<td>0.17</td>
<td>0.17</td>
<td>0.05</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)GW+BSE results are based on eigenvalue self-consistent \( evG_W \).\(^b\)Experimental results in diethyl ether from ref 8.

Figure 2. Colored bars denote the average difference between calculated and experimental excitation energies for \( evG_W @\text{LDA+BSE} \) (blue), \( evG_W @\text{PBE+BSE} \) (red), and TDDFT (green) with \( \omega \text{PBE} \). The dots represent the maximum deviation in each case.
(50%). As before, evG,W_r+BSE is in excellent agreement with experiment and almost independent of the underlying xc functional.

The experimental results reported in Tables 2 and 3 are based on measurements in diethyl ether, whereas our calculations are for gas-phase molecules. To approximately account for the effect of the solvent, we extracted experimental reference values for Q_x and Q_y excitations from a study by Limantara et al., in which electronic absorption spectroscopy was used to obtain Q_x and Q_y for a large number of nonpolar and polar solvents at room temperature. This study reports regression lines for Q_x and Q_y excitations of BCL_a as a function of ε/2, where ε is the dielectric constant of the solvent. The extrapolated values for ε = 4.33 (vacuum) are 1.68 eV (nonpolar) and 1.67 eV (polar) for the Q_y index of the solvent. The extrapolated values for ε = 1 (vacuum) are 1.68 eV (nonpolar) and 1.67 eV (polar) for the Q_x and 2.25 eV (nonpolar) and 2.21 eV (polar) for the Q_y excitation. Based on these regression parameters, we estimate that the experimental reference values in Table 2 lie ~50–70 meV below the gas-phase excitation energies. We also calculated the Q_x and Q_y excitation energies of BCL_a with TDDFT (using ωPBE), approximating solvent effects with the COSMO approach as implemented in Turbomole. We used a dielectric constant of 4.33ε_0, corresponding to the value in diethyl ether. COSMO red-shifts the Q_x and Q_y excitation energies by 70 and 50 meV, respectively, supporting our estimate. We conclude that solvent effects are small—with the numerical accuracy of our GW+BSE calculations—and do not change our main conclusions. Note that we also neglect the effects of temperature and the 0–0 vibrational energy contribution in our comparison with experimental results. Exact agreement of our calculated results with experiment is therefore not expected.

**Bacteriochlorophyll a.** In the remainder of this paper, we will use the BCL_a molecule as a case study to compare to available computational literature results for this pigment, discuss the origin of differences between our evG,W_r+BSE and TDDFT results, and comment on the effects of the choice of structure on excitation energies.

**Comparison with RASPT2.** For the Q_x and Q_y excitations of BCL_a, we compare our GW+BSE and TDDFT calculations to multistate, second-order perturbation theory (RASPT2) calculations by Anda et al. For this comparison, we use the molecular geometry reported in ref 16, which is a BCL_a unit from the light-harvesting system LH2 of *Rhodoblastus acidophilus*. This structure was extracted from an experimental X-ray crystallographic structure of the LH2 complex (unit 302 within the structure 1NKZ in the RCSB Protein Data Bank). The phytyl tail was truncated and replaced by a hydrogen atom, and no further geometry optimization was carried out. In the following, we will call this structure “A”. Our geometry-optimized version of “A”, which we relaxed using DFT as implemented in the Turbomole code with a de2-TZVP basis set and B3LYP, will be called “R”. A visual comparison between “A” and “R” is shown in Figure 3. The large differences that we observe between these two structures are unsurprising, given that we perform our geometry optimizations without taking into account the protein environment in which BCL_a “A” is embedded in vivo. Table 3 shows our GW+BSE and TDDFT results for “A” in comparison with the RASPT2 excitation energies from refs. We find, as before, that when eigenvalue self-consistency is used in GW, HOMO–LUMO gaps and Q_x excitation energies differ by a maximum of 0.1 eV. Most notably, however, our GW+BSE excitation energies substantially differ from those calculated with RASPT2, with Q_x 0.4 eV and Q_y 0.5 eV lower than the RASPT2 result.

| Table 3. Difference between Q_x and Q_y Excitation Energies (in eV) Using TDDFT and evG,W_r+BSE for BCL_a |
|---|---|---|
| method | xc functional | ΔQ_x−Q_y (meV) |
| evG,W_r+BSE | LDA | 0.57 |
| | B3LYP | 0.54 |
| | BHLYP | 0.54 |
| | ωPBE | 0.59 |
| TDDFT | LDA | 0.25 |
| | B3LYP | 0.40 |
| | BHLYP | 0.64 |
| | ωPBE | 0.43 |
| exp^a | | 0.55 |

Figure 3. Overlay of structures “A” (red) and “R” (blue) in (a) top view and (b) side view.

**Table 4. HOMO–LUMO Gaps and Q_x and Q_y Excitation Energies (in eV) for BCL_a Structure “A”**

<table>
<thead>
<tr>
<th>method/basis set</th>
<th>xc functional</th>
<th>H−L gap (eV)</th>
<th>Q_x (eV)</th>
<th>Q_y (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>evG,W_r+BSE/6-311++G(2d,2p) LDA</td>
<td>3.62</td>
<td>1.17</td>
<td>1.90</td>
<td></td>
</tr>
<tr>
<td>B3LYP</td>
<td>3.67</td>
<td>1.19</td>
<td>1.90</td>
<td></td>
</tr>
<tr>
<td>BHLYP</td>
<td>3.72</td>
<td>1.23</td>
<td>1.92</td>
<td></td>
</tr>
<tr>
<td>ωPBE</td>
<td>3.68</td>
<td>1.16</td>
<td>1.91</td>
<td></td>
</tr>
<tr>
<td>evG,W_r+BSE/ANO-RCC-vDZP TDDFT/6-311++G(2d,2p) LDA</td>
<td>3.76</td>
<td>1.38</td>
<td>2.18</td>
<td></td>
</tr>
<tr>
<td>B3LYP</td>
<td>0.92</td>
<td>1.59</td>
<td>1.99</td>
<td></td>
</tr>
<tr>
<td>BHLYP</td>
<td>1.80</td>
<td>1.64</td>
<td>2.17</td>
<td></td>
</tr>
<tr>
<td>ωPBE</td>
<td>2.61</td>
<td>1.57</td>
<td>2.34</td>
<td></td>
</tr>
<tr>
<td>RASPT2/ANO-RCC-vDZP</td>
<td>3.70</td>
<td>1.48</td>
<td>2.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.61</td>
<td>2.40</td>
<td></td>
</tr>
</tbody>
</table>

https://dx.doi.org/10.1021/acs.jpca.1c01240
*J. Phys. Chem. A* 2021, 125, 2163–2172
wide variety of organic compounds as compared to global hybrid functionals\textsuperscript{10–12} and is more than 0.1 eV lower in energy than the RASPT2 \(Q_e\) excitation energy. Similar trends have also been shown for CL \(a\), where DFT-based multireference CI, just as TDDFT with global hybrid functionals, tends to overestimate experiment by \(\sim 0.2\) eV for the \(Q_e\) and \(Q_\epsilon\) excitation.\textsuperscript{53} All in all, given that comparisons with experimental data are complicated for an \textit{in vivo} structure such as “\(A\)”, we consider it most likely that our GW+BSE calculations underestimate the excitation energies of structure “\(A\)” by \(\sim 0.1\) eV, similar to our results for gas-phase BCL \(a\) (Table 2). The remaining deviations could be attributed to the multireference character of the \(Q_e\) excitation\textsuperscript{16} and the choice of the RAS.

We also note that our GW+BSE results reproduce the energetic order and relative energy differences of the \(Q_e\) excitation of other BCL units within the LH2 ring that RASPT2 predicts, when using the ANO-RCC-vDZP basis set. However, the use of the significantly larger 6-311++G(2d,2p) basis leads to substantially larger excitation energy differences between these units (Table S7). Finally, it is worth mentioning that our GW+BSE calculations reproduce the relatively large energy difference \(\Delta Q_\epsilon - Q_\epsilon \approx 0.8\) eV that RASPT2 predicts, whereas TDDFT excitation energy differences are much less sensitive to details of the structure, with \(\Delta Q_\epsilon - Q_\epsilon \approx 0.5\) eV (using \(\omega\text{-}PBE\)) similar to the gas-phase structure of BCL \(a\). We speculate that a geometry optimization of structure “\(A\)” within its protein environment would result in a smaller \(\Delta Q_\epsilon - Q_\epsilon\) for both RASPT2 and GW+BSE.

\textbf{Role of the Electron–Hole Kernel.} We find that the difference between GW+BSE and TDDFT excitation energies can be traced back almost entirely to differences in how electron–hole interactions are described in both schemes. The \(Q_e\) excitation is primarily (~90%) a HOMO \(	ext{\rightarrow}\) LUMO transition, and the HOMO–LUMO gaps, as calculated with DFT-\(\omega\)-PBE and \(G_{\text{\omega}}W_0\text{-}\omega\text{-}PBE\), differ by only 0.02 meV (Table 3). In fact, the density of states (DOS) in the energy range relevant for both the \(Q_\epsilon\) and the \(Q_e\) excitations based on \(G_{\text{\omega}}W_0\text{-}\omega\text{-}PBE\) and DFT-\(\omega\)-PBE eigenvalues are almost identical (see Figure 4). To further test our hypothesis, we construct the statically screened Coulomb interaction \(W_0\) (see eq 4) and solve the BSE based on a DFT-\(\omega\)-PBE eigensystem (instead of first computing \(Q\) eigenvalues using eq 2). We obtain values for the \(Q_\epsilon\) and \(Q_e\) excitation that are only 20 meV higher and 40 meV lower than the full GW+BSE solution, respectively, for structure “\(A\)”. Similarly, for structure “\(R\)”, the results are within less than 10 and 50 meV for the \(Q_\epsilon\) and \(Q_e\) excitation, respectively. This observation confirms that differences between the GW+BSE and TDDFT excitation energies are primarily due to differences in the xc and the BSE kernel. Generally, the overestimation of excitation energies that we observe with TDDFT is in line with results for other organic \(\pi\) chromophores such as rhodamine and rosamine\textsuperscript{54} and of phenothiazine dyes,\textsuperscript{55} for which it has been linked to an insufficient treatment of differential electron correlation between the ground and excited states by most TDDFT xc kernels.\textsuperscript{56}

\textbf{Charge Transfer Excitations with TD-LDA and \(G_{\text{\omega}}W_0\text{-}\text{LDA+BSE}\).} Finally, motivated by the excellent performance of \(G_{\text{\omega}}W_0\text{-}\text{LDA+BSE}\), we compare \(G_{\text{\omega}}W_0\text{-}\text{LDA+BSE}\), \(G_{\text{\omega}}W_0\text{-}\text{LDA+BSE}\), and TD-LDA results for structures “\(A\)” and “\(R\)” of BCL \(a\). Figure 5 shows the excitation spectrum calculated at these levels of theory. TD-LDA’s severe underestimation of charge transfer excitations is well known\textsuperscript{7} and leads to spurious excitations with charge transfer character at energies between \(Q_\epsilon\) and \(Q_e\) for BCL \(a\).\textsuperscript{7} Our comparison of structures “\(A\)” and “\(R\)” shows that while the energy of \(Q_e\) and \(Q_\epsilon\) is changing only slightly when TD-LDA is used, the relative position of these spurious low-oscillator strength excitations depends strongly on the structure. \(G_{\text{\omega}}W_0\text{-}\text{LDA+BSE}\) results in a very different, albeit no more reassuring, picture. For both structures, the first excitation already appears at energies below or around 1 eV and its oscillator strength is considerably lower than with TD-LDA; for structure “\(A\)”, the oscillator strength of \(Q_\epsilon\) is even lower than that of \(Q_e\). For structure “\(R\)”, excitations 2, 3, and 4 have similar, very low, oscillator strength. However, already at the \(G_{\text{\omega}}W_0\text{-}\text{LDA+BSE}\) level, no charge transfer excitations are found between \(Q_\epsilon\) and \(Q_e\)—a consequence of the inherent nonlocality of the BSE kernel. Finally, for both structures, eigenvalue self-consistency pushes all excitations to significantly higher energies and results in a quantitatively correct description of \(Q_\epsilon\) and \(Q_e\).

Inspection of the DOS calculated with DFT-xc, \(G_{\text{\omega}}W_0\text{-xc}\), and \(G_{\text{\omega}}W_0\text{-xc} (\text{xc} = \text{LDA}, \omega\text{-}PBE)\) shown in Figure 4 is instructive for understanding the contribution of eigenvalue differences to the TDDFT and GW+BSE excitation energies. The \(G_{\text{\omega}}W_0\text{-}\text{LDA DOS}\) underestimates the HOMO–LUMO gap and the energy difference between the HOMO and HOMO \(\sim 1\). In contrast, there is virtually no difference between the HOMO, HOMO \(\sim 1\), and LUMO energies as calculated with DFT-\(\omega\)-PBE, \(G_{\text{\omega}}W_0\text{-}\omega\text{-}PBE\), \(G_{\text{\omega}}W_0\text{-}\text{LDA+BSE}\), and \(G_{\text{\omega}}W_0\text{-}\text{LDA}\). As expected, the DFT-LDA DOS is markedly different, not only underestimating the HOMO–LUMO gap but also significantly underestimating the energy differences between the HOMO \(\sim 1\), HOMO \(\sim 2\), and HOMO \(\sim 3\). Notably, the spurious dark states between \(Q_\epsilon\) and \(Q_e\) that TD-LDA predicts have significant contributions from transitions involving these lower occupied states.
CONCLUSIONS

In this article, we performed a systematic first-principles study of the electronic structure and excitations of seven members of the (bacterio)chlorophyll family, which we validated through comparison with calculated and experimental literature results. The GW+BSE approach, when used in a partially self-consistent fashion, is in excellent agreement with experiment for excitations in the visible and near-ultraviolet part of the spectrum. GW+BSE also correctly predicts the energy difference between the low-energy Q_y and Q_x excitations of these pigments, relevant for the description of the coupling between pigment complexes, present in the light harvesting units and reaction centers of plants and bacteria and crucial for excitation energy and charge transfer. Most importantly, our results are almost entirely independent of the DFT eigensystem used as an input for the GW+BSE calculations. A computationally inexpensive LDA starting point leads to similar results as a more involved optimally tuned ωPBE starting point.

It should be noted that the GW approach, despite its implementation using Gaussian basis functions and the use of the RI approximation in MOLGW and other codes, remains a major bottleneck of these calculations due to its $O(N^4)$ scaling with system size. Furthermore, our results highlight that the GW approach, more so than DFT, requires careful convergence with respect to the basis set size. This limits its applicability to systems with a few (B)CL pigments at most, until algorithms with better scaling become more widely available.\(^{57-59}\) Our study joins a growing number of results, demonstrating that the GW+BSE approach can accurately predict neutral excitations of complex molecules without empirical parameters.\(^{59}\) With new approaches for combining GW+BSE with large-scale molecular mechanics simulations\(^{28}\) and polarizable continuum embedding\(^{60}\) emerging, an accurate prediction of excitation energy and charge transfer in complex molecular environments is within reach.

ASSOCIATED CONTENT

Supporting Information
The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.1c01240.

Additional convergence data, tuned range-separation parameters, and excitation energies of all molecules discussed in the main manuscript (PDF)

AUTHOR INFORMATION

Corresponding Author
Linn Leppert — MESA+ Institute for Nanotechnology, University of Twente, 7500 AE Enschede, The Netherlands; Institute of Physics, University of Bayreuth, Bayreuth 95440, Germany; orcid.org/0000-0002-4361-4382; Email: l.leppert@utwente.nl

Author
Zohreh Hashemi — Institute of Physics, University of Bayreuth, Bayreuth 95440, Germany

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.1c01240

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are grateful for helpful discussions with C. Filippi. This work was supported by the Bavarian State Ministry of Science and the Arts through the Collaborative Research Network Solar Technologies go Hybrid (SoITech), the Elite
Network Bavaria (ENB), and through computational resources provided by the Bavarian Polymer Institute (BPI).

REFERENCES