



## OPEN ACCESS

## EDITED BY

Zi-Piao Ye,  
Jinggangshan University, China

## REVIEWED BY

Elisabetta Collini,  
University of Padua, Italy  
Stefano Santabarbara,  
National Research Council (CNR), Italy

## \*CORRESPONDENCE

Jevgenij Chmeliov  
✉ jevgenij.chmeliov@ff.vu.lt

RECEIVED 05 April 2024

ACCEPTED 02 July 2024

PUBLISHED 07 August 2024

## CITATION

Rankelytė G, Gelzinis A, Robert B, Valkunas L and Chmeliov J (2024) Environment-dependent chlorophyll–chlorophyll charge transfer states in Lhca4 pigment–protein complex. *Front. Plant Sci.* 15:1412750. doi: 10.3389/fpls.2024.1412750

## COPYRIGHT

© 2024 Rankelytė, Gelzinis, Robert, Valkunas and Chmeliov. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Environment-dependent chlorophyll–chlorophyll charge transfer states in Lhca4 pigment–protein complex

Gabrielė Rankelytė<sup>1,2</sup>, Andrius Gelzinis<sup>1,2</sup>, Bruno Robert<sup>3</sup>, Leonas Valkunas<sup>2</sup> and Jevgenij Chmeliov<sup>1,2\*</sup>

<sup>1</sup>Institute of Chemical Physics, Faculty of Physics, Vilnius University, Vilnius, Lithuania,

<sup>2</sup>Department of Molecular Compound Physics, Center for Physical Sciences and Technology, Vilnius, Lithuania, <sup>3</sup>Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell, Gif-sur-Yvette, France

Photosystem I (PSI) light-harvesting antenna complexes LHCI contain spectral forms that absorb and emit photons of lower energy than that of its primary electron donor, P700. The most red-shifted fluorescence is associated with the Lhca4 complex. It has been suggested that this red emission is related to the inter-chlorophyll charge transfer (CT) states. In this work we present a systematic quantum-chemical study of the CT states in Lhca4, accounting for the influence of the protein environment by estimating the electrostatic interactions. We show that significant energy shifts result from these interactions and propose that the emission of the Lhca4 complex is related not only to the previously proposed a603<sup>+</sup>–a608<sup>−</sup> state, but also to the a602<sup>+</sup>–a603<sup>−</sup> state. We also investigate how different protonation patterns of protein amino acids affect the energetics of the CT states.

## KEYWORDS

LHCI, chlorophyll, amino acid, atomic partial charge, quantum chemistry, charge transfer states

## 1 Introduction

In oxygenic photosynthesis, the storage of the light energy in the form of chemical bonds requires two photosystems: photosystem II (PSII), which extracts electrons from water, and photosystem I (PSI), which produces the low redox potential electrons that ultimately reduce NADP into NADPH. In all organisms, PSI is constituted of a huge multi-protein architecture that binds a large number of chlorophylls and ensures a photo-induced electron transfer reactions across the photosynthetic membrane. In higher plants and green algae, the core of PSI is associated with an ensemble of light-harvesting chlorophyll–protein complexes, termed LHCI antennae (Haworth et al., 1983). In the early 80's, it was shown that PSI in *Chlamydomonas* binds two types of LHCI proteins: one fluorescing at 686 nm and the other around 715 nm (Ish-

Shalom and Ohad, 1983). Later, a similar situation was observed in higher plants as well, with even longer-wavelength (around 730 nm) emission of the redder spectral species (Lam et al., 1984). These two observed complexes with distinct fluorescence spectra are actually composed of four polypeptides, Lhca1–4, belonging to the LHC superfamily, which all bind chlorophylls *a* and *b*. Electron microscopy studies showed that LHCI complexes bind on one side of the PSI core (Boekema et al., 2001), and later X-ray crystallographic studies allowed a precise description of the LHCI–PSI core interactions (Ben-Shem et al., 2003).

Initially, the heterodimer comprised of the Lhca1 and Lhca4 complexes was shown to be responsible for the observed red emission (Knoetzel et al., 1992). More recent work, however, actually revealed that the native heterodimer formed by other two complexes, Lhca2/3, also displays a red-shifted emission (Wientjes and Croce, 2011). Because of distinct spectroscopic differences from the major light-harvesting complexes of PSII, LHCI, the LHCI antennae in general and its constituents Lhca1 and Lhca4 in particular attracted intense attention. Particularly, it was noted that the excited state energies of some of LHCI Chls are lower than that of the PSI primary electron donor, which, albeit extending the PSI absorption toward the red spectral region, also slows down the excited energy trapping by the reaction center. The mechanisms underlying this spectral red shift had to be clarified.

The study of these red forms largely benefited from approaches using LHC reconstitution (Plumley and Schmidt, 1987). Reconstituted complexes indeed led to the conclusion that the Chl molecules displaying highly red-shifted fluorescence exclusively belong to the Lhca4 complex (Croce et al., 2002). Moreover, reconstitution of the pertinently mutated Lhca4 showed that the red fluorescence of these complexes disappears when the asparagine ligand of Chl *a*603 (using the Qin et al. (2015) notation) is replaced by the histidine (Morosinotto et al., 2003). This was taken as a straightforward indication that this Chl *a*603 (together with the neighboring Chl *b*609 it interacts with) was responsible for the nearly 50-nm fluorescence redshift. Recently, genetic engineering of Lhca4 in *Arabidopsis* fully supported the results obtained from the reconstituted complexes (Li et al., 2023).

Detailed spectroscopic characterization of Lhca4 reconstituted from wild type and mutated polypeptides showed that the low-energy electronic transitions display a large bandwidth, which were proposed to arise from the mixing of an excitonic state, involving Chl *a*603 and Chl *b*609 molecules, with an inter-pigment charge transfer (CT) state (Croce et al., 2007). The determination of the PSI structure from higher plants at high resolution (Ben-Shem et al., 2003; Mazor et al., 2015; Qin et al., 2015, 2017) opened the way to model the electronic structure of LHCI complexes and Lhca4 in particular. By performing molecular dynamics simulations using the Quantum mechanics/Molecular mechanics (QM/MM) approach, the influence of a possible CT state in the Chl *a*603/*b*609 exciton manifold was investigated (Sláma et al., 2023). It was showed that such a state is highly sensitive to the precise relative geometry of the Chl pair, also explaining its disappearance in the asparagine mutant. Using an exciton model including a CT state, Novoderezhkin and Croce (2023) were able to simulate the

ensemble of Lhca4 electronic properties on the basis of an exciton-type mixing of a CT state with the excited state manifold.

In this paper, we also aim to characterize the CT states present in Lhca4 complex using quantum mechanical calculations. The CT states are considered only energetically, and possible mixing with the excitonic states of the pigments is not included. However, we do not focus solely on the Chl *a*603/*b*609 pair but rather try to pinpoint the ensemble of possible CT states present in this complex. Indeed, the asparagine-to-histidine (N98H, based on the notation of Qin et al. (2015)) mutation is likely to eventually perturb not only this pair, and, as underlined by Sláma et al. (2023), even minimal structural changes may result in dramatic changes in CT states manifold. Moreover, a second red form of Lhca4 exhibiting emission at ~705 nm exists, which is also present in the N98H mutant. Although the atomic structure of this complex is known, there are still some parameter values that should be specified in order to obtain a unique model of this system. Among these parameters, the exact protonation state of the protonable aminoacids in Lhca4 remains relatively poorly investigated. Accordingly, we tested the influence of several possible protonation states on the energetics of the CT states in this complex. Our results show that the  $a602^+ - a603^-$  state should also be related to the far-red fluorescence. Moreover, (de)protonation of a few amino acids could significantly alter the CT state energetics.

## 2 Methods

### 2.1 Quantum chemical calculations

The structure of Lhca4 was obtained as the 4th chain of the pea PSI–LHCI supercomplex structure (Qin et al., 2015) (PDB ID: 4XK8) from the PDB database (see Figure 1). In this work, we seek to investigate the energetics of the inter-chlorophyll charge transfer states that are expected to be present in this complex. Short distance between the pigments in the pigment–protein complex is a prerequisite for a CT state formation. In order to find possible locations of the CT states in Lhca4, we first determined which chlorophyll dimers have the shortest (up to 12 Å) Mg–Mg distance and found 12 dimers that satisfy this condition.

For all single chlorophylls that form these dimers, geometry optimization *in vacuo* was performed. The phytol tail of these chlorophylls was truncated by replacing the phytol group with the methyl group, similarly to the previous works (Jornet-Somoza et al., 2015; Sirohiwal et al., 2020a; Mikalčiūtė et al., 2022), in order to reduce the calculation time and prevent extreme structural changes of the tail during the geometry optimization. Missing hydrogen atoms were added using GaussView software (Dennington et al., 2019). Ground state geometry optimization was performed using Gaussian 16 (Frisch et al., 2016). Density Functional Theory (DFT) calculations were performed using CAM-B3LYP functional (Yanai et al., 2004) and 6–31G(d) basis set. The DFT level of theory with the aforementioned functional has been commonly used for chlorophyll systems (Yin et al., 2007; Higashi et al., 2014; Frankcombe, 2015; Bold et al., 2020).

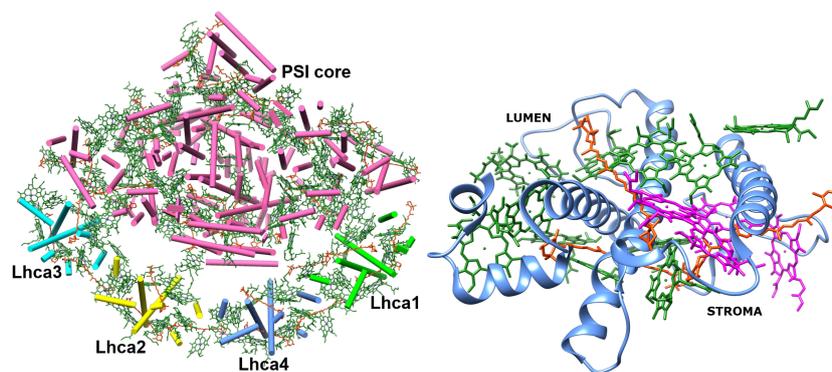


FIGURE 1

Left: PSI–LHCI supercomplex (PDB ID: 4XK8) (Qin et al., 2015). Right: Lhca4 light-harvesting complex. Chlorophylls *a* are depicted in green and Chlorophylls *b* are depicted in magenta. Carotenoids (BCR, XAT and LUT) are depicted in orange.

We formed 12 selected dimers by mapping individual chlorophylls with optimized geometry onto the Lhca4 structure using UCSF Chimera software (Pettersen et al., 2004). TD-DFT calculations were conducted for the lowest eight excited states of these dimers, treating them as a single supermolecule, employing the CAM-B3LYP functional and a 6–31G(d) basis set. Our examination included various parameters, such as energy, static dipole moment, transition dipole moment, and the sum of the Mulliken partial charges for each molecule in the dimer. These findings unveiled the presence of 19 potential inter-chlorophyll charge transfer states within the Lhca4 pigment–protein complex.

## 2.2 Modelling of pigment–protein complex environment

In order to account for the environment that surrounds the dimers of interest and calculate its impact on energy values of the CT states, we chose to follow the methodology developed by Renger and co-workers, which was previously applied for local excitations (Adolphs et al., 2007; Renger and Müh, 2013). The environment (protein, other single chlorophyll molecules, and carotenoids) is divided into the environmental blocks. These blocks are coupled through the Coulomb interactions. The dimer of interest  $m$  is described by the wavefunction  $|A_a^{(m)}\rangle$ , where  $a$  indicates the electronic state of the dimer. The dimer is surrounded by  $N - 1$  environmental blocks with the wavefunctions  $|B_b^{(\eta)}\rangle$ , where  $\eta = 1, \dots, N - 1$  is the number of the block and  $b$  is the electronic state of this block. These wavefunctions satisfy the following Schrödinger equations:

$$\begin{aligned} H_A^{(m)}|A_a^{(m)}\rangle &= E_a^{(m)}|A_a^{(m)}\rangle, \\ H_B^{(\eta)}|B_b^{(\eta)}\rangle &= F_b^{(\eta)}|B_b^{(\eta)}\rangle. \end{aligned} \quad (1)$$

Here in Equation 1  $H_A^{(m)}$  and  $H_B^{(\eta)}$  are the Hamiltonians of the environmental blocks of the system *in vacuo*,  $E_a^{(m)}$  is the energy of the  $a$ th electronic state of the dimer,  $F_b^{(\eta)}$ —the energy of the  $b$ th

electronic state of the environmental block  $\eta$ . The total Hamiltonian of the system is given by

$$H = H_A^{(m)} + \sum_{\eta} H_B^{(\eta)} + \sum_{\eta} V_{AB}^{(m,\eta)} + \frac{1}{2} \sum_{\eta \neq \eta'} V_{BB}^{(\eta,\eta')}. \quad (2)$$

Here  $V_{AB}^{(m,\eta)}$  describes the Coulomb coupling between the dimer and  $\eta$  block and  $V_{BB}^{(\eta,\eta')}$  describes the interaction between two environmental blocks  $\eta$  and  $\eta'$ . The last two terms define the interaction Hamiltonian  $V$ . We assume that the blocks do not undergo electron exchange, thus the wavefunctions of the system can be combined using the Hartree product:

$$|\psi_{ab}^{(m)}\rangle = |A_a^{(m)}\rangle \prod_{\eta} |B_b^{(\eta)}\rangle, \quad (3)$$

where  $|\psi_{ab}^{(m)}\rangle$  is the eigenfunction of the Schrödinger equation

$$\left( H_A^{(m)} + \sum_{\eta} H_B^{(\eta)} \right) |\psi_{ab}^{(m)}\rangle = \left( E_a^{(m)} + F_b \right) |\psi_{ab}^{(m)}\rangle, \quad (4)$$

$\mathbf{b} = \{b_1, b_2, \dots, b_{\eta}, \dots\}$  enumerates the electronic states of the environment, and  $F_b = \sum_{\eta} F_{b_{\eta}}^{(\eta)}$ . Equations 3 and 4 can be used to determine the energy shift  $\Delta E_a^{(m)}$  that the  $a$ th state of the dimer acquires after experiencing the electrostatic interaction with the environmental blocks, given by the two last terms in Equation 2. Within the first order perturbation theory (the charge density coupling (CDC) method), the energy shift is given as:

$$\Delta E_a^{(m)} = \langle \psi_{a\mathbf{0}}^{(m)} | V | \psi_{a\mathbf{0}}^{(m)} \rangle, \quad (5)$$

here index  $\mathbf{0} = \{0, 0, \dots\}$  denotes the electronic ground state of all the environmental blocks.

The energy of the transition between the ground and the  $s$ th excited state is given by:

$$E_s^{(m)} = E_0^{(m)} + \Delta E_s^{(m)} - \Delta E_0^{(m)}. \quad (6)$$

Here  $E_0^{(m)}$  is the transition energy between the ground and the  $s$ th excited state *in vacuo*,  $\Delta E_s^{(m)}$  and  $\Delta E_0^{(m)}$  are the energy shifts for the excited and the ground state, respectively. Energy shifts of electronic

state  $a$  of dimer  $m$  are determined by calculating the values of the Coulomb matrix elements given in Equation 5:

$$\Delta E_a^{(m)} = \sum_{\eta} \langle A_a^{(m)} B_0^{(\eta)} | V_{AB}^{(m,\eta)} | A_a^{(m)} B_0^{(\eta)} \rangle + \frac{1}{2} \sum_{\eta \neq \eta'} \langle B_0^{(\eta)} B_0^{(\eta')} | V_{BB}^{(\eta,\eta')} | B_0^{(\eta)} B_0^{(\eta')} \rangle. \quad (7)$$

Here, in Equation 7, the first term on the right hand side denotes the Coulomb interaction between the charge densities of the dimer  $m$  of state  $a$  and the environmental block  $\eta$  in its ground state, and the second one denotes the Coulomb interaction between the charge densities of environmental blocks  $\eta$  and  $\eta'$  in the ground state. The elements of the Coulomb matrix can be approximated as the interaction between the atomic partial charges (Adolphs et al., 2007; Renger and Müh, 2013):

$$\langle A_a^{(m)} B_0^{(\eta)} | V_{AB}^{(m,\eta)} | A_a^{(m)} B_0^{(\eta)} \rangle = \frac{1}{4\pi\epsilon_0} \sum_{I,J} \frac{q_I^{(m)}(a, a) q_J^{(\eta)}(0, 0)}{|R_I^{(m)} - R_J^{(\eta)}|}, \quad (8)$$

where in Equation 8  $\epsilon_0$  denotes the permittivity of vacuum,  $q_I^{(m)}(a, a)$  and  $q_J^{(\eta)}(0, 0)$  are the atomic partial charges of the  $I$ -th atom of the dimer and  $J$ -th atom of the environment, respectively. The positions of these charges are  $R_I^{(m)}$  and  $R_J^{(\eta)}$ .

The possible errors in the values of the atomic partial charges can be compensated by introducing the effective dielectric constant  $\epsilon_{\text{eff}}$  to scale the Coulomb interaction. Its adjustment can thus approximately account for the higher order terms not included in the CDC approximation. Then, following Equation 6, the energy of the transition between the ground and the  $s$ -th excited state of the dimer  $m$  is given as:

$$\begin{aligned} E_s^{(m)} &= E_0^{(m)} + \frac{1}{\epsilon_{\text{eff}}} \sum_{\eta} \left( \langle A_s^{(m)} B_0^{(\eta)} | V_{AB}^{(\eta)} | A_s^{(m)} B_0^{(\eta)} \rangle - \langle A_0^{(m)} B_0^{(\eta)} | V_{AB}^{(\eta)} | A_0^{(m)} B_0^{(\eta)} \rangle \right) \\ &= E_0^{(m)} + \frac{1}{4\pi\epsilon_0\epsilon_{\text{eff}}} \sum_I \sum_J \frac{\left( q_I^{(m)}(s, s) - q_I^{(m)}(0, 0) \right) q_J^{(\eta)}(0, 0)}{|R_I^{(m)} - R_J^{(\eta)}|}. \end{aligned} \quad (9)$$

Here, in Equation 9,  $q_I^{(m)}(s, s)$  and  $q_I^{(m)}(0, 0)$  are the atomic partial charges of the dimer  $m$  in its excited and ground state, respectively. The values of these charges were obtained by fitting the electrostatic potential of the dimers *in vacuo*.  $q_J^{(\eta)}(0, 0)$  are the atomic partial charges of the environmental blocks. The partial charges of single chlorophylls and carotenoids in their ground state were determined by fitting the electrostatic potential as well. The calculation of the partial charges was performed using Gaussian 16 package (Frisch et al., 2016) at the TD-DFT/CAM-B3LYP level of theory with 6-31G(d) basis set and the Multiwfn software (Lu and Chen, 2011; Zhang and Lu, 2021). After assuming that single tailless chlorophylls of optimized geometry have very similar spatial structure, we chose two reference chlorophylls (*a602* and *b605*) for the calculation of atomic partial charges. The values of partial charges of reference chlorophylls were used for all single chlorophylls (when chlorophylls were considered to be part of the environment) present in the pigment-protein complex (see Supplementary Table S-I in Supplementary Material). The calculation of the atomic partial charges of carotenoids was based on the original geometry of the molecules (see Supplementary Table S-II in Supplementary Material). While choosing non-optimized geometry might induce some error in the calculated partial charges, this was

done to avoid possible geometry changes that would be hindered in the protein environment. Partial charges of the chlorophyll monomers (used for calculating energy shifts for monomers) are given in Supplementary Tables S4–S33 in Supplementary Material. Partial charges for the chlorophyll dimers are given in Supplementary Tables S34–S74 in Supplementary Material. The charges of hydrogen atoms for all the molecules (except amino acids) were set to 0. The atomic partial charges of amino acid molecules were taken from the CHARMM (Brooks et al., 2009) force field, including the charges of the hydrogen atoms [they were added using AMBER 2020 software (Case et al., 2020)].

## 2.3 Determining the protonation pattern of the protein

The largest part of the Lhca4 complex is the protein, consisting of 196 amino acid molecules. In order to simulate the environment properly, it is necessary to determine the protonation pattern of the protein. Besides the N-terminus and C-terminus of the protein chain, Lhca4 protein contains 6 types of titratable amino acids: aspartic acid (ASP), glutamic acid (GLU), tyrosine (TYR), histidine (HIS), lysine (LYS) and arginine (ARG). The first three types are acidic titratable groups that are prone to releasing a proton to the environment and the latter three types are basic titratable groups that are likely to accept a proton from the environment. The corresponding Henderson–Hasselbalch equation for the acids is:

$$\text{pH} = \text{p}K_a + \lg \frac{[A^-]}{[\text{HA}]}, \quad (10)$$

and for bases:

$$\text{pH} = \text{p}K_b - \lg \frac{[\text{BH}^+]}{[\text{B}]}, \quad (11)$$

where the brackets denote the concentrations of the corresponding species at equilibrium.  $\text{p}K_b = -\lg K_b$  and  $\text{p}K_a = -\lg K_a$  are the logarithms of the equilibrium constants for the reaction of basic and acidic amino acids, respectively. If the equilibrium constant for the dissociation and association reactions as well as the pH of the environment are known, the most probable protonation pattern for every titratable group of the protein can be determined. In most cases, acidic titratable group is considered fully protonated if the difference between pH and the equilibrium constant is  $\lesssim -2$ , and fully deprotonated if this difference is  $\gtrsim 2$  (for basic titratable groups the rule is inverse).

## 3 Results

### 3.1 Estimating the protonation pattern and modelling the protein

The termini of the protein chain were modelled as follows: N-terminus was capped with a positively charged acetyl group and C-terminus was modelled with negatively charged  $\text{COO}^-$  group. For every titratable group present in the protein, the values of the

equilibrium constant in neutral pH were estimated using PROPKA 3 tool (Olsson et al., 2011; Søndergaard et al., 2011) (see Supplementary Table S-III in Supplementary Material). There were 9 amino acids that had the absolute value of the difference between  $pK_a$  and pH values less or equal to 2. These residues are ASP159, GLU95, GLU113, GLU145, GLU153, GLU202, HIS222, HIS236, and HIS242. Note that throughout the paper we use the numbering and the nomenclature of the PDB file 4XK8 (Qin et al., 2015) for the amino acids and chlorophylls, contrary to some of the literature, where often the nomenclature of the LHCI crystal structure (Liu et al., 2004) is used instead. Using the Henderson–Hasselbalch Equations 10 and 11 we determined that there are three amino acids that are in their non-standard protonation state (in neutral pH): GLU145, GLU153 and HIS242. These residues are considered protonated and their total charge is neutral for glutamic acids and positive for histidine. The rest of the histidines were considered to be  $\epsilon$ -tautomers. After setting the non-standard charges for the titratable groups mentioned above and the termini, the total charge of the Lhca4 protein chain is  $-5$  (for full description of the protein chain see Supplementary Table S1 in Supplementary Material).

### 3.2 Energy shifts of chlorophyll monomers

Lhca4 complex contains 15 chlorophyll monomers (11 chlorophylls *a* and 4 chlorophylls *b*). We calculated  $Q_y$  energies for these tailless monomers *in vacuo* as well as their  $Q_y$  energy shifts determined by the environmental blocks (other chlorophyll

monomers, carotenoids, and the protein chain) using CDC method. The values are given in Table 1. The energy shifts are also depicted in Figure 2. The energy shifts are of the similar magnitude as was previously estimated for the PSI core using the same approach (Adolphs et al., 2010). The lowest  $-$ energy Chl *a* is *a602*, according to our calculations. Interestingly, in the previous modeling work on Lhca, where site energies were obtained by the fits of the linear optical spectra using the modified Redfield theory (Novoderezhkin et al., 2016), this was not the case. The current results indirectly supports the earlier assessment by some of us that the site energies obtained by fitting the optical lineshapes using the modified Redfield theory needs to be reevaluated (Gelzinis et al., 2015). It is also of note that the energy shifts for Chl *b* molecules are somewhat smaller than for Chl *a*, which should be related to the smaller difference values between the static dipole moments of the ground and the  $Q_y$  states. Interestingly, our calculations slightly overestimate the energy gap between the Chl *a* and Chl *b* pigments (at  $\sim 800 \text{ cm}^{-1}$ ), contrary to some of the previous work in literature, where the same functional and basis set was used, but the environmental effects were accounted for in a more advanced way (Jurinovich et al., 2015).

### 3.3 Energy shifts of chlorophyll dimers

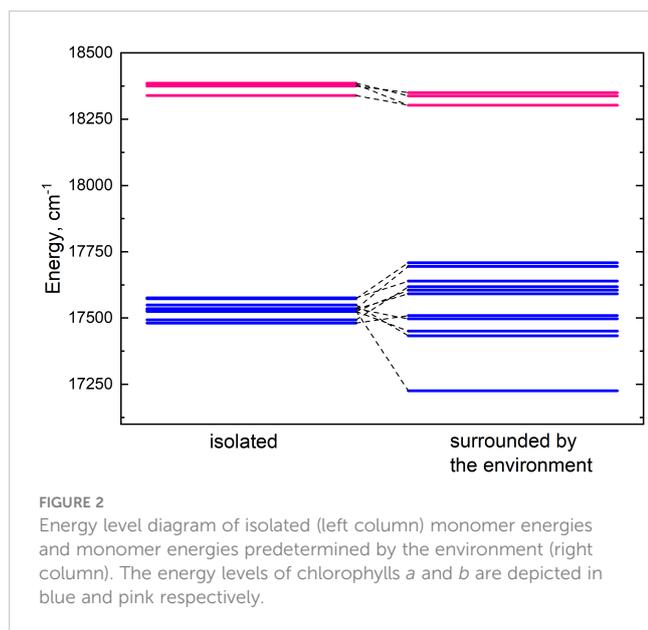
#### 3.3.1 Selected dimers and their CT states

We selected chlorophyll dimers that have the Mg–Mg distance less than  $12\text{\AA}$ . For each selected dimer, we performed quantum chemical calculations and compared static and transition from the

TABLE 1  $Q_y$  energies of the tailless isolated chlorophyll monomers, energy shifts for these states calculated with CDC method ( $\epsilon_{\text{eff}} = 1$ ), and their total  $Q_y$  energies in the Lhca4 complex.

Monomer	$E_{Q_y}$ , $\text{cm}^{-1}$	$\Delta E_{Q_y}$ , $\text{cm}^{-1}$	$E_{Q_y} + \Delta E_{Q_y}$ , $\text{cm}^{-1}$	$ \mu_1 - \mu_0 $ , D	$ \mu_{0 \rightarrow 1} $ , D	$\Delta E_{Q_y}^{N98H}$ , $\text{cm}^{-1}$	$E_{Q_y} + \Delta E_{Q_y}^{N98H}$ , $\text{cm}^{-1}$
<i>a601</i>	17576	64	17639	0.6062	5.1466	59	17635
<i>a602</i>	17527	-301	17226	0.6330	5.0764	-283	17244
<i>a603</i>	17481	28	17509	0.5303	5.1435	-151	17330
<i>a604</i>	17549	-117	17433	0.5986	5.1148	-114	17435
<i>a608</i>	17535	160	17695	0.5971	5.1646	169	17704
<i>a609</i>	17493	125	17618	0.6149	5.1338	120	17613
<i>a610</i>	17535	-38	17497	0.5936	5.1644	-37	17498
<i>a611</i>	17525	80	17605	0.6096	5.1453	81	17606
<i>a612</i>	17525	-74	17451	0.6049	5.0521	-70	17455
<i>a613</i>	17573	136	17709	0.8073	5.2488	138	17711
<i>a614</i>	17534	57	17591	0.5972	5.1900	53	17587
<i>b605</i>	18381	-43	18338	0.2503	4.1736	-40	18341
<i>b606</i>	18339	-36	18303	0.1315	4.1228	-39	18300
<i>b607</i>	18386	-48	18338	0.1844	4.0244	-43	18343
<i>b615</i>	18375	-25	18350	0.1548	3.9901	-26	18349

The difference between the static dipole moment values of the ground and the first excited states and the transition from the ground to the excited state dipole moment values for every monomer are given as well. The last two columns represent energy shifts and total  $Q_y$  energies in N98H mutant of Lhca4 complex (see Section 3.4).



ground to the excited state dipole moments, as well as the sum of Mulliken partial charges for each monomer of eight lowest excited states. Large values of static dipole moment and small values of transition dipole moment indicate the CT state (Premvardhan et al., 2008), and the sum of Mulliken partial charges for such state will be approximately equal to 1 for one pigment of the dimer and to  $-1$  for another pigment (Voityuk, 2013). The results of these calculations for the dimer *a610–a611* are given in Table 2. The values of the physical quantities mentioned above significantly differ for the fifth and sixth excited states. Therefore, these states are CT states.

The results for the rest of the selected dimers can be found in Supplementary Material. After analyzing the results for all 12 selected dimers, 19 low energy CT states were found in Lhca4 complex. It is crucial to stress that after performing calculations *in vacuo*, the  $Q_y$  energies of monomers (see Table 1) are far lower than

the energies of CT states of dimers and it is only the environmental electrostatics effects that can make them comparable.

### 3.3.2 Estimated protonation of the protein

For every CT state found in the Lhca4 complex we calculated the energy shift influenced by the environment. The protein is considered to be in the estimated protonation state where the total charge of the chain is  $-5$ . The environmental chlorophylls and carotenoids are in their ground state while their net charge is set to neutral. The energy shifts were calculated using the CDC method. The CT state energy values of the isolated dimers and energy shifts caused by the environment are given in Table 3. The energies of the CT states *in vacuo* and in the Lhca4 environment are compared in Figure 3 as well.

There are four dimers whose total energy of the CT state is lower than the  $Q_y$  energy of monomers: *a602–a603*, *a603–a608*, *b607–a609* and *b607–b615*. In principle, these CT states might contribute to the formation of the red-shifted peak in the experimental fluorescence spectra (Morosinotto et al., 2003; Wientjes and Croce, 2011). The former two dimers are more probable candidates, since our calculated electrostatic shifts were done by assuming  $\epsilon_{\text{eff}} = 1$ , while in proteins this value is expected to be slightly higher.

The lists of environmental blocks that create the largest energy shifts for monomers (see Table 4) and dimers mentioned above (see Table 5) reveal that the protein influences both monomer  $Q_y$  energies and dimer CT energies more than remaining pigments. While looking at Table 4, one can notice that the majority of monomers listed are the ones that form the four dimers of interest and that the largest energy

shifts are calculated for these exact monomers (*a602*, *a608*, *a609*, *b607* and *b615*). Amino acids that have significant influence on these monomers also have effect on the corresponding dimers that these monomers form (see Table 5). Amino acids that significantly influence both *a602* monomer and *a602–a603* dimer are GLU95, ASP77 and ARG209. For the dimer *a603–a608* and

TABLE 2 The comparison of the static (SDM) and transition (TDM) dipole moments for the first eight excited states of the dimer *a610–a611*.

State	$E, \text{cm}^{-1}$	SDM, D		TDM, D	Mulliken charges	
		$ \mu_n $	$ \mu_n - \mu_0 $	$ \mu_{0 \rightarrow n} $	<i>a610</i>	<i>a611</i>
$S_0$	–	4.127	–	–	-0.02	0.02
$S_1$	17180	4.075	0.103	7.814	-0.02	0.02
$S_2$	17583	4.006	0.149	2.215	-0.02	0.02
$S_3$	20406	3.570	1.123	1.161	-0.02	0.02
$S_4$	20648	3.415	1.111	1.742	-0.02	0.02
$S_5$	<b>24197</b>	<b>30.872</b>	<b>32.649</b>	<b>0.811</b>	<b>0.90</b>	<b>-0.90</b>
$S_6$	<b>24762</b>	<b>36.266</b>	<b>34.022</b>	<b>0.280</b>	<b>-0.97</b>	<b>0.97</b>
$S_7$	27504	2.888	1.528	0.656	0.01	-0.01
$S_8$	28149	1.769	17.568	6.353	0.48	-0.48

The last two columns give the sum of Mulliken partial charges (in atomic units) for each pigment of the dimer. The values given in bold indicate the CT state.

TABLE 3 Energy shifts for the CT states determined using CDC method ( $\epsilon_{\text{eff}} = 1$ ).

No.	Dimer (Chl <sub>1</sub> -Chl <sub>2</sub> )	R, Å	S <sub>CT</sub>	E <sub>0</sub> , cm <sup>-1</sup>	ΔE, cm <sup>-1</sup>	E <sub>0</sub> + ΔE, cm <sup>-1</sup>	ΔE <sup>N98H</sup> , cm <sup>-1</sup>	E <sub>0</sub> + ΔE <sup>N98H</sup> , cm <sup>-1</sup>
1	a601-a610	11.75	S <sub>5</sub> (+-)	27505	8379	35884	8311	35816
2	a601-a610		S <sub>6</sub> (-+)	27505	-8281	19224	-8217	19288
3	a602-a603	11.87	S <sub>5</sub> (-+)	27505	15756	43261	16399	43904
4	a602-a603		S <sub>6</sub> (+-)	27666	-15428	<b>12238</b>	-16366	<b>11300</b>
5	a603-a608	8.93	S <sub>6</sub> (+-)	24521	-10292	<b>14229</b>	-8840	<b>15681</b>
6	a604-b605	8.67	S <sub>5</sub> (+-)	23633	-3598	20035	-3536	20097
7	a604-b605		S <sub>8</sub> (+-)	27263	-2831	24432	-2789	24474
8	a604-b605		S <sub>6</sub> (+-)	26537	-3488	23049	-3429	23108
9	b607-a608	10.46	S <sub>5</sub> (+-)	26698	1019	27717	276	26974
10	b607-a609	11.75	S <sub>5</sub> (+-)	25488	-9189	<b>16299</b>	-8990	<b>16498</b>
11	b607-b615	11.43	S <sub>7</sub> (-+)	27586	-9652	<b>17933</b>	-9398	<b>18188</b>
12	a608-a614	10.23	S <sub>5</sub> (+-)	25811	1231	27042	403	26214
13	a608-a614		S <sub>6</sub> (-+)	26940	-1012	25928	-254	26687
14	a609-a611	11.86	S <sub>5</sub> (-+)	27344	10	27354	299	27642
15	a610-a611	9.90	S <sub>5</sub> (+-)	24198	-4484	19714	-4460	19738
16	a610-a611		S <sub>6</sub> (-+)	24762	4274	29036	4253	29015
17	a612-a613	9.40	S <sub>5</sub> (+-)	23714	-1805	21908	-1622	22092
18	a612-a613		S <sub>6</sub> (+-)	25892	-1975	23916	-1812	24080
19	a612-a613		S <sub>7</sub> (-+)	27182	1889	29071	1710	28892

The energies of the CT states of isolated pigments E<sub>0</sub> and total site energies are given as well. Column R denotes the Mg-Mg distance between the pigments of the dimer. Column S<sub>CT</sub> denotes the type of the CT state. The last two columns represent energy shifts and total CT state energies in N98H mutant of Lhca4 complex (see Section 3.4). The cases where monomer Q<sub>y</sub> energy value exceeds the value of the CT state of the dimer are given in bold.

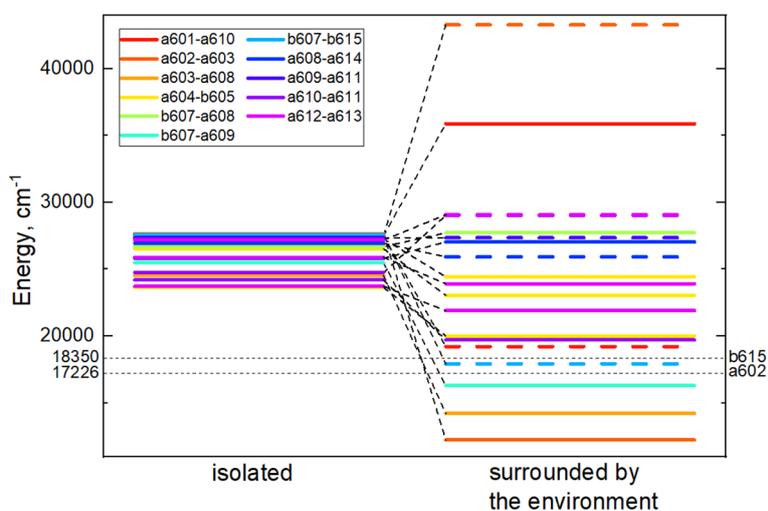


FIGURE 3

Energy level diagram of the isolated (left column) dimer CT state energies and CT energies predetermined by the environment (right column). In the right column, energies of the CT states, whose first pigment is negatively charged and the second one is positively charged, are given in dash. Cases, where the first pigment is positively charged and the second one is negatively charged, are given in solid lines. Black dashed lines mark the lowest and the highest monomer Q<sub>y</sub> energies predetermined by the environment.

TABLE 4 List of the environmental blocks that have the largest ( $\geq 80\text{cm}^{-1}$ ) influence on the monomer  $Q_y$  energies (in descending order).

Monomer	Env. block	$\Delta E, \text{cm}^{-1}$	Monomer	Env. block	$\Delta E, \text{cm}^{-1}$
a608	ARG156	244	a601	GLY54	118
b615	ARG155	203	b615	ASP169	-111
a609	ARG100	201	a610	LYS203	111
b607	ARG100	-193	a613	HIE236	111
a602	ARG209	165	a609	LYS201	98
b607	ARG156	-152	b606	SER138	-97
a602	GLU95	-148	a610	GLU199	-88
a611	LYS203	146	b606	GLU132	85
b607	ARG155	144	a613	PRO240	-84
a602	ASP77	-118			

TABLE 5 List of the environmental blocks that have the largest ( $\geq 4000\text{cm}^{-1}$ ) influence on the CT state energies of dimers a602–a603, a603–a608, b607–a609 and b607–b615 (in descending order).

Dimer	Env. block	$\Delta E, \text{cm}^{-1}$	Dimer	Env. block	$\Delta E, \text{cm}^{-1}$
b607–b615	<b>ASP169</b>	-10990	b607–a609	<b>ARG155</b>	-5931
a602–a603	<b>GLU95</b>	-10679	a603–a608	<b>ARG156</b>	-5354
a602–a603	<b>ARG209</b>	10505	a602–a603	<b>ASP77</b>	-5209
b607–a609	GLU204	-9338	a602–a603	ASP73	-4887
b607–b615	<b>ARG156</b>	-8191	b607–a609	ASP159	4823
b607–a609	<b>ARG156</b>	-7407	b607–b615	ASP159	4425
b607–b615	<b>ARG100</b>	-6861	b607–a609	LYS203	4041

The blocks that have large influence on corresponding monomer  $Q_y$  energies as well are given in bold.

monomer a608 the most significant impact comes from ARG156.  $Q_y$  energy of the monomer b607 and CT state energy of the dimer b607–a609 are

both shifted by ARG155 and ARG156 amino acids. The last dimer b607–b615 and monomer b615 are strongly influenced by ASP169, and the same dimer and monomer b607 are influenced by ARG156 and ARG100.

### 3.3.3 Possible non-standard protonation of the protein

The Lhca4 protein chain contains nine amino acids with an absolute difference between their  $pK_a$  (or  $pK_b$ ) and pH values less than or equal to 2. In principle, these amino acids could easily gain/lose a proton under physiological conditions. We calculated the energy shifts for the CT states of selected dimers considering these various protonation patterns of the protein. There were nine distinct protonation states, with one of the nine amino acid residues exhibiting a protonation state opposite to the one estimated and all other residues exhibiting previously estimated protonation state. The differences in the CT state energy shifts calculated in estimated protonation environment and all non-standard protonation environments are given in the Figure 4 (similar representation of the results in various protonation

patterns for chlorophyll monomers is given in Supplementary Figure S1 in Supplementary Material). The most significant energy shift differences after changing the protonation pattern of the protein are obtained for the dimers a602–a603, a603–a608, a604–b605, b607–a608, b607–a609, b607–b615, and a608–a614. Amino acids that are mostly responsible for these differences are protonated GLU95 (for the dimer a602–a603), standard GLU145 (for the dimer a604–b605), and standard GLU153 (for the dimers a603–a608, b607–a608, b607–a609, b607–b615, and a608–a614). Four dimers, whose CT state energies are lower than the monomer  $Q_y$  energies, together with amino acids, that create large energy shifts for these pigments, are shown in Figure 5.

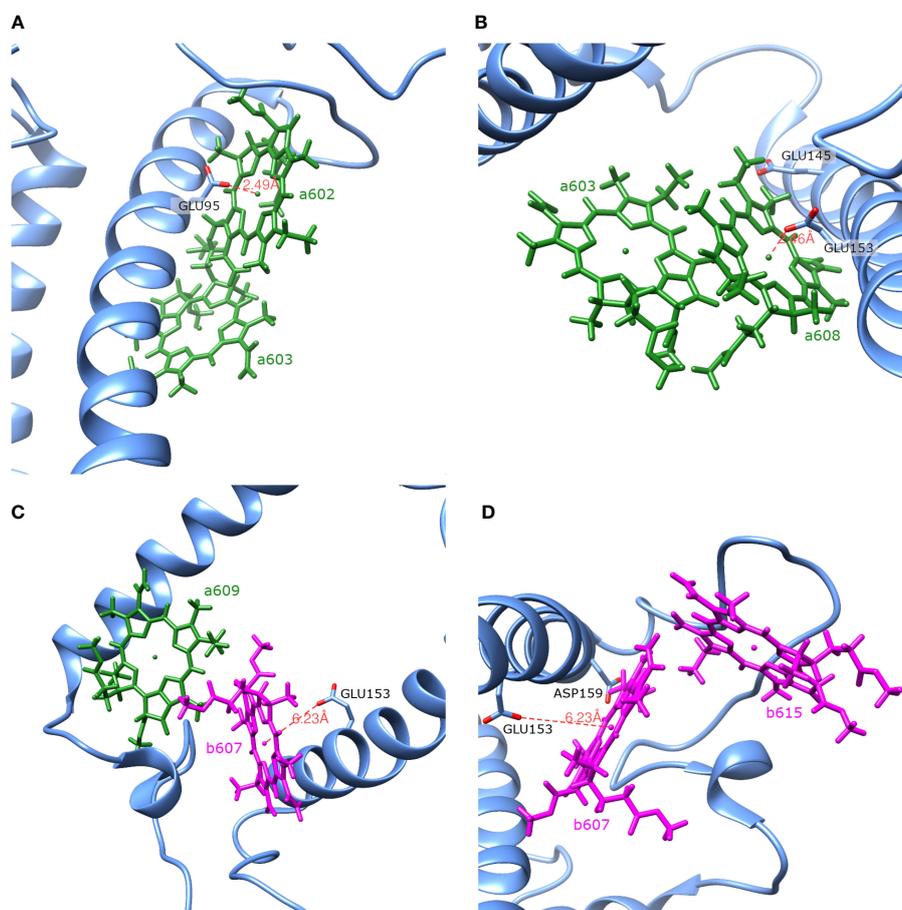
## 3.4 N98H mutant

Apart from the WT environment of Lhca4, we also checked whether the energy shifts for the CT states of the Chl dimers and the  $Q_y$  energies of the monomers would still lead to similar results in N98H mutant environment. In N98H mutant, the asparagine (ASN) in position 45 (counting from N-terminus) is substituted for histidine (HIS). We chose the most probable rotamer of histidine (probability ~36%) using UCSF Chimera software

a601-a610s5	350	-682	704	-273	-259	3230	97	368	26
a601-a610s6	-345	684	-685	273	261	-3098	-91	-324	-40
a602-a603s5	165	-11956	701	-3444	-4227	-2918	1671	-169	87
a602-a603s6	-196	11711	-664	3269	4242	2875	-1638	175	-89
a603-a608s6	-1319	892	546	1755	6984	157	1022	509	-324
a604-b605s5	-397	145	2138	8798	1609	324	418	650	-382
a604-b605s6	-367	151	2049	8340	1488	305	460	618	-365
a604-b605s8	-291	177	1932	7723	1317	343	125	584	-335
b607-a608s5	-6031	637	359	-4163	-7960	-547	990	269	-205
b607-a609s5	-5554	1227	585	1580	5164	2460	167	552	-325
b607-b615s7	-5124	-1997	-762	2166	6080	-1069	-834	-531	350
a608-a614s5	1641	2247	898	-3793	-10994	1012	953	596	-373
a608-a614s6	-1405	-1955	-770	3112	9364	-873	-821	-514	321
a609-a611s5	-3309	-1014	1503	1034	2607	-71	563	1617	-857
a610-a611s5	-283	647	-1161	641	513	2049	-524	-46	-71
a610-a611s6	272	-763	1251	-659	-514	-2263	569	122	44
a612-a613s5	342	807	1414	-747	-647	130	2410	-3468	1824
a612-a613s6	319	750	1422	-693	-598	135	2215	-3130	1673
a612-a613s7	-355	-859	-1370	751	659	-173	-2374	3996	-1876
	ASP159	GLU95	GLU113	GLU145	GLU153	GLU202	HIS222	HIS236	HIS242



**FIGURE 4**  
The difference between the energy shifts of the CT states, calculated in estimated protonation environment and the non-standard protonation state of the specific amino acid indicated at the bottom of each column.



**FIGURE 5**  
Spatial arrangement of the dimers a602–a603 (A), a603–a608 (B), b607–a609 (C) and b607–b615 (D) together with amino acids that have the largest influence on the Qy energies of the monomers and the CT energies of the dimers.

(Pettersen et al., 2004). The same methods were used for the mutated protein chain to find that the newly added amino acid is neutral  $\epsilon$ -tautomer and thus the net charge of the protein remains unchanged.

Using the CDC method, energy shifts in the new environment for both monomers and dimers were calculated. The results are given in the last two columns of Tables 1 and 3, respectively. Compared to the results obtained in the environment of WT Lhca4, energies of the monomer  $Q_y$  and dimer CT states in N98H environment show no significant changes. After comparing the energies of the dimer CT states and the corresponding monomer  $Q_y$  states in N98H environment, it is clear that there are 4 CT states whose energy is lower than the  $Q_y$  energy of the monomers that form these dimers. These four states are the same as were obtained in WT Lhca4 environment.

## 4 Discussion

The presence of CT states in the Lhca4 complex has been suggested based on the significantly red-shifted peaks in the emission spectra. LHCII trimers, for instance, fluoresce at  $\sim 680\text{nm}$  (Chmeliov et al., 2016), and almost the same fluorescence wavelength is observed for the monomeric CP29 complexes (Mascoli et al., 2019). For the light-harvesting antenna of PSI, the experimental investigations are hindered by the fact that Lhca1–4 monomers cannot be isolated. Nonetheless, both Lhca1/4 heterodimer (Wientjes and Croce, 2011), and reconstituted Lhca4 complexes (Morosinotto et al., 2003) show red-shifted fluorescence, with peaks at  $\sim 730\text{nm}$ . While initially this was thought to be related to excitonic interaction (Morosinotto et al., 2003), it was soon proposed that CT states should play a role (Ihalainen et al., 2005), which was confirmed later by Stark absorption measurements (Romero et al., 2009). Another point to consider is that Lhca4 complexes might exist in multiple conformations. This is supported by single-molecule spectroscopy fluorescence measurements (Krüger et al., 2011) and by time-resolved fluorescence measurements (Wientjes et al., 2011). All this implies that the spectroscopic features of Lhca4 might be influenced by more than one CT state. Therefore, in this work we have performed a systematic quantum-chemical investigation of the possible CT states, paying particular attention to the electrostatic environment-induced shifts of the CT state energies. Below we will discuss the implications of our work in more detail.

We first note that we have found 4 possible CT states that, in principle, could contribute to the experimental fluorescence signal:  $a602^+-a603^-$ ,  $a603^+-a608^-$ ,  $b607^-a609^+$ , and  $b607^-b615^+$ . The calculated energies of the latter two states are reasonably close to the  $Q_y$  manifold, as the mean  $Q_y$  energy for Chls  $a$  is  $17543\text{cm}^{-1}$  and the lowest energy, corresponding to  $a602$ , is  $17226\text{cm}^{-1}$ . For simplicity, all calculations in this work were done using  $\epsilon_{\text{eff}} = 1$ . Note that  $\epsilon_{\text{eff}}$  is an effective scaling factor, accounting for both the dielectric screening of the protein environment and possible systematic errors in the evaluation of the atomic partial charges. Previous theoretical work often assumed  $\epsilon_{\text{eff}} = 1.5$ ; therefore, after performing calculations with  $\epsilon_{\text{eff}} = 1$ , we have also checked the

effect of its variation within the interval between 1 and 1.5. Assuming this value, the estimated energies of the latter two CT states would be much higher than the  $Q_y$  manifold, while the energies of the former two states would become  $E_{a602^+-a603^-} = 17381\text{cm}^{-1}$  and  $E_{a603^+-a608^-} = 17659\text{cm}^{-1}$ , which would put them in the middle of the  $Q_y$  manifold. A slightly smaller value of  $\epsilon_{\text{eff}} = 1.3$  would result in  $E_{a602^+-a603^-} = 15799\text{cm}^{-1}$  and  $E_{a603^+-a608^-} = 16604\text{cm}^{-1}$ , and the corresponding energy gap to the  $Q_y$  manifold would then become comparable to the red-shifted fluorescence peak. We thus propose that both of these states could contribute to the experimental fluorescence signal, although fine-tuning of the  $\epsilon_{\text{eff}}$  value in the future detailed theoretical modeling is needed to precisely describe the whole spectrum. The latter two (energetically higher) CT states will hardly contribute to the red-shifted fluorescence, but their presence might still be needed to correctly reproduce all the features of the experimental data, especially those observed in the Stark measurements (Romero et al., 2009).

Interestingly, previously only the  $a603$ – $a608$  dimer (note that it is often called  $a603$ – $a609$  dimer due to the prevalent usage of the LHCII nomenclature) was considered to be the origin of the relevant CT state in both experimental considerations (Ihalainen et al., 2005; Romero et al., 2009) and theoretical calculations (Novoderezhkin et al., 2016; Sláma et al., 2023). One of the reasons for that is that the Mg–Mg distance between Chls  $a603$  and  $a608$  is much shorter than between  $a602$  and  $a603$  (see Table 3). Clearly, a short enough distance between two pigments is a necessary, but not sufficient condition for a spectroscopically active CT state to occur therein. Indeed, even for the Lhca4 complex, the shortest distance between the two pigments corresponds to the  $a604$ – $b605$  dimer, yet no low-lying CT state is present there. At this point it is also worthwhile to point out a recent quantum-chemical work on the PSII reaction center (Sirohiwal et al., 2020b), where it was shown that no low-energy CT states could be identified between the so-called special pair pigments  $P_{D1}$  and  $P_{D2}$ , despite a very small spatial separation. Clearly, the environmental effects on the CT state energetics are of the paramount importance.

There is another argument that is often employed to claim that the  $a603$ – $a608$  dimer is the spatial origin of the CT states in the Lhca4 complex. It comes from site-directed mutagenesis. The amino acid coordinating the  $a603$  pigment can be changed from ASN to HIS (N98H mutant in our notation, often called N47H elsewhere in literature). It was demonstrated that for this mutant the fluorescence spectrum of the reconstituted Lhca4 loses the red-shifted peak almost entirely (Morosinotto et al., 2003). Very recently, a partial loss of the red-shifted fluorescence in this mutant was also demonstrated for intact PSI–LHCI supercomplexes (Li et al., 2023). In addition, recent advanced quantum-chemical calculations also demonstrated the loss of the red-shifted fluorescence for this Lhca4 mutant (Sláma et al., 2023). We have performed calculations for this mutant as well, but our results do not show any huge differences for the CT state energies. At first glance, this seems to make our results to be at odds with the literature data. That is not the case, however. What our calculations show, is that solely replacing ASN98 to HIS98 while keeping the

geometry of the nearby Chls the same, cannot account for huge energy shifts, neither for  $a602^+ - a603^-$ , nor for  $a603^+ - a608^-$  dimers. Therefore, protein electrostatics is not the main factor in this case. Indeed, Sláma et al. (2023) demonstrated that this mutation induces spatial shifts for  $a603$  pigment. It is these shifts that are the reason for different CT state energetics. We must stress that changes in the spatial position of Chl  $a603$  would also induce significant changes in the  $a602^+ - a603^-$  energetics. Thus, it might be that the N98H lacks both of these CT states, and this is the reason for the disappearance of the red-shifted fluorescence.

We must also highlight that our suggestion that the  $a602^+ - a603^-$  state also contributes to the experimentally observable red-shifted fluorescence of the Lhca4 complex is indirectly supported by experimental evidence. Morosinotto et al. (2005) investigated several Lhca4 mutants, one of them being E95V/R209L double mutant (E44V/R158L in the original notation of Ref (Morosinotto et al., 2005)). GLU95 coordinates  $a602$  and ARG209 is quite close to it, thus these mutations should affect the geometry of this pigment. Indeed, this mutant no longer fluoresces at  $\sim 730$  nm, though fluorescence at  $\sim 700$  nm remains (Morosinotto et al., 2005). Thus, it is very probable that  $a602$  is involved in one of the spectroscopically active CT states, which our work shows to be  $a602^+ - a603^-$ .

Let us now discuss the protonation pattern of the Lhca4 protein. According to our estimations, there are 31 charged amino acids at neutral pH, not counting the N- and C-terminus. Almost all the charged amino acids are quite close to the stromal or luminal surfaces. The majority of the charged amino acids are on the stromal side of the membrane. Two of the three amino acids not exhibiting the standard protonation pattern (GLU145 and GLU153) are neutral and are located deeper inside the protein, while HIS242 is positively charged and located near the luminal surface. Of all the charged amino acids, only the guanidino group of ARG156 quite far from the stromal surface and oriented towards the inside of the protein. We thus find that the estimated protonation should be reliable since both methods and software used are time-tested and proven to be credible and that the resulting protonation pattern does not contain any unexpected features.

Under physiological conditions, the protonation pattern of the amino acids results from a statistical process, and some of the amino acids could be easily (de)protonated. This, in turn, could shift the energy levels of the chlorophyll excited states, leading to different conformational states of the whole pigment-protein complex. Indeed, for Lhca4 the single-molecule spectroscopy measurements (Krüger et al., 2011) and time-resolved fluorescence measurements (Wientjes et al., 2011) suggest the coexistence of a few conformational states. We have investigated the possible non-standard protonation states of the Lhca4 complex by considering amino acids with an absolute difference between their  $pK_a$  and pH values not exceeding 2. The resulting energy shifts for the Chl  $Q_y$  states are presented in Supplementary Figure S1, while the shifts for the identified CT states are presented in Figure 4. We will now discuss these possibilities in more detail.

Our identification of the easily protonable amino acids was based on the corresponding  $pK_a$  values. Some additional consideration should be applied, however, because some of these

amino acids are coordinating the pigments in the Lhca4 complex. GLU95 coordinates Chl  $a602$ , GLU153 coordinates Chl  $a608$ , and HIS236 coordinates Chl  $a613$ . Thus we hold that it is rather unlikely for these amino acids to change their protonation state, since that might affect the surrounding geometry. GLU145 does not directly coordinate the Mg atom of any Chl, yet it coordinates the water molecule that coordinates Chl  $b605$  and it also coordinates the C7-formyl of Chl  $b606$ , thus it is also unlikely for this amino acid to change its protonation state. GLU113 and HIS242 are both on the periphery of the complex on the luminal side. According to our calculations, the change of the protonation state of these amino acids would not result in any significant energy shifts for both the Chl  $Q_y$  states and the CT states that could contribute to the red emission. ASP159 is on the stromal side of the complex, and if it were to acquire a proton, this would result in significant decrease of the  $b607^- - a609^+$  and  $b607^- - b615^+$  state energies, making them comparable to the energies of  $a602^+ - a603^-$  and  $a603^+ - a608^-$ . This could result in an increase of the red emission. On the other hand, the effects of the protonation state to the Chl  $Q_y$  state energies are minor. GLU202 and HIS222 are both on the periphery of the Lhca4 complex (on the stromal and luminal sides, respectively). The changes of their protonation states would result in either higher (in case of GLU202) or lower (in case of HIS222) energy of the  $a602^+ - a603^-$  state, which could affect the low energy emission in the fluorescence spectra. Regarding the Chl  $Q_y$  states, all the calculated energy shifts are less than  $100\text{cm}^{-1}$ , with the largest effect being an increase of  $b606$   $Q_y$  energy by  $84\text{cm}^{-1}$  in case of protonated HIS202.

There is one other issue that should be addressed. In the present work, similarly to many other works in literature (Frankcombe, 2015; Kavanagh et al., 2020; Sirohiwal et al., 2020b), we have considered chlorophyll dimers as a large supermolecule, and the identified CT states are then the eigenstates of this supersystem. The obtained CT states are thus not pure, but mixed (at least partially) with the excitonic states. This approach allows us to easily identify the energetically relevant CT states, but it must be kept in mind, that if the mixing with bright states is vanishingly small, such states would not receive oscillator strength; hence, these CT states would not be visible spectroscopically. Future studies thus should also consider the degree of mixing of pure CT states with pure site-basis excitations, which could provide the couplings of the CT states to the bright states that could then in principle be used to model the population transfer between all possible excited states. Such couplings could be obtained via the diabaticization procedure, as in Ref (Nottoli et al., 2018).

## 5 Conclusion

In this work, we have performed a systematic study of possible CT states in the Lhca4 complex. In vacuum, all the CT states have energies much higher than the  $Q_y$  states of the pigments. However, when the electrostatic effects of the protein environment are taken into account, the energies of some of CT states drop down significantly. Based on our calculations, we propose that in addition to the  $a603^+ - a608^-$  CT state, also the  $a602^+ - a603^-$  state

contributes to the far-red fluorescence signals of the Lhca4 complexes, which is supported by the experimental work on site-directed mutants. We have also investigated the possible protonation patterns of the titrable amino acids, and found that changes in protonation states of ASP159, GLU202 and HIS222 could affect the low energy emission. Future work should investigate the couplings between the pure CT states and molecular excited states, which would allow one to estimate the CT state populations.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Author contributions

GR: Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. AG: Conceptualization, Investigation, Methodology, Writing – original draft. BR: Conceptualization, Funding acquisition, Supervision, Writing – original draft. LV: Conceptualization, Methodology, Supervision, Writing – review & editing. JC: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Research Council of Lithuania (LMT grant

No. S-MIP-23–31 (AG, LV, JC) and LMT students' research grant No. S-ST-23–195 (GR)) and French National Agency for Research (Project Excit ANR-20-CE11–0022) (BR).

## Acknowledgments

Computations were performed using the resources of the High Performance Computing Center “HPC Sauletekis” at Faculty of Physics, Vilnius University.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1412750/full#supplementary-material>

## References

- Adolphs, J., Müh, F., Madjet, M. E.-A., and Renger, T. (2007). Calculation of pigment transition energies in the FMO protein: From simplicity to complexity and back. *Photosynth. Res.* 95, 197–209. doi: 10.1007/s11120-007-9248-z
- Adolphs, J., Müh, F., Madjet, M. E.-A., Schmidt am Busch, M., and Renger, T. (2010). Structure-based calculations of optical spectra of photosystem I suggest an asymmetric light-harvesting process. *J. Am. Chem. Soc.* 132, 3331–3343. doi: 10.1021/ja907222z
- Ben-Shem, A., Frolow, F., and Nelson, N. (2003). Crystal structure of plant photosystem I. *Nature* 426, 630–635. doi: 10.1038/nature02200
- Boekema, E. J., Jensen, P. E., Schlodder, E., van Breemen, J. F. L., van Roon, H., Scheller, H. V., et al. (2001). Green plant photosystem I binds light-harvesting complex I on one side of the complex. *Biochemistry* 40, 1029–1036. doi: 10.1021/bi0015358
- Bold, B. M., Sokolov, M., Maity, S., Wanko, M., Dohmen, P. M., Kranz, J. J., et al. (2020). Benchmark and performance of long-range corrected time-dependent density functional tight binding (LC-TD-DFTB) on rhodopsins and light-harvesting complexes. *Phys. Chem. Chem. Phys.* 22, 10500–10518. doi: 10.1039/C9CP05753F
- Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., et al. (2009). CHARMM: The biomolecular simulation program. *J. Comput. Chem.* 30, 1545–1614. doi: 10.1002/jcc.21287
- Case, D. A., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E. III, et al. (2020). *AMBER 2020* (San Francisco: University of California).
- Chmeliov, J., Gelzinis, A., Songaila, E., Augulis, R., Duffy, C. D. P., Ruban, A. V., et al. (2016). The nature of self-regulation in photosynthetic light-harvesting antenna. *Nat. Plants* 2, 16045. doi: 10.1038/nplants.2016.45
- Croce, R., Chojnicka, A., Morosinotto, T., Ihalainen, J. A., van Mourik, F., Dekker, J. P., et al. (2007). The low-energy forms of photosystem I light-harvesting complexes: Spectroscopic properties and pigment-pigment interaction characteristics. *Biophys. J.* 93, 2418–2428. doi: 10.1529/biophysj.107.106955
- Croce, R., Morosinotto, T., Castelletti, S., Breton, J., and Bassi, R. (2002). The Lhca antenna complexes of higher plants photosystem I. *Biochim. Biophys. Acta Bioenerg.* 1556, 29–40. doi: 10.1016/S0005-2728(02)00304-3
- Dennington, R., Keith, T. A., and Millam, J. M. (2019). *GaussView Version 6* (Shawnee Mission, KS: Semichem Inc.).
- Frankcombe, T. J. (2015). Explicit calculation of the excited electronic states of the photosystem II reaction centre. *Phys. Chem. Chem. Phys.* 17, 3295–3302. doi: 10.1039/C4CP04468A
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2016). *Gaussian 16 Revision C.01* (Wallingford CT: Gaussian Inc.).
- Gelzinis, A., Abramavicius, D., and Valkunas, L. (2015). Absorption lineshapes of molecular aggregates revisited. *J. Chem. Phys.* 142, 154107. doi: 10.1063/1.4918343
- Haworth, P., Watson, J. L., and Arntzen, C. J. (1983). The detection, isolation and characterization of a light-harvesting complex which is specifically associated with photosystem I. *Biochim. Biophys. Acta Bioenerg.* 724, 151–158. doi: 10.1016/0005-2728(83)90036-1
- Higashi, M., Kosugi, T., Hayashi, S., and Saito, S. (2014). Theoretical study on excited states of bacteriochlorophyll a in solutions with density functional assessment. *J. Phys. Chem. B* 118, 10906–10918. doi: 10.1021/jp507259g

- Ihalainen, J. A., Croce, R., Morosinotto, T., van Stokkum, I. H. M., Bassi, R., Dekker, J. P., et al. (2005). Excitation decay pathways of Lhca proteins: A time-resolved fluorescence study. *J. Phys. Chem. B* 109, 21150–21158. doi: 10.1021/jp0519316
- Ish-Shalom, D., and Ohad, I. (1983). Organization of chlorophyll–protein complexes of photosystem I in *Chlamydomonas reinhardtii*. *Biochim. Biophys. Acta Bioenerg.* 722, 498–507. doi: 10.1016/0005-2728(83)90066-X
- Jornet-Somoza, J., Alberdi-Rodriguez, J., Milne, B. F., Andrade, X., Marques, M. A. L., Nogueira, F., et al. (2015). Insights into colour-tuning of chlorophyll optical response in green plants. *Phys. Chem. Chem. Phys.* 17, 26599–26606. doi: 10.1039/C5CP03392F
- Jurinovich, S., Viani, L., Prandi, I. G., Renger, T., and Mennucci, B. (2015). Towards an *ab initio* description of the optical spectra of light-harvesting antennae: Application to the CP29 complex of photosystem II. *Phys. Chem. Chem. Phys.* 17, 14405–14416. doi: 10.1039/C4CP05647G
- Kavanagh, M. A., Karlsson, J. K. G., Colburn, J. D., Barter, L. M. C., and Gould, I. R. (2020). A TDDFT investigation of the photosystem II reaction center: Insights into the precursors to charge separation. *Proc. Natl. Acad. Sci.* 117, 19705–19712. doi: 10.1073/pnas.1922158117
- Knoetzel, J., Svendsen, I., and Simpson, D. J. (1992). Identification of the photosystem I antenna polypeptides in barley: Isolation of three pigment-binding antenna complexes. *Eur. J. Biochem.* 206, 209–215. doi: 10.1111/j.1432-1033.1992.tb16918.x
- Krüger, T. P. J., Wientjes, E., Croce, R., and van Grondelle, R. (2011). Conformational switching explains the intrinsic multifunctionality of plant light-harvesting complexes. *Proc. Natl. Acad. Sci. U. S. A.* 108, 13516–13521. doi: 10.1073/pnas.1105411108
- Lam, E., Ortiz, W., and Malkin, R. (1984). Chlorophyll *a/b* proteins of photosystem I. *FEBS Lett.* 168, 10–14. doi: 10.1016/0014-5793(84)80197-0
- Li, X., Zhu, L., Song, J., Wang, W., Kuang, T., Yang, G., et al. (2023). LHCA4 residues surrounding red chlorophylls allow for fine-tuning of the spectral region for photosynthesis in *Arabidopsis thaliana*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1118189
- Liu, Z. F., Yan, H. C., Wang, K. B., Kuang, T. Y., Zhang, J. P., Gui, L. L., et al. (2004). Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature* 428, 287–292. doi: 10.1038/nature02373
- Lu, T., and Chen, F. (2011). Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* 33, 580–592. doi: 10.1002/jcc.22885
- Mascoli, V., Liguori, N., Xu, P., Roy, L. M., van Stokkum, I. H., and Croce, R. (2019). Capturing the quenching mechanism of light-harvesting complexes of plants by zooming in on the ensemble. *Chem* 5, 2900–2912. doi: 10.1016/j.chempr.2019.08.002
- Mazor, Y., Borovikova, A., Caspy, I., and Nelson, N. (2017). Structure of the plant photosystem I supercomplex at 2.6 Å resolution. *Nat. Plants* 3. doi: 10.1038/nplants.2017.14
- Mazor, Y., Borovikova, A., and Nelson, N. (2015). The structure of plant photosystem I super-complex at 2.8 Å resolution. *eLife* 4, e07433. doi: 10.7554/eLife.07433
- Mikalciūtė, A., Gelzinis, A., Mačernis, M., Büchel, C., Robert, B., Valkunas, L., et al. (2022). Structure-based model of fucoxanthin–chlorophyll protein complex: Calculations of chlorophyll electronic couplings. *J. Chem. Phys.* 156, 234101. doi: 10.1063/5.0092154
- Morosinotto, T., Breton, J., Bassi, R., and Croce, R. (2003). The nature of a chlorophyll ligand in Lhca proteins determines the far red fluorescence emission typical of photosystem I. *J. Biol. Chem.* 278, 49223–49229. doi: 10.1074/jbc.M309203200
- Morosinotto, T., Mozzo, M., Bassi, R., and Croce, R. (2005). Pigment–pigment interactions in Lhca4 antenna complex of higher plants photosystem I. *J. Biol. Chem.* 280, 20612–20619. doi: 10.1074/jbc.M500705200
- Nottoli, M., Jurinovich, S., Cupellini, L., Gardiner, A. T., Cogdell, R., and Mennucci, B. (2018). The role of charge–transfer states in the spectral tuning of antenna complexes of purple bacteria. *Photosynth. Res.* 137, 215–226. doi: 10.1007/s11120-018-0492-1
- Novoderezhkin, V. I., and Croce, R. (2023). The location of the low-energy states in Lhca1 favors excitation energy transfer to the core in the plant PSI-LHCI supercomplex. *Photosynth. Res.* 156, 59–74. doi: 10.1007/s11120-022-00979-8
- Novoderezhkin, V. I., Croce, R., Wahadoszamen, M., Polukhina, I., Romero, E., and van Grondelle, R. (2016). Mixing of exciton and charge-transfer states in light-harvesting complex Lhca4. *Phys. Chem. Chem. Phys.* 18, 19368–19377. doi: 10.1039/C6CP02225A
- Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011). PROPKA3: Consistent treatment of internal and surface residues in empirical pK<sub>a</sub> predictions. *J. Chem. Theory Comput.* 7, 525–537. doi: 10.1021/ct100578z
- Petersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612. doi: 10.1002/jcc.20084
- Plumley, F. G., and Schmidt, G. W. (1987). Reconstitution of chlorophyll *a/b* light-harvesting complexes: Xanthophyll-dependent assembly and energy transfer. *Proc. Natl. Acad. Sci. U. S. A.* 84, 146–150. doi: 10.1073/pnas.84.1.146
- Premvardhan, L., Sandberg, D. J., Fey, H., Birge, R. R., Büchel, C., and van Grondelle, R. (2008). The charge-transfer properties of the S<sub>2</sub> state of fucoxanthin in solution and in fucoxanthin chlorophyll-*a* c<sub>2</sub> protein (FCP) based on Stark spectroscopy and molecular-orbital theory. *J. Phys. Chem. B* 112, 11838–11853. doi: 10.1021/jp802689p
- Qin, X., Suga, M., Kuang, T., and Shen, J.-R. (2015). Structural basis for energy transfer pathways in the plant PSI-LHCI supercomplex. *Science* 348, 989–995. doi: 10.1126/science.aab0214
- Renger, T., and Müh, F. (2013). Understanding photosynthetic light-harvesting: A bottom up theoretical approach. *Phys. Chem. Chem. Phys.* 15, 3348–3371. doi: 10.1039/C3CP43439G
- Romero, E., Mozzo, M., van Stokkum, I. H., Dekker, J. P., van Grondelle, R., and Croce, R. (2009). The origin of the low-energy form of photosystem I light-harvesting complex Lhca4: Mixing of the lowest exciton with a charge-transfer state. *Biophys. J.* 96, L35–L37. doi: 10.1016/j.bpj.2008.11.043
- Sirohiwal, A., Berraud-Pache, R., Neese, F., Izsák, R., and Pantazis, D. A. (2020a). Accurate computation of the absorption spectrum of chlorophyll *a* with pair natural orbital coupled cluster methods. *J. Phys. Chem. B* 124, 8761–8771. doi: 10.1021/acs.jpcc.0c05761
- Sirohiwal, A., Neese, F., and Pantazis, D. A. (2020b). Protein matrix control of reaction center excitation in photosystem II. *J. Am. Chem. Soc.* 142, 18174–18190. doi: 10.1021/jacs.0c08526
- Sláma, V., Cupellini, L., Mascoli, V., Liguori, N., Croce, R., and Mennucci, B. (2023). Origin of low-lying red states in the Lhca4 light-harvesting complex of photosystem I. *J. Phys. Chem. Lett.* 14, 8345–8352. doi: 10.1021/acs.jpclett.3c02091
- Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M., and Jensen, J. H. (2011). Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK<sub>a</sub> values. *J. Chem. Theory Comput.* 7, 2284–2295. doi: 10.1021/ct200133y
- Voityuk, A. A. (2013). Estimation of electronic coupling for photoinduced charge separation and charge recombination using the fragment charge difference method. *J. Phys. Chem. C* 117, 2670–2675. doi: 10.1021/jp311891h
- Wientjes, E., and Croce, R. (2011). The light-harvesting complexes of higher-plant photosystem I: Lhca1/4 and Lhca2/3 form two red-emitting heterodimers. *Biochem. J.* 433, 477–485. doi: 10.1042/BJ20101538
- Wientjes, E., van Stokkum, I. H. M., van Amerongen, H., and Croce, R. (2011). Excitation-energy transfer dynamics of higher plant photosystem I light-harvesting complexes. *Biophys. J.* 100, 1372–1380. doi: 10.1016/j.bpj.2011.01.030
- Yanai, T., Tew, D. P., and Handy, N. C. (2004). A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* 393, 51–57. doi: 10.1016/j.cplett.2004.06.011
- Yin, S., Dahlbom, M. G., Canfield, P. J., Hush, N. S., Kobayashi, R., and Reimers, J. R. (2007). Assignment of the q<sub>y</sub> absorption spectrum of photosystem-I from *Thermosynechococcus elongatus* based on CAM-B3LYP calculations at the PW91-optimized protein structure. *J. Phys. Chem. B* 111, 9923–9930. doi: 10.1021/jp070030p
- Zhang, J., and Lu, T. (2021). Efficient evaluation of electrostatic potential with computerized optimized code. *Phys. Chem. Chem. Phys.* 23, 20323–20328. doi: 10.1039/D1CP02805G