

# Determination of the protonation preferences of bilin pigments in cryptophyte antenna complexes

Marina Corbella,<sup>a</sup> Zi S. D. Toa,<sup>b</sup> Gregory D. Scholes,<sup>b</sup> F. Javier Luque,<sup>c</sup> and Carles Curutchet <sup>\*a</sup>

Received 00th January 20xx,  
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

The light-harvesting mechanisms of cryptophyte antenna complexes have attracted considerable attention due to their ability to exhibit maximal photosynthetic activity at very low-light conditions and to display several colors, as well as the observation of vibronic coherent features in their two-dimensional electronic spectra. However, detailed investigations on the interplay between the protein environment and their light-harvesting properties are hampered by the uncertainty related to the protonation state of the underlying bilin pigments. Here we study the protonation preferences of four types of bilin pigments including 15,16-dihydrobiliverdin (DBV), phycoerythrobilin (PEB), phycocyanobilin (PCB) and mesobiliverdin (MBV), which are found in phycoerythrin PE545 and phycocyanin PC577, PC612, PC630 and PC645 complexes. We apply quantum chemical calculations coupled to continuum solvation calculations to predict the intrinsic acidity of bilins in aqueous solution, and then combine molecular dynamics simulations with empirical  $pK_a$  estimates to investigate the impact of the local protein environment in the acidity of the pigments. We also report measurements of the absorption spectra of the five complexes measured in a wide range of  $pH$  in order to validate our simulations and investigate possible changes in the light harvesting properties of the complexes in the range of physiological  $pH$  found in the lumen ( $pH \sim 5-7$ ). The results suggest a  $pK_a > 7$  for DBV and MBV pigments in the  $\alpha$  polypeptide chains of PE545 and PC630/PC645 complexes, which are not coordinated to a negatively charged amino acid. For the other PEB, DBV and PCB pigments, which interact with a Glu or Asp side chain, higher  $pK_a$  values ( $pK_a > 8$ ) are estimated. Overall, the results support a preferential population of the fully protonated state for bilins in cryptophyte complexes at physiological conditions regardless of the specific type of pigment and local protein environment.

## 1 Introduction

In photosynthesis specialized light-harvesting pigment-protein complexes (PPCs) are used to capture incident sunlight and funnel its energy to the reaction center.<sup>1,2</sup> Cryptophytes are eukaryotic algae that live in marine and freshwater environments and represent important primary producers, able to capture sunlight through tunable linear tetrapyrrole chromophores known as bilins (see Fig. 1). Despite the apparent simplicity of their light-harvesting antennae

compared to other classes of algae, cryptophytes display several colors and exhibit maximal photosynthetic activity at very low-light conditions.<sup>3-5</sup> Depending on their absorption maximum they are divided in two classes, phycoerythrins (PE, red colored) and phycocyanins (PC, blue colored), and so far four types of PEs (PE545, PE555, two types of PE566) and five types of PCs (PC569, PC577, PC612, PC630 and PC645) have been reported.<sup>5</sup> The variability in spectral properties of these phycobiliproteins mainly arises from differences in the underlying bilin composition and content. Indeed, several types of bilins are present in cryptophytes, which differ in the number of double bonds and the degree of conjugation of the  $\pi$ -electron system (Fig. 1): 15,16-dihydrobiliverdin (DBV), phycoerythrobilin (PEB), phycocyanobilin (PCB), mesobiliverdin (MBV), bilin 584 and bilin 618.<sup>5</sup> Furthermore, the protein sequence also impacts the spectral properties of these complexes, as illustrated by differences in PC577 and PC612 spectra, which contain the same set of bilin pigments, six PCBs and two DBVs.

The remarkable photosynthetic activity of these algae has motivated a variety of studies aimed at unraveling their light

<sup>a</sup> Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica and Institut de Química Teòrica i Computacional (IQTCUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona Spain.

<sup>b</sup> Department of Chemistry, Princeton University, Washington Road, Princeton, New Jersey 08544, United State.

<sup>c</sup> Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Institut de Biomedicina (IBUB) and Institut de Química Teòrica i Computacional (IQTCUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Santa Coloma de Gramenet, Spain.

<sup>†</sup> Electronic Supplementary Information (ESI) available. See DOI: 10.1039/x0xx00000x

harvesting mechanisms,<sup>2,6</sup> in the quest for principles to be used in the development of artificial devices.<sup>7</sup> Among them, the observation of coherence phenomena at room temperature for the PE545 and PC645 complexes<sup>8</sup> and other photosynthetic systems has fueled an ongoing debate on the molecular origin of the observed oscillations and their role in the photosynthetic function.<sup>9</sup> An emerging consensus indicates that vibronic coherences can be supported by resonant vibrations, although their electronic or vibrational nature is still under debate.<sup>10–12</sup> Recent simulations on the PC645 complex, for example, suggest that energy transport occurs in an incoherent vibronic regime, a finding explained by the rather large reorganization energies of the pigments compared to the electronic couplings among them.<sup>13</sup> Cryptophyte antenna complexes are also interesting because of the observation of persistent vibronic coherences in "closed" forms of these antenna proteins (PC630, PC645) compared to "open" ones (PC577, PC612), where the insertion of a single residue induces a rotation of  $\sim 73$  degrees of the protein subunits leading to an almost two-fold symmetry in the quaternary structure.<sup>14</sup>

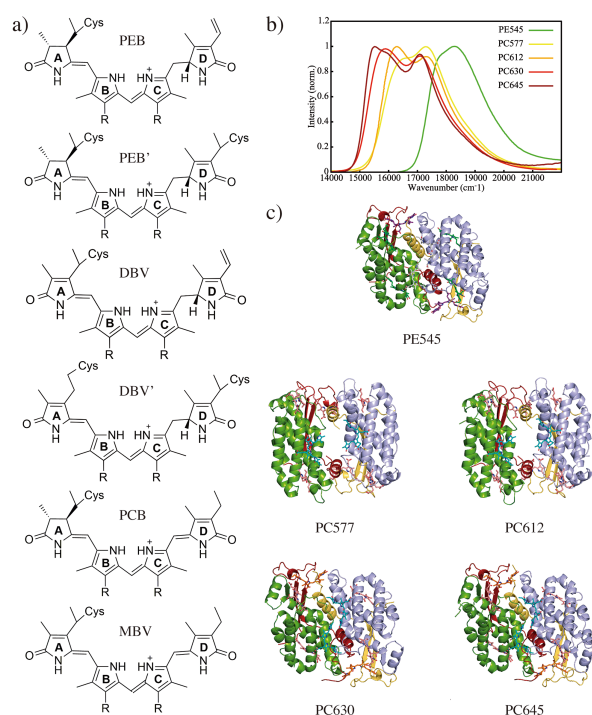


Figure 1. Bilin pigments present in the cryptophyte antenna complexes studied in this work. a) Bilin structures, b) absorption spectra, and c) protein structures of the PE545, PC577, PC612, PC630 and PC645 complexes. R stands for the propionic group.

Theoretical studies of energy transfer dynamics in antenna complexes and the role of coherence, however, still face considerable challenges related to the accurate determination of the ingredients needed to construct the exciton Hamiltonian of the system: the site energies, the electronic couplings and the spectral densities of the corresponding pigments.<sup>6</sup> An additional complication in the study of phycobiliproteins is the need to understand the protonation preferences of bilin pigments in the complex, as they can considerably impact their

electronic transition properties, and also their degree of coupling to the environment.<sup>15</sup> Moreover, protonation/deprotonation of the pigments is also important because, unlike other photosynthetic organisms where antenna complexes are bound to the thylakoid membrane, cryptophyte biliproteins are suspended in the lumen,<sup>16</sup> where the *pH* varies on the range  $\sim 5$ –7, depending on the prolongation of the incident sunlight.<sup>17</sup> The uncertainty regarding this issue leaves open questions: can the protonation state of the pigments change upon variations of available sunlight?, and can changes in the bilin protonation patterns alter the light-harvesting pathways and dynamics in the system?

Cryptophyte antennae are organized either as  $\alpha_1\beta\alpha_2\beta$  heterodimers or  $(\alpha\beta)_2$  homodimers. The crystal structures solved for several of these complexes (PE545, PE555, PC577, PC612, PC630 and PC645) indicate that all bilins in the highly-conserved  $\beta$  polypeptide chains present an Asp residue interacting with the central pyrrole rings B and C, thus suggesting a fully protonated form of the pyrrole rings.<sup>14,18,19</sup>

This would be consistent with research that supports a protonated form for the PCB in the C-Phycocyanin complex of cyanobacteria, where an Asp counterion also coordinates the B and C central pyrrole rings.<sup>20–22</sup> Spectroscopic studies performed on the similar phytochromobilin pigment in plant phytochromes also support a fully protonated tetrapyrrole.<sup>23</sup> Compared to the  $\beta$  chains, however, the structures of the  $\alpha$  subunits in cryptophyte antennae are much more divergent. In this case, the bilins in the  $\alpha$  subunits of PE555, PC577 and PC612 interact with a Glu side chain, but the DBVs in PE545 and the MBVs in PC645 and PC630 lack the presence of a negatively charged amino acid. In PE545, indeed, difference electron density maps obtained from ultrahigh resolution data at 0.97 Å seem to suggest a protonated form for the PEBs in the complex, which interact with an Asp residue, but can not distinguish among protonated or neutral forms for the DBVs.<sup>19</sup> The uncertainty regarding the protonation preferences of the bilins in cryptophyte antennae is reflected in a variety of theoretical studies performed on these systems, where the pigments are either assumed to be all protonated or unprotonated at physiological conditions.<sup>13,14,19,24–33</sup> Recently, the Coker group has studied the protonation state of the MBVs in PC645 by comparing experimental and theoretical spectra derived from quantum chemical calculations.<sup>15</sup> Their results indicate that the PC645 complex with unprotonated MBVs displays better maximum intensities in the two main absorption peaks compared to the protonated case. However, the low-energy edge of the predicted spectra is slightly too broad, so that the assignment of the protonation preferences still remains open.

In this contribution, we study the protonation preferences of the PEB, DBV, PCB and MBV pigments found in five cryptophyte phycobiliproteins (PE545, PC577, PC612, PC630 and PC645), shown in Fig. 1, by combining structure-based predictions of the  $pK_a$ 's of the pigments based on thermodynamic cycles with *pH*-dependent measures of their absorption spectra. Our results strongly suggest that all bilins

in these proteins are present in their protonated form at biologically relevant  $pH$  in the lumen ( $pH \sim 5-7$ ). We find that the MBVs in PC630 and PC645 as well as the DBVs in PE545, located at the  $\alpha$  polypeptide chains of the complexes and lacking the presence of a Glu counterion, are the most acid bilins in the complexes, but still have  $pK_a$  values  $> 7$ . Nevertheless, the changes observed in the absorption spectra suggest that a small fraction of deprotonated species of these pigments could be populated at  $pH$ 's  $\sim 7$ .

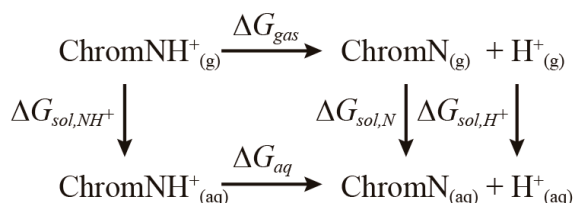
## 2 Methods

### 2.1 Thermodynamic cycles

Acid dissociation constants,  $K_a$ , are related to the reaction where an acid dissociates into its conjugate base and a proton in solution. The corresponding  $pK_a$  can thus be obtained from the Gibbs free energy change of this reaction,  $\Delta G_{aq}$ :

$$pK_a = \frac{\Delta G_{aq}}{RT \ln(10)} \quad (1)$$

In order to estimate accurate values of  $\Delta G_{aq}$ , we adopt the common thermodynamic cycle illustrated in Scheme 1, which combines quantum-mechanical (QM) calculations of the reaction free energy in vacuum,  $\Delta G_{gas}$ , with solvation free energies in water estimated using quantum-chemical continuum solvation models, as described in the next section.<sup>34,35</sup>



Scheme 1. Thermodynamic cycle used to calculate the aqueous deprotonation free energies of bilin pigments. ChromNH<sup>+</sup> and ChromN stand for the protonated and deprotonated states of the chromophore.

Thus, we calculate  $\Delta G_{aq}$  values using the following expression:

$$\Delta G_{aq} = \Delta G_{gas} + \Delta \Delta G_{sol} \quad (2)$$

where

$$\Delta G_{gas} = G_{gas}(\text{ChromN}) + G_{gas}(\text{H}^+) - G_{gas}(\text{ChromNH}^+) \quad (3)$$

and

$$\Delta \Delta G_{sol} = \Delta G_{sol}(\text{ChromN}) + \Delta G_{sol}(\text{H}^+) - \Delta G_{sol}(\text{ChromNH}^+) \quad (4)$$

We adopted the well-accepted value of  $G_{gas}(\text{H}^+)$  equal to -6.28 kcal/mol defined for a standard state of 1 atm, as well as the value of  $\Delta G_{sol}(\text{H}^+)$  equal to -265.9 kcal/mol, defined for the transfer from the gas-phase to aqueous solution at 1M standard state.<sup>34</sup>  $G_{gas}$  values computed for protonated/unprotonated forms of the bilins using QM methods also correspond to a standard state of 1 atm, whereas the corresponding  $\Delta G_{sol}$  terms calculated using continuum solvation models are defined for a gas-phase 1M

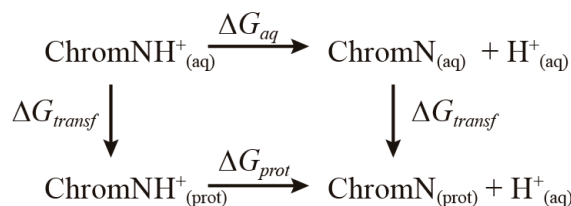
standard state. Thus,  $\Delta G_{gas}$  values were converted to a 1M standard state by adding the usual 1.89 kcal/mol term.<sup>34</sup>

Because the propionic acid groups can adopt multiple conformations with respect to the pyrrole rings, and therefore impact in different ways their deprotonation properties, the aqueous  $pK_a$  values corresponding to the deprotonation reaction on rings A, B, C or D of each type of bilin was computed for model pigments where the propionic groups were substituted by methyl units, as shown in Fig. 2. Later, the impact of the negatively charged propionic groups on the  $pK_a$  was introduced when computing the  $pK_a$  shift associated to the transfer of the bilin from water solution to the protein. Note that a similar strategy was recently adopted in order to compute the  $pK_a$  of biliverdin in the IFP2.0 fluorescent protein.<sup>36</sup> We thus compute the  $pK_a$ s first in water and then introducing the influence of the protein environment. This strategy has the advantage that carefully parametrized strategies to determine protonation states in biological environments are mostly based on water reference values and the determination of water-to-protein  $pK_a$  shifts.

The prediction of the  $pK_a$  of titratable groups or ligands inside a protein environment is a challenging task, and a variety of approaches have been proposed.<sup>37-39</sup> The usual strategy relies on estimating the free energies associated to the transfer of the titratable group from aqueous solution to the protein, as illustrated in Scheme 2. Then, these free energy contributions are combined to estimate the  $\Delta pK_a$  shift compared to an aqueous reference value:

$$pK_a(\text{prot}) = pK_a(\text{wat}) + \Delta pK_a(\text{wat} \rightarrow \text{prot}) \quad (5)$$

where  $pK_a(\text{wat})$  is the reference value adopted for the group and the  $\Delta pK_a$  term mainly involves electrostatic contributions due to interactions established with the protein environment as well as the desolvation cost.



Scheme 2. Thermodynamic cycle used to calculate the deprotonation free energies of bilin pigments in the protein environment.

Different strategies have been developed to compute the  $\Delta pK_a(\text{wat} \rightarrow \text{prot})$  term, mainly based on i) continuum solvation models, like Generalized Born (GB) or Poisson-Boltzmann (PB) electrostatic calculations, ii) molecular dynamics (MD)-based techniques coupled to free energy simulations, and iii) empirical models in which the shift is given as a sum of effective perturbation contributions, for example hydrogen bonds, charge-charge interactions or desolvation effects, whose weights are optimized by comparison to large sets of experimental data, like the popular PROPKA

method.<sup>40,41</sup> Here we adopt the fast PROPKA method to compute the  $\Delta pK_a$  shifts, which has been shown to provide  $pK_a$  values with similar accuracy compared to more costly molecular dynamics/generalized Born/thermodynamic integration (MD/GB/TI) techniques, with root-mean-squared deviations of  $\sim 1.4$   $pK_a$  units.<sup>38</sup> In PROPKA, the  $\Delta pK_a$  shift is expressed as the sum of a Coulomb contribution due to charge-charge interactions with all other charged or ionizable groups, and a self-energy contribution, representing the remaining contributions obtained when all other charged or ionizable groups are kept in their neutral form. The latter involves desolvation and intrinsic electrostatic energy contributions, like hydrogen-bonding interactions and other unfavorable electrostatic interactions. The resulting coupled titrations are then solved iteratively using a Monte Carlo scheme. Moreover, we average  $pK_a$  estimates over conformational ensembles sampled from MD simulations of the complexes in order to account for thermal fluctuations. For the sake of comparison, in addition to PROPKA estimates we have also computed the  $pK_a$  shifts using an electrostatic continuum solvation model-based approach, the APBS method developed by Baker and co-workers, which relies on continuum PB electrostatic calculations.<sup>42</sup>

## 2.2 Computational details

**Structures of the cryptophyte complexes.** Calculations were performed based on the X-ray crystal structures reported by Curmi and co-workers for phycoerythrin PE545 from *Rhodomonas* sp. CS24 (PDB code 1XG0, resolution 0.97 Å),<sup>18,19</sup> phycocyanin PC645 from *Chroomonas* sp. CCMP270 (PDB code 4LMS, resolution 1.35 Å),<sup>14</sup> phycocyanin PC612 from *Hemiselmis virescens* CCAC 1635 B (PDB code 4LM6, resolution 1.70 Å),<sup>14</sup> as well as phycocyanin PC577 from *Hemiselmis pacifica* CCMP706 (resolution 1.00 Å) and PC630 from *Chroomonas* CCAC 1627 B (resolution 1.60 Å), which have not been deposited yet. PE545, PC630 and PC645 are organized as

$\alpha_1\beta\alpha_2\beta$  heterodimers, whereas PC577 and PC612 are  $(\alpha\beta)_2$  homodimers and display an almost perfect two-fold symmetry. In PE545, each  $\alpha$  chain binds one DBV covalently linked to one Cys residue (DBV<sub>19A</sub> and DBV<sub>19B</sub>), whereas each  $\beta$  chain binds two PEBs linked to a single Cys (PEB<sub>158C</sub>, PEB<sub>158D</sub>, PEB<sub>82C</sub> and PEB<sub>82D</sub>) and one PEB linked to two Cys residues (PEB<sub>50/61C</sub> and PEB<sub>50/61D</sub>). In PC577 and PC612, on the other hand, each  $\alpha$  chain binds one singly linked PCB (PCB<sub>20A</sub> and PCB<sub>20C</sub>) and each  $\beta$  chain binds two singly linked PCBs (PCB<sub>158B</sub>, PCB<sub>158D</sub>, PCB<sub>82B</sub> and PCB<sub>82D</sub>) and a doubly linked DBV (DBV<sub>50/61B</sub> and DBV<sub>50/61D</sub>). Finally, in PC630 and PC645 the  $\beta$  chains bind the same pigments as in PC577 and PC612, whereas the  $\alpha$  chains binds one singly linked MBV (MBV<sub>19A</sub> and MBV<sub>19C</sub>) instead of a PCB. In PE545, the missing residues in the crystal were added as reported previously,<sup>26</sup> whereas in the PC577 and PC612  $(\alpha\beta)_2$  homodimers missing residues were added by homology with the other symmetric polypeptide chain. In PC630 and PC645, we simply capped polypeptide chains using an acetyl group due to missing residues in the N-terminal end of the chains.

**Quantum chemical calculations.** Starting from the geometries of bilins as found in the crystal structures, we defined simplified models in order to compute their acidity/basicity in gas phase and in water solution. Such models included the full tetrapyrrolic structure, in which the propionic side chains of rings B and C were replaced by methyl groups to avoid the folding of the chromophore and thus maintain its biological conformation, whereas the thioether linkage was replaced by a methyl group. In order to perform the aqueous  $pK_a$  calculations, we selected a single geometry for each type of pigment PEB, PEB', DBV, DBV', PCB and MBV, where the prime indicates bilins doubly-linked to the protein through Cys residues. The optimized geometries of each model bilin, aligned with the multiple conformations found for each type of pigment in the crystal structures, are shown in Fig. 2. As can be observed, these geometries keep the conformations of the

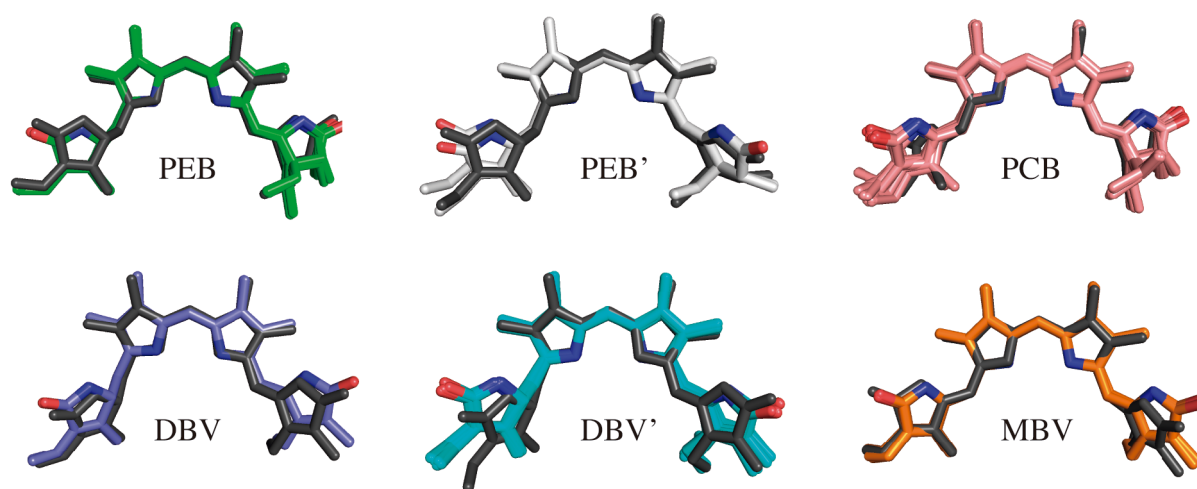


Figure 2. Structures of the bilin pigments found in PE545, PC577, PC612, PC630 and PC645 aligned to the quantum-chemical optimized geometries used for the prediction of the aqueous  $pK_a$ . The optimized geometries are displayed in black.

pigments in their specific protein environments.

The geometries of fully protonated and deprotonated (at rings A, B, C or D) forms of each model chromophore (PEB, PEB', DBV, DBV', PCB and MBV) were optimized at the B3LYP/6-31G(d) level of theory. The resulting minima were verified by vibrational frequency analysis, and corresponding free-energy corrections ( $G_{corr}$ ) at 298.15 K were obtained by scaling the vibrational frequencies by 0.977, as recommended for this level of theory.<sup>43</sup> Subsequently, the electronic energies were refined by combining SCS-MP2 energies extrapolated at the complete basis set (CBS) limit with CCSD corrections obtained with a reduced basis set. The spin-component-scaled version of MP2 provides a significant improvement on ground state energies by scaling parallel and antiparallel-spin pair correlation energies:<sup>44</sup>

$$E_{SCS-MP2} = E_{HF} + \frac{1}{3}(E_{corr(\alpha-\alpha)} + E_{corr(\beta-\beta)}) + \frac{6}{5}E_{corr(\alpha-\beta)} \quad (6)$$

The CBS energy was then extrapolated from SCS-MP2 correlation energies computed using Dunning's cc-pVDZ and cc-pVTZ basis sets according to the formula proposed by Halkier:<sup>45</sup>

$$E_{SCS-MP2/CBS} = E_{HF/cc-pVTZ} + \frac{E_{SCS-MP2/cc-pVTZ}^3 - E_{SCS-MP2/cc-pVDZ}^3}{3^3 - 2^3} \quad (7)$$

Finally, in order to recover the correlation energy between MP2 and CCSD, we added a correction given by the difference between CCSD/6-31G(d) and MP2/6-31G(d) calculations.

$$\Delta E_{CCSD} = E_{CCSD/6-31G(d)} - E_{SCS-MP2/6-31G(d)} \quad (8)$$

Thus, the final gas-phase free energies for the protonated and deprotonated species were given by the following expression:

$$G_{gas} = E_{SCS-MP2/CBS} + \Delta E_{CCSD} + G_{corr} \quad (9)$$

Gas-phase free energies were then combined with hydration free energies computed using the MST<sup>46,47</sup> and SMD<sup>48</sup> continuum solvation models. MST and SMD calculations were performed at the B3LYP/6-31G(d) level of theory, in order to be consistent with the parametrization of these methods. In order to test the accuracy of the computational protocol we also computed the known  $pK_a$ 's of the heteroaromatic molecules pyridinium and imidazolium, which can also deprotonate an aromatic  $NH^+$  group like the bilins, following the same strategy. All quantum chemical calculations were performed using the Gaussian 09 software.<sup>49</sup>

**MD simulations and protein  $pK_a$  shifts.** We performed MD simulations for PC577, PC612, PC630 and PC645 solvated in pre-equilibrated TIP3P water boxes (buffer zone 15 Å) using the Leap module of the Amber14 suite of programs.<sup>50</sup> The ff14SB<sup>51</sup> and GAFF<sup>52</sup> force fields were used for the proteins and the bilins, respectively. Charges for PCB, MBV, and DBV' bilins in their fully protonated forms were derived from multiconformational RESP<sup>50</sup> fits performed at the HF/6-31G(d) level of theory on the multiple geometries of the pigments found in the complexes, where hydrogens were previously optimized and Cys links were treated as described previously for PE545.<sup>26</sup> Protein hydrogen atoms were added according to

PROPKA predictions,<sup>40</sup> which indicated standard protonation states for all titratable residues except for His22A in PC630 and His21A in PC645, which were considered in their fully protonated form. The four systems were initially minimized and then thermalized running 50 ps NVT and 150ps NPT simulations applying constraints on the protein and bilin atoms with a harmonic potential (500.0 kcal/mol-Å<sup>2</sup>). Then, we thermalized the complete system from 50K to 300K by running 50 ps NVT and 50 ps NPT simulations. Finally, we extended the simulations for 10 ns for production purposes. All runs were performed with Amber 14<sup>50</sup> using an integration time step of 2 fs together with the SHAKE algorithm to restrain all bonds involving hydrogen, periodic boundary conditions, the particle-mesh Ewald approach to account for long-range electrostatics and a nonbonded cutoff equal to 10 Å. For PE545 we used the MD simulation described in ref.<sup>26</sup> We then extracted 50 snapshots equally spaced along the trajectories to estimate the bilin  $pK_a$  values in the complexes. Water-to-protein  $pK_a$  shifts were computed using the empirical rules implemented in the PROPKA3.0<sup>40</sup> server, which have been extended to  $pK_a$  shifts of active site residues and ionizable ligands/cofactors in PROPKA3.1.<sup>41</sup> In these calculations, the propionic acid groups of the pigments were treated as additional titratable groups with a reference  $pK_a$  value of 4.5. In addition, we also computed  $pK_a$  shifts using a physics-based approach, the APBS method developed by Baker and co-workers, which uses electrostatic Poisson-Boltzmann calculations to evaluate the transfer free energies for the protonated and unprotonated chromophores.<sup>42</sup> In this case, the propionic acid groups were described as single formic acid groups, in order to account for the electrostatic effect exerted by them on the bilin  $pK_a$ s. We adopted the point charges of the Amber ff14SB force field for the protein, whereas the charges of the protonated and neutral bilin pigments, as well as the formic acid groups, were computed at the HF/6-31G(d) level of theory using the RESP approach<sup>50</sup> on the optimized geometries of the model pigments described in the previous section. We used the APBS recommended values for the dielectric constants of water and the protein equal to 78.5 and 20, respectively.

### 2.3 Absorption spectra

Concentrated samples of purified PE545, PC577, PC630 and PC645 isolated from *Rhodomonas* sp. CS24 (PE545), *Hemiselmis pacifica* CCMP706 (PC577), *Chroomonas* CCAC 1627 B (PC630) and *Chroomonas mesostigmatica* CCMP269 (PC645) were frozen at -20°C in 0.050 M sodium phosphate buffer (pH 6.5 for PE545; pH 7.5 for PC577, PC630 and PC645) until required for spectroscopic measurements.<sup>53</sup> We then prepared buffer solutions at varying pH levels by adjusting relative amounts of Na<sub>2</sub>HPO<sub>4</sub> and citric acid (pH 4.0, 4.6, 5.4), Na<sub>2</sub>HPO<sub>4</sub> and NaH<sub>2</sub>PO<sub>4</sub> (pH 6.5, 7.8) trizma HCl and trizma base (pH 8.2, 8.5, 9.0) and Na<sub>2</sub>CO<sub>3</sub> and NaHCO<sub>3</sub> (pH 9.4). We measured the pH levels to be 4.0 ± 0.1, 4.6 ± 0.1, 5.4 ± 0.1, 6.5 ± 0.1, 7.8 ± 0.1, 8.2 ± 0.1, 8.5 ± 0.1, 9.0 ± 0.1 and 9.4 ± 0.1 using a calibrated microelectrode (Mettler Toledo). 100 μL aliquots of concentrated protein were diluted into the different 1.5 mL volume buffer solutions. The protein and buffer solutions were

measured and dispensed using a micropipette to make the concentrations as equal as possible. Proteins were diluted to an optical density (O.D.) of  $\sim 0.05/\text{mm}$ . Solution-phase steady-state linear absorption spectra for each phycobiliprotein was then recorded using a Varian Cary 6000i UV-vis spectrometer with a resolution of 1 nm at room temperature.

### 3. Results and discussion

#### 3.1 Aqueous $pK_a$

The first-principles determination of aqueous  $pK_a$  values is considerably challenging due to potential inaccuracies in the calculation of the gas-phase free energies of the acid and the conjugated base, as well as on their solvation free energies, as their counterparts for the proton,  $G_{\text{gas}}(H^+)$  and  $\Delta G_{\text{sol}}(H^+)$ , are well tabulated. The main source of inaccuracies may be ascribed to the calculation of the solvation free energies of the ionic species, as the gas-phase free energies and the solvation free energies of neutrals can be computed with errors close to  $\sim 1$  kcal/mol if high-level correlated methods are combined with modern parametrizations of solvation models.<sup>34</sup> In contrast to neutrals, solvation free energies of charged species are larger in magnitude due to strong electrostatic solute-solvent interactions, thus increasing associated errors. In addition, this can lead to persistent specific interactions, like hydrogen bonds, not properly accounted for in continuum solvation models. Thus, the hydration free energy of ions estimated from continuum models may lead to increased errors of  $\sim 4$ –5 kcal/mol, though prediction of relative free energies may benefit from cancellation of errors. Because of this, we have chosen to use two alternative solvation models, the SMD model developed by Marenich and co-workers<sup>48</sup> and the MST model developed in Barcelona.<sup>46,47</sup>

Table 1. Gas-phase and aqueous proton dissociation free energies and corresponding  $pK_a$  values computed for pyridinium and imidazolium.

	MST			SMD		
	$\Delta G_{\text{gas}}^a$	$\Delta G_{\text{aq}}^a$	$pK_a$	$\Delta G_{\text{aq}}^a$	$pK_a$	$pK_{a,\text{exp}}^b$
Pyridinium	215.6	5.4	3.9	5.4	3.9	5.2
Imidazolium	219.2	10.1	7.4	8.7	6.4	7.0

<sup>a</sup>Free energies (kcal/mol) corresponding to standard states of 1 atm in gas phase ( $\Delta G_{\text{gas}}^0$ ) and 1 M in aqueous solution ( $\Delta G_{\text{aq}}^*$ ). <sup>b</sup>Experimental values from Ref<sup>54</sup>.

In order to test the validity of our computational protocol, we first computed the  $pK_a$  of two aromatic heterocyclic compounds with known acidity, pyridinium and imidazolium, which, like the bilins, can be deprotonated on  $NH^+$  groups without breaking the aromaticity of the heterocycle. The results, reported in Table 1, support the reliability of the computational protocol, with absolute mean errors of  $\sim 1$   $pK_a$  unit for both solvation models (contributions to the proton dissociation free energies are reported in Table S1 of the Supporting Information). This indicates the absence of important systematic deviations in the prediction of absolute

$pK_a$  values, an error often found to depend on the nature of the functional group.<sup>34,55</sup> Thus,  $pK_a$  predictions for the bilin chromophores can be roughly expected to have an accuracy of  $\sim 1$ –2  $pK_a$  units.

For the bilins, we have computed the  $pK_a$  using the same protocol and adopting solvation free energies averaged over MST and SMD-based predictions. Table 2 shows the aqueous  $pK_a$  predicted for each pyrrole ring of the bilin chromophores, as well as the corresponding gas-phase and aqueous free energies (the individual contributions to these free energies are shown in Table S2). In all cases the outer pyrrole rings A and D present larger  $pK_a$  values (from 11.9 to 31.8) than those of the central B and C rings, which lie in the range 5.8–7.4. The amide character of the nitrogens in the outer rings explains this result, as deprotonation would prevent the delocalization of the nitrogen lone pair toward the carbonyl group and the corresponding resonance stabilization. However, the  $pK_a$  values of outer rings show three characteristic behaviours, depending on i) whether the ring is linked through a methine or a methylene bridge to the central pyrrole ring, and ii) the oxidation state of the ring. In the first group, the D rings of PEB and DBV (as well as doubly linked PEB' and DBV') are not conjugated with the other rings because of the methylene bridge, leading to a  $pK_a$  of 30–32. This can be explained by the lack of the positive charge delocalization onto this ring in PEB and DBV, whereas in PCB and MBV the positive charge can delocalize along the four pyrroles, including ring D (note here, however, that for the sake of simplicity only a single Lewis resonance structure is shown in Fig. 1). The A rings of PEB, PEB' and PCB, in contrast, are the least basic outer pyrroles, given their less oxidized state compared to the A rings in the other bilins, which leads to  $pK_a$  values of 12–13. Finally, the other outer rings (A ring of DBV, DBV' and MBV, D ring of PCB and MBV) display a  $pK_a$  of 16–17.

Overall, the results clearly indicate that the central B and C rings are the most acidic ones, with a  $pK_a$  of 6–7 in aqueous solution. The PCB pigment seems to be the least acidic, with a  $pK_a$  value of 7.1 (7.4) for the B (C) ring. The other pigments show similar acidities, the DBV one being slightly more acidic, especially the doubly linked DBV' one, in which the  $pK_a$  of ring C is lowered from 6.3 to 5.8 compared to the singly linked DBV. The results also indicate minor differences regarding the acidity of rings B and C, suggesting that at basic  $pH$  there could be a mixture of both deprotonated tautomers. On the other hand, the protein environment might play a key role in determining the relative stability of these tautomers as well as of the fully protonated form, in a situation reminiscent of the influence exerted by the local environment on the  $pK_a$  of His side chains in proteins.

Recent predictions by Feliks and Field on the biliverdin pigment of the IFP2.0 fluorescent protein found slightly more acidic  $pK_a$  values by about  $\sim 4$ –5  $pK_a$  units compared to our results for the similar MBV bilin.<sup>36</sup> Nevertheless, biliverdin differs from MBV by the presence of two vinyl groups instead of ethyl groups linked to rings A and D as well as by the orientation of the A ring. Moreover, their calculations included the propionic groups and involved an empirical correction

calibrated on a training set of acids. Nevertheless, it is worth noting that a similar protocol was adopted in that study based on DFT and single point MP2/CBS calculations. However, here we further refine the MP2/CBS values by adopting the SCS variant of MP2 and by adding CCSD corrections, two effects that contribute by ~50% each in increasing the  $pK_a$  estimates of the MBV central rings by ~3.5  $pK_a$  units, thus illustrating the importance of such corrections and partially explaining the discrepancy with our results. For example, our predictions for pyridinium and imidazolium show absolute mean errors of ~1  $pK_a$  unit, as discussed before, but degrade considerably to ~3  $pK_a$  units if SCS and CCSD corrections are omitted.

Table 2. Gas-phase and aqueous proton dissociation free energies and  $pK_a$  values computed for the deprotonation reaction on different pyrrole rings of bilin molecules.

	Ring	$\Delta G_{gas}^{\circ}$ <sup>a</sup>	$\Delta G_{aq}^{\circ}$ <sup>a</sup>	$pK_a$
PEB	A	249.4	18.2	13.3
	B	235.4	9.3	6.8
	C	234.1	8.9	6.5
	D	290.4	41.0	30.0
PEB'	A	248.2	17.0	12.4
	B	234.9	9.5	6.9
	C	233.9	8.8	6.4
	D	292.1	42.3	31.0
DBV	A	255.3	23.7	17.4
	B	234.3	9.2	6.7
	C	233.1	8.6	6.3
	D	288.4	40.7	29.9
DBV'	A	255.9	24.1	17.6
	B	234.5	9.2	6.7
	C	232.6	7.9	5.8
	D	290.4	43.4	31.8
PCB	A	250.2	16.2	11.9
	B	238.4	9.7	7.1
	C	238.7	10.1	7.4
	D	258.1	23.1	17.0
MBV	A	256.6	22.9	16.7
	B	236.7	9.1	6.6
	C	236.6	9.0	6.6
	D	256.2	21.9	16.1

<sup>a</sup>Free energies (kcal/mol) corresponding to standard states of 1 atm in gas phase ( $\Delta G_{gas}^{\circ}$ ) and 1 M in aqueous solution ( $\Delta G_{aq}^{\circ}$ ), the latter computed by using the solvation free energies averaged from MST and SMD calculations.

### 3.2 Protonation preferences in the protein complex

In this section we investigate the impact of the specific environment of the bilins in five different cryptophyte complexes. Our predictions suggest that the model bilin pigments (lacking the propionic acid groups) have aqueous  $pK_a$  values around ~6–7, corresponding to deprotonation of either B or C central pyrrole rings. Thus, the protein environment and the relative orientation of the propionic acid groups may play a key role in determining the relative stability of the protonated form and of the two unprotonated tautomers. Table 3 reports the  $pK_a$  values computed for bilin pyrrole rings B and C in their protein environments, computed as the sum of the aqueous  $pK_a$  results discussed in the previous section and the  $\Delta pK_a(wat \rightarrow prot)$  shifts computed by using PROPKA and APBS methods (Eq. 5). Because MD-averaged values computed using PROPKA displayed minor differences compared to the crystal (see Table S3 of the Supporting Information), we performed APBS calculations based on the crystal structure. In addition, in Tables S4 and S5 of the Supporting Information we show the  $pK_a$  values estimated using PROPKA for the bilin propionic groups and selected titratable residues directly interacting with the bilins.

The predicted  $pK_a$  values for the bilin pyrroles in the five cryptophyte complexes span a range between ~5 and 8  $pK_a$  units, according to PROPKA, and between ~7 and 11  $pK_a$  units, according to APBS. Thus, we find a systematic disagreement of ~2  $pK_a$  units between the two methods. Despite this discrepancy, there is high similarity in the relative  $pK_a$  values determined using PROPKA and APBS calculations. The range of values estimated thus indicates that the protein environment and the propionic acid groups tune in slightly different ways the aqueous  $pK_a$ 's of bilins, which are in the range ~6–7 (Table 2). Interestingly, in most bilins the acidity of rings B and C is very similar, with differences less than 0.5  $pK_a$  units. However, for doubly linked pigments DBV<sub>50/61B</sub> and DBV<sub>50/61D</sub> (PC577, PC612, PC630 and PC645 complexes) there is a clear trend indicating a  $pK_a$  smaller by ~1 unit for ring C compared to B, thus suggesting that deprotonation should only occur in pyrrole unit C, a preference that was already observed in the acidities of DBV' in water (Table 2). Note also that in the symmetric proteins PC577 and PC612, which are organized as ( $\alpha\beta$ )<sub>2</sub> homodimers, the symmetric pairs of chromophores located in the A/C and B/D polypeptide chains present very similar acidity constants, as one would expect. The impact of thermal fluctuations on the acidity constants is found to be small. Indeed, most PROPKA  $pK_a$  values presented standard deviations of ~0.1–0.2 units along the MD trajectories, and the estimates are close to those derived from the crystal structure of the complexes, with differences below 0.5 units in most cases (see Table S3). The only remarkable change is found for PEB<sub>158D</sub> in PE545, in which the crystal structure leads to an unexpected  $pK_a$  of 5.2 compared to the MD-averaged value of 7.2, which is closer to the value of 7.5 found for the similar PEB<sub>158C</sub> bilin. In contrast, APBS estimates based on the crystal structure were consistent throughout the whole set of bilins. On the other hand, we find  $pK_a$  values in the range 2.5–5 for all

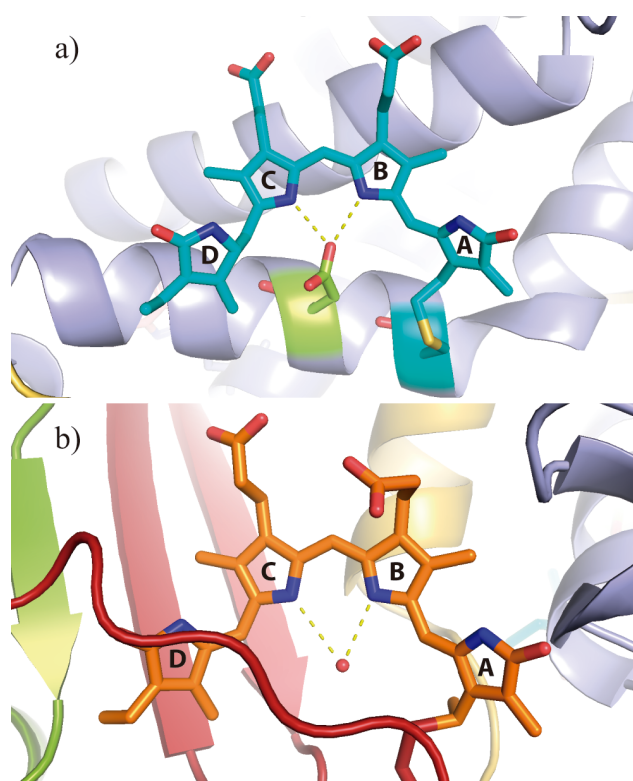


Figure 3. Representation of the environment surrounding bilin pigments. a) DBV<sub>50/61D</sub> pigment in the PC645 complex with central pyrrole rings interacting with Asp54D. b) MBV<sub>19A</sub> pigment in the PC645 complex with central pyrrole rings interacting with a crystallographic water.

bilin propionic acid groups (Table S4), suggesting that they are unprotonated at physiological conditions.

By analyzing the details of the environment surrounding the bilin pigments on each complex, we can identify two different situations regardless of the nature of the chromophore that largely explain the trends in  $pK_a$  values. As it also happens in the C-PC complex of cyanobacteria,<sup>20–22</sup> the central pyrrole rings of all PEB, PCB and DBV bilins located in the highly-conserved  $\beta$  polypeptide chains are coordinated with an Asp, which stabilizes the protonated form. A similar explanation applies to PCB<sub>20</sub> bilins in PC577 and PC612 located in the more divergent  $\alpha$  subunits, which interact with a Glu. As reported in Table S5, our PROPKA simulations indicate  $pK_a$  values in the range 3.3–6.9 for these Asp and Glu residues, in all cases lower than those estimated for the bilin pyrrole rings they interact with. In these situations, the bilin  $pK_a$  values are in the range ~6–8 and ~8–11 as estimated by PROPKA and APBS calculations, respectively. In contrast, MBVs in PC630 and P645 and DBVs in PE545 lack the coordination with Asp or Glu, leading to  $pK_a$  values of ~5 and ~7–8 for PROPKA and APBS. In Fig. 3 we show the environment surrounding DBV<sub>50/61D</sub> in PC645, with its central pyrrole rings interacting with Asp54D, and MBV<sub>19A</sub>, in which the pyrroles are coordinated with a crystallographic water. It is also worth noting that in DBVs in PE545 and MBV<sub>19C</sub> in PC630 and PC645 there is a His residue interacting with the bilins, as shown in Fig. S1. In this case, a  $pK_a$  of ~4 is estimated using PROPKA for His16A/His16B in PE545, in agreement with the neutral state suggested by difference

electron density maps of the complex.<sup>19</sup> On the other hand, the  $pK_a$  of ~7.5 computed for His22A and His21A in PC630 and PC645 indicate a fully protonated form, consistent with its simultaneous interaction with a propionic group and a Glu observed in the crystal structure.

Table 3.  $pK_a$  values computed using MD-averaged PROPKA calculations and APBS calculations based on the crystal structure for bilin pyrrole rings B and C in their specific protein environments.

	PROPKA		APBS	
	$pK_a(B)$	$pK_a(C)$	$pK_a(B)$	$pK_a(C)$
PE545 ( $\alpha\beta\alpha\beta$ )				
DBV <sub>19A</sub>	5.6	5.7	8.1	7.7
DBV <sub>19B</sub>	5.8	5.6	8.2	7.8
PEB <sub>50/61C</sub>	7.3	7.1	10.7	10.1
PEB <sub>158C</sub>	7.5	7.6	9.6	9.3
PEB <sub>82C</sub>	6.7	6.9	9.0	8.8
PEB <sub>50/61D</sub>	7.4	7.5	10.9	10.5
PEB <sub>158D</sub>	7.2	7.2	10.2	9.9
PEB <sub>82D</sub>	6.9	7.0	9.1	8.8
PC577 ( $\alpha\beta\alpha\beta$ )				
PCB <sub>20A</sub>	8.0	7.8	10.7	10.7
DBV <sub>50/61B</sub>	7.6	6.7	11.4	10.5
PCB <sub>158B</sub>	7.9	8.2	10.2	10.4
PCB <sub>82B</sub>	6.9	7.3	9.5	9.8
PCB <sub>20C</sub>	7.9	7.8	10.4	11.0
DBV <sub>50/61D</sub>	7.5	6.6	11.6	10.9
PCB <sub>158D</sub>	7.9	8.2	10.3	10.5
PCB <sub>82D</sub>	7.0	7.4	9.4	9.7
PC612 ( $\alpha\beta\alpha\beta$ )				
PCB <sub>20A</sub>	8.1	7.9	9.3	9.4
DBV <sub>50/61B</sub>	7.5	6.6	9.2	8.8
PCB <sub>158B</sub>	7.9	8.1	9.3	9.4
PCB <sub>82B</sub>	7.0	7.4	8.9	9.2
PCB <sub>20C</sub>	8.1	8.0	9.4	9.4
DBV <sub>50/61D</sub>	7.7	6.8	9.6	8.9
PCB <sub>158D</sub>	7.9	8.1	9.4	9.5
PCB <sub>82D</sub>	7.0	7.4	8.8	9.1
PC630 ( $\alpha\beta\alpha\beta$ )				
MBV <sub>19A</sub>	5.6	5.0	7.2	7.1
DBV <sub>50/61B</sub>	7.3	6.3	9.3	8.4
PCB <sub>158B</sub>	7.9	8.3	9.3	9.6
PCB <sub>82B</sub>	7.0	7.3	8.8	9.0
MBV <sub>19C</sub>	5.4	4.9	8.0	8.1
DBV <sub>50/61D</sub>	7.2	6.4	9.9	9.2
PCB <sub>158D</sub>	7.6	7.9	9.5	9.7
PCB <sub>82D</sub>	7.1	7.3	9.0	9.3
PC645 ( $\alpha\beta\alpha\beta$ )				
MBV <sub>19A</sub>	5.9	5.3	7.9	7.8
DBV <sub>50/61B</sub>	7.3	6.3	9.1	7.9
PCB <sub>158B</sub>	7.8	8.0	9.5	9.7
PCB <sub>82B</sub>	7.1	7.4	9.0	9.2
MBV <sub>19C</sub>	5.5	5.0	8.3	8.3
DBV <sub>50/61D</sub>	7.3	6.3	9.6	8.7
PCB <sub>158D</sub>	7.6	8.0	9.4	9.6
PCB <sub>82D</sub>	7.0	7.4	9.3	9.5

### 3.3 Absorption spectra

Because cryptophyte biliproteins are immersed in the lumen,<sup>16</sup> where the  $pH$  varies in the range  $\sim 5$ – $7$  depending on the prolongation of the incident sunlight,<sup>17</sup> light exposure could potentially tune the light harvesting properties of these complexes given that protonation/deprotonation could be expected to occur in these conditions. However, as discussed in Sections 2.1 and 3.1, the accurate prediction of both absolute  $pK_a$  values in water and the corresponding shift upon transfer from water to a protein environment are still challenging, which may lead to uncertainties of  $\sim 1$ – $2$   $pK_a$  units on the acidities reported in Table 3. The relative  $pK_a$  values for the pigments, however, are expected to be more accurate, with errors within 1  $pK_a$  unit, as illustrated by the systematic discrepancy between PROPKA and APBS estimates, which otherwise lead to rather similar relative  $pK_a$ s. In order to validate our simulations and identify potential systematic errors in the predicted acidity constants, we have recorded the absorption spectra of four of the complexes (PE545, PC577, PC630 and PC645) at  $pH$  values ranging from 4.0 to 9.4. Our aim is to identify variations in the absorption spectra upon varying conditions and correlate them with changes in the protonation pattern of the pigments inside the proteins. In addition, these measurements allow us to directly test whether  $pH$  changes in the range  $\sim 5$ – $7$  as those occurring in the lumen could impact the absorption spectra of the complexes and thus their photosynthetic function upon prolonged light exposure.

The  $pH$ -dependent absorption spectra measured for the different complexes are shown in Fig. 4, whereas Fig. 5 shows the difference spectra relative to the  $pH$  6.5 curves in order to amplify the spectral changes. We additionally show the spectra obtained at  $pH$  4.0 and 9.4, in which the trends observed are less regular, in Fig. S2 of the Supporting Information. Note that at these  $pH$  values protonation/deprotonation of protein residues and partial unfolding of the complexes may affect the absorption spectra, making it more difficult to correlate absorption changes to variations in the protonation pattern of the pigments.<sup>56</sup> Indeed, the absorption spectra of all the proteins varies at  $pH$  4, matching the approximate  $pK_a$  of Asp and Glu side chains, but after the eventual deprotonation of these aminoacids all spectra remain mostly unchanged as  $pH$  is increased from 4.6 to 6.5. Thus, it seems unlikely that any change in protonation pattern occurs in this range of  $pH$ . In contrast, we find some appreciable changes in all absorption spectra in basic conditions as the  $pH$  increases until 9, especially for PC630 and PC645 complexes, but also to some extent in PE545 and PC577.

The changes observed in PC645 indicate that the relative intensity of the low energy band is reduced. This is consistent with the spectral changes predicted recently by Lee and co-workers arising from deprotonation of MBVs,<sup>15</sup> which increases the energy of the  $\pi \rightarrow \pi^*$  transition. As expected by the lack of a coordinating anionic residue, MBV pigments are predicted to be the least basic bilins in the complex. In this case, PROPKA calculations estimate a  $pK_a \sim 5$  for MBVs,

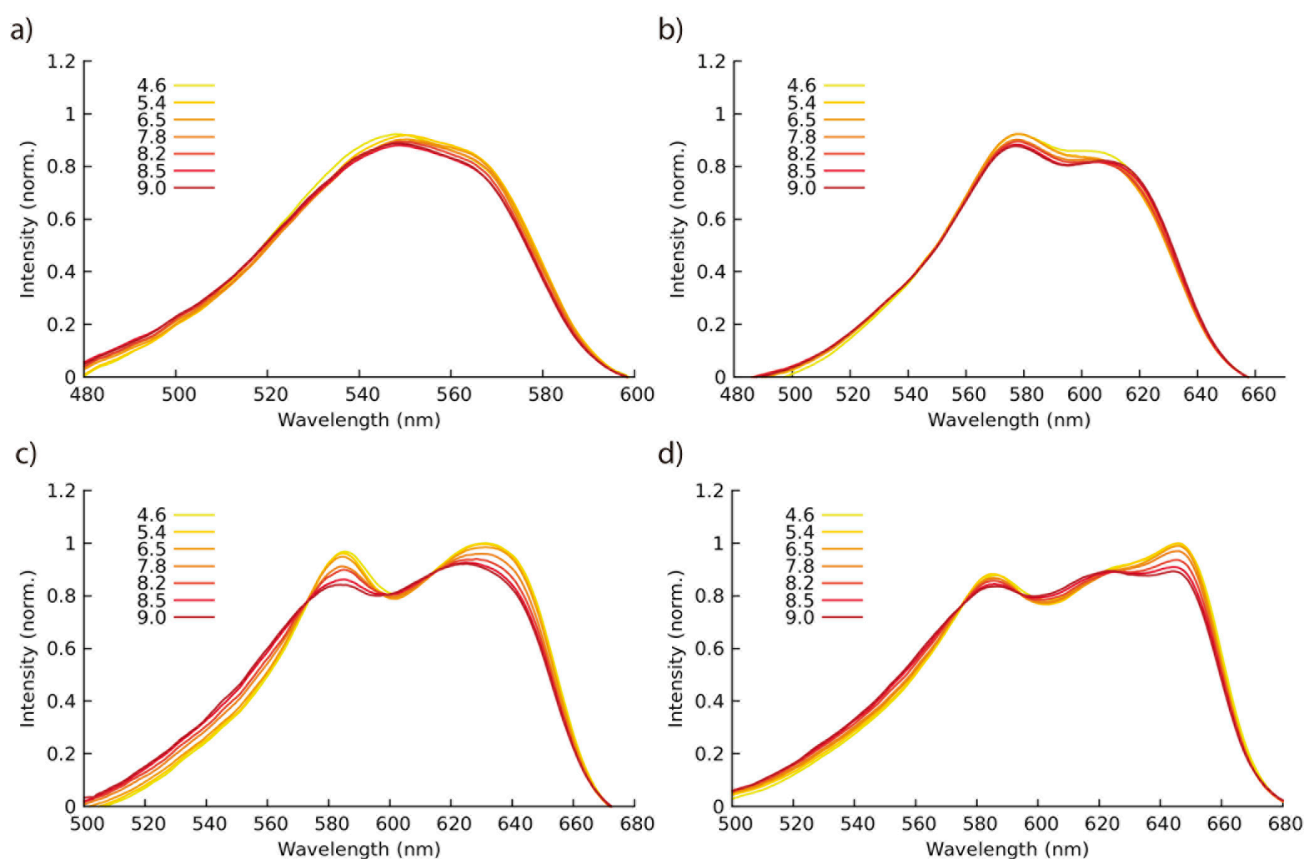


Figure 4. Normalized absorption spectra of cryptophyte antenna complexes at different  $pH$  values. a) PE545, b) PC577, c) PC630 and d) PC645.

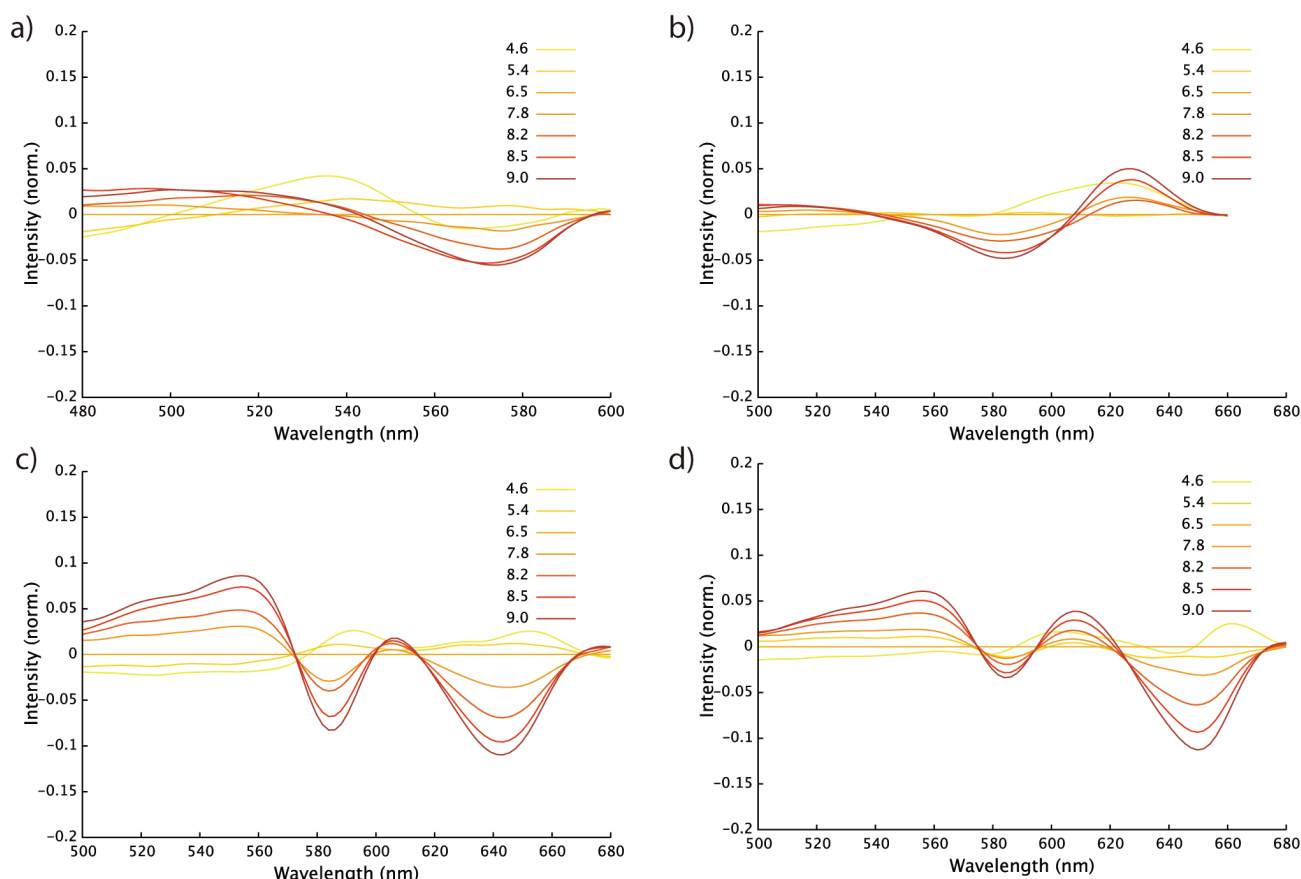


Figure 5. Normalized difference absorption spectra of cryptophyte antenna complexes at different  $pH$  values compared to the spectra at  $pH$  6.5. a) PE545, b) PC577, c) PC630 and d) PC645.

whereas APBS electrostatic calculations suggest a  $pK_a \sim 8$ , consistent with the spectral changes occurring at  $pH > 7$ . Thus, the observed spectral changes suggest that PROPKA predictions are underestimating the  $pK_a$  of these pigments by at least 2 units, which we deem reasonable, whereas APBS values seem to be rather accurate. Another possibility could be that MBVs are indeed already deprotonated in acidic conditions, as suggested by PROPKA  $pK_a$  estimates  $\sim 5$ , so that the spectral changes might arise from an eventual deprotonation of the central DBV's in the complex, predicted to be slightly more basic than MBVs with a  $pK_a \sim 6.3$ . This possibility, however, is inconsistent with the fact that the spectra remains almost unchanged in the 4.6 – 7.8  $pH$  range, as this would suggest a larger difference in acidity among DBVs and MBVs. Moreover, the DBVs contribute to the high-energy band of the absorption,<sup>15,24,25</sup> so we would expect the largest spectral changes to occur there instead of the low-energy band. Thus, all bilins, including MBVs lacking the anionic counterion that interacts with the central pyrrole rings, seem to be protonated at the  $pH$  range found in the lumen for PC645. These findings are in line with negligible changes in both spectra and energy transfer dynamics measured for PC645 using two-dimensional electronic spectroscopy at  $pH$  values 5.7, 6.5 and 7.4.<sup>57</sup>

In the PC630 complex, we find a similar trend in spectral changes compared to PC645, with appreciable changes in the spectra occurring only at  $pH$  values  $> 7$ . As for PC645, we

expect that deprotonation of MBVs in the complex, rather than the central DBVs, is responsible for the changes observed, given that protonation of the DBVs and the consequent increase in transition energy should mostly impact the high-energy band of the spectra, whereas MBVs are expected to contribute to the low energy bands. In contrast to PC645, in PC630 the rise in  $pH$  decreases the relative intensities of both high and low-energy bands compared to the medium-energy region. In addition, there is a slight shift to shorter wavelengths both at the red and blue edges of the spectrum, leading to an increase in the overall broadening. Because the spectra in PC630 is more congested than for PC645, the shift in MBV energies seems to have a larger impact on the total broadening of the absorption, thus causing an overall decrease of the intensity in the normalized spectra shown in Fig. 4, which affects both low and high-energy bands.

In contrast to PC630 and PC645, in PE545 and PC577 the absorption is found to be less sensitive to changes in acidity. Nevertheless, the trends are similar given that appreciable changes in the spectra are only found when the  $pH$  is increased over 7. This finding supports the idea that most bilins are fully protonated in acidic conditions, whereas some of them can deprotonate at  $pH$  values  $\sim 7$  or larger. As reported in Table 3, the more acidic bilins in PE545 are the central DBV<sub>19A</sub> and DBV<sub>19B</sub> pigments, which lack an anionic amino acid counterion, with a  $pK_a$  5.6 and 7.7 according to PROPKA and APBS calculations, respectively. The variation

found at basic conditions on the low energy band are in agreement with the expected impact of deprotonation of the DBVs, which mainly contribute to the red-edge.<sup>28,58</sup> This interpretation again suggests that PROPKA  $pK_a$  predictions are underestimated by at least  $\sim 2$  units, whereas APBS results seem to be rather accurate. Otherwise, if most PEB pigments in PE545 would have a  $pK_a \sim 7$  as suggested by PROPKA results reported in Table 3, we would expect more drastic changes in the spectra when the  $pH$  is increased from 4.6 to 9.0. As shown in Fig. S2, a drastic change is indeed observed at  $pH$  9.4, which could be partially caused by deprotonation of most PEBs in the complex. Thus, our results support a fully protonated form of the DBV pigments lacking the anionic counterion at the  $pH$  ranges  $\sim 5$ –7 found in the lumen, although some small fraction of these pigments could be present in their deprotonated form.

On the other hand, in PC577, as in PC612, all pigments in the complex are coordinated either to Asp or Glu. In this case, the central DBV<sub>50/61B</sub> and DBV<sub>50/61D</sub> pigments are predicted to have the lowest  $pK_a$ , close to  $\sim 6.7$  according to PROPKA, whereas APBS simulations predict  $pK_a$ s  $\sim 8.8$  and  $\sim 10.7$  in PC612 and PC577, respectively. In PC577, indeed, several PCBs are found to be more acidic than the DBVs. The changes observed in the PC577 spectra are again consistent with our overall interpretation, as we find almost no changes when the  $pH$  is increased from 4.6 to 6.5, and only a small decrease in the relative intensity of the maximum absorption band located at  $\sim 577$  nm when the  $pH$  is further increased until 9.0. Because the central DBVs are expected to contribute to this high-energy band,<sup>5</sup> these spectral changes could be caused by a progressive deprotonation of the DBV pigments at basic conditions, but it is also possible that the changes reflect partial deprotonation of the PCBs. Nevertheless, the results support again the accuracy of APBS estimates. Due to the high similarity between PC612 and PC577 in terms of structure, spectra and predicted  $pK_a$ , we expect the same situation to happen in PC612.

## Conclusions

In this study we have addressed the protonation preferences of four types of bilin pigments (PEB, PCB, MBV and DBV) found in different antenna complexes from cryptophyte algae, including three "closed" (PE545, PC630 and PC645) and two "open" forms (PC577, PC612). We applied SCS-MP2 quantum chemical calculations extrapolated to the CBS limit, supplemented with CCSD corrections, as well as SMD and MST continuum solvation calculations to predict the intrinsic acidity of model bilins (lacking the propionate side chains), leading to  $pK_a$  values of  $\sim 6$ –7 in aqueous solution. The results point out that deprotonation of bilins is expected to occur at any of the central B and C pyrrole rings, which are predicted to have similar acidities, given the amide character of the nitrogens in the outer rings (A and D). We then investigated the impact of the propionate side groups and of the specific protein environment on the  $pK_a$ s using PROPKA and APBS calculations, and critically assessed the predictions through comparison

with the absorption spectra of the complexes measured in the  $pH$  range from 4.0 to 9.4. Our simulations predict a strong impact of the local environment in the protonation preferences of bilins, leading to both increases as well as decreases in their acidities according to PROPKA and APBS estimates. These two methods lead to consistent relative  $pK_a$  values, but present a systematic difference of  $\sim 2$   $pK_a$  units. The analysis of the absorption spectral changes observed upon increasing  $pH$  values, however, support the accuracy of APBS estimates and suggest that PROPKA predictions are underestimated by  $\sim 2$   $pK_a$  units, given that appreciable variations in the spectra of the complexes are only observed at  $pH \sim 7$  and found to be consistent with deprotonation of the DBV and MBV pigments lacking a coordinated anionic residue. Thus, spectral changes support our APBS results, which indicate approximate  $pK_a$  values  $\sim 8$ –11 for PCBs, PEBs and DBVs in which the central pyrrole rings interact with an anionic Asp or Glu counterion, as also found for PCBs in the C-PC complex of cyanobacteria, whereas a  $pK_a \sim 7$ –8 is estimated for DBVs and MBVs lacking such counterion in PE545 and PC630/PC645 complexes, respectively.

The present results thus suggest that most bilins in cryptophyte complexes are expected to be fully protonated when the  $pH$  varies on the range  $\sim 5$ –7 in the lumen, depending on the prolongation of light exposure, although some minor fraction of DBV and MBVs could also be deprotonated at  $pH \sim 7$ . Nevertheless, protonation/deprotonation is not expected to change the light harvesting pathways and dynamics in the complexes at physiological conditions. This information will be useful to derive improved models aimed at understanding the interplay between the protein environment and energy transfer dynamics in cryptophytes, including the observation of persistent vibronic coherences in "closed" forms of these antenna proteins.

## Conflicts of interest

There are no conflicts of interest to declare.

## Acknowledgements

We are grateful to the Consorci de Serveis Universitaris de Catalunya for providing access to computational resources (Molecular Recognition project). Financial support from the Agència de Gestió d'Ajuts Universitaris i de Recerca from Generalitat de Catalunya (GENCAT; SGR2014-1189 and SGR2017-1746) and from the Spanish Agencia Estatal de Investigación and the European Regional Development Fund (AEI/FEDER/UE; grants CTQ2012-36195, CTQ2017-89924-P and SAF2017-88107-R) are acknowledged. C.C. is a Serra Húnter Fellow (Generalitat de Catalunya) and M.C. acknowledges a fellowship from the Spanish Ministerio de Economía y Competitividad (MINECO; grant BES-2013-064088). F.J.L. acknowledges the financial support from Icrea Academia. We are thankful to Paul M. G. Curmi for providing the crystal structures of the PC577 and PC630 complexes. Z.S.D.T. and

G.D.S. would like to acknowledge the generous funding by Photosynthetic Antenna Research Center, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0001035.

## References

- 1 Y.-C. Cheng and G. R. Fleming, *Annu. Rev. Phys. Chem.*, 2009, **60**, 241–262.
- 2 T. Mirkovic, E. E. Ostroumov, J. M. Anna, R. van Grondelle, Govindjee and G. D. Scholes, *Chem. Rev.*, 2017, **117**, 249–293.
- 3 E. Collini, C. Curutchet, T. Mirkovic and D. Gregory, in *Energy Transfer Dynamics in Biomaterial Systems*, eds. I. Burghardt, V. May, D. A. Micha and E. R. Bittner, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, vol. 93, pp. 3–34.
- 4 R. J. Ritchie, *Photosynth. Res.*, 2008, **96**, 201–215.
- 5 G. D. Scholes, T. Mirkovic, D. B. Turner, F. Fassioli and A. Buchleitner, *Energy Environ. Sci.*, 2012, **5**, 9374–9393.
- 6 C. Curutchet and B. Mennucci, *Chem. Rev.*, 2017, **117**, 294–343.
- 7 G. D. Scholes, G. R. Fleming, A. Olaya-Castro and R. van Grondelle, *Nat. Chem.*, 2011, **3**, 763–774.
- 8 E. Collini, C. Y. Wong, K. E. Wilk, P. M. G. Curmi, P. Brumer and G. D. Scholes, *Nature*, 2010, **463**, 644–647.
- 9 A. Chenu and G. D. Scholes, *Annu. Rev. Phys. Chem.*, 2015, **66**, 69–96.
- 10 J. C. Dean, T. Mirkovic, Z. S. D. Toa, D. G. Oblinsky and G. D. Scholes, *Chem*, 2016, **1**, 858–872.
- 11 H.-G. Duan, V. I. Prokhorenko, R. J. Cogdell, K. Ashraf, A. L. Stevens, M. Thorwart and R. J. D. Miller, *Proc. Natl. Acad. Sci.*, 2017, **114**, 8493–8498.
- 12 B. S. Rolczynski, H. Zheng, V. P. Singh, P. Navotnaya, A. R. Ginzburg, J. R. Caram, K. Ashraf, A. T. Gardiner, S.-H. Yeh, S. Kais, R. J. Cogdell and G. S. Engel, *Chem*, 2018, **4**, 138–149.
- 13 S. M. Blau, D. I. G. Bennett, C. Kreisbeck, G. D. Scholes and A. Aspuru-Guzik, *Proc. Natl. Acad. Sci.*, 2018, **115**, E3342–E3350.
- 14 S. J. Harrop, K. E. Wilk, R. Dinshaw, E. Collini, T. Mirkovic, C. Y. Teng, D. G. Oblinsky, B. R. Green, K. Hoef-Emden, R. G. Hiller, G. D. Scholes and P. M. G. Curmi, *Proc. Natl. Acad. Sci.*, 2014, **111**, E2666–E2675.
- 15 M. K. Lee, K. B. Bravaya and D. F. Coker, *J. Am. Chem. Soc.*, 2017, **139**, 7803–7814.
- 16 L. Spear-bernstein and K. R. Miller, *J. Phycol.*, 1989, **3**, 412–419.
- 17 B. Demming-Adams and W. W. Adams, *New Phytol.*, 2006, **172**, 11–21.
- 18 K. E. Wilk, S. J. Harrop, L. Jankova, D. Edler, G. Keenan, F. Sharples, R. G. Hiller and P. M. G. Curmi, *Proc. Natl. Acad. Sci.*, 1999, **96**, 8901–8906.
- 19 A. B. Doust, C. N. J. Marai, S. J. Harrop, K. E. Wilk, P. M. G. Curmi and G. D. Scholes, *J. Mol. Biol.*, 2004, **344**, 135–153.
- 20 H. Kikuchi, T. Sugimoto and M. Mimuro, *Chem. Phys. Lett.*, 1997, **274**, 460–465.
- 21 C. Kneip, P. Hildebrandt, F. Mark and K. Schaffner, *Chem. Phys. Lett.*, 1999, **311**, 479–484.
- 22 M. A. Mrogiński, F. Mark, W. Thiel and P. Hildebrandt, *Biophys. J.*, 2007, **93**, 1885–1894.
- 23 F. Velazquez Escobar, C. Lang, A. Takiden, C. Schneider, J. Balke, J. Hughes, U. Alexiev, P. Hildebrandt and M. A. Mrogiński, *J. Phys. Chem. B*, 2017, **121**, 47–57.
- 24 A. B. Doust, K. E. Wilk, P. M. G. Curmi and G. D. Scholes, *Photochem. Photobiol. A Chem.*, 2006, **184**, 1–17.
- 25 T. Mirkovic, A. B. Doust, J. Kim, K. E. Wilk, C. Curutchet, B. Mennucci, R. Cammi, P. M. G. Curmi and G. D. Scholes, *Photochem. Photobiol. Sci.*, 2007, **6**, 964–75.
- 26 C. Curutchet, J. Kongsted, A. Muñoz-Losa, H. Hossein-Nejad, G. D. Scholes and B. Mennucci, *J. Am. Chem. Soc.*, 2011, **133**, 3078–3084.
- 27 H. Hossein-Nejad, C. Curutchet, A. Kubica and G. D. Scholes, *J. Phys. Chem. B*, 2011, **115**, 5243–5253.
- 28 C. Curutchet, V. I. Novoderezhkin, J. Kongsted, A. Muñoz-Losa, R. van Grondelle, G. D. Scholes and B. Mennucci, *J. Phys. Chem. B*, 2013, **117**, 4263–4273.
- 29 L. Viani, C. Curutchet and B. Mennucci, *J. Phys. Chem. Lett.*, 2013, **4**, 372–377.
- 30 L. Viani, M. Corbella, C. Curutchet, E. J. O'Reilly, A. Olaya-Castro and B. Mennucci, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16302–16311.
- 31 M. Aghtar, J. Strümpfer, C. Olbrich, K. Schulten and U. Kleinekathöfer, *J. Phys. Chem. Lett.*, 2014, **5**, 3131–3137.
- 32 M. Aghtar and U. Kleinekathöfer, *J. Lumin.*, 2016, **169**, 406–409.
- 33 S. Chandrasekaran, K. R. Pothula and U. Kleinekathöfer, *J. Phys. Chem. B*, 2017, **121**, 3228–3236.
- 34 K. S. Alongi and G. C. Shields, in *Annual Reports in Computational Chemistry*, Elsevier, 2010, pp. 113–138.
- 35 R. Casasnovas, J. Ortega-Castro, J. Frau, J. Donoso and F. Muñoz, *Int. J. Quantum Chem.*, 2014, **114**, 1350–1363.
- 36 M. Feliks and M. J. Field, *J. Chem. Inf. Model.*, 2015, **55**, 2288–2296.
- 37 M. N. Davies, C. P. Toseland, D. S. Moss and D. R. Flower, *BMC Biochem.*, 2006, **7**, 18–29.
- 38 C. L. Stanton and K. N. Houk, *J. Chem. Theory Comput.*, 2008, **4**, 951–966.
- 39 E. Alexov, E. L. Mehler, N. Baker, A. M. Baptista, Y. Huang, F. Milletti, J. Erik Nielsen, D. Farrell, T. Carstensen, M. H. M. Olsson, J. K. Shen, J. Warwicker, S. Williams and J. M. Word, *Proteins Struct. Funct. Bioinform.*, 2011, **79**, 3260–3275.
- 40 M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525–537.
- 41 C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 2284–2295.
- 42 N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *Proc. Natl. Acad. Sci.*, 2001, **98**, 10037–10041.
- 43 I. M. Alecu, J. Zheng, Y. Zhao and D. G. Truhlar, *J. Chem. Theory Comput.*, 2010, **6**, 2872–2887.
- 44 S. Grimme, *J. Chem. Phys.*, 2003, **118**, 9095–9102.
- 45 A. Halkier, T. Helgaker, P. Jørgensen and W. Klopper, *Chem. Phys. Lett.*, 1998, **286**, 243–252.

- 46 I. Soteras, C. Curutchet, A. Bidon-Chanal, M. Orozco and F. J. Luque, *J. Mol. Struct. THEOCHEM*, 2005, **727**, 29–40.
- 47 C. Curutchet, A. Bidon-Chanal, I. Soteras, M. Orozco and F. J. Luque, *J. Phys. Chem. B*, 2005, **109**, 3565–3574.
- 48 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
- 49 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision D.1*, Gaussian Inc., Wallingford CT, 2009.
- 50 D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, I. T.E. Cheatham, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P.A. Kollman, *AMBER 14*, University of California, San Francisco, 2014.
- 51 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
- 52 J. M. Wang, R. M. Wolf, J. W. Caldwell, P. a Kollman and D. a Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
- 53 C. C. Jumper, P. C. Arpin, D. B. Turner, S. D. McClure, S. Rafiq, J. C. Dean, J. A. Cina, P. A. Kovac, T. Mirkovic and G. D. Scholes, *J. Phys. Chem. Lett.*, 2016, **7**, 4722–4731.
- 54 P. M. Dewick, *Essentials of organic chemistry : for students of pharmacy, medicinal chemistry and biological chemistry*, J. Wiley, 2006.
- 55 A. D. Bochevarov, M. A. Watson, J. R. Greenwood and D. M. Philipp, *J. Chem. Theory Comput.*, 2016, **12**, 6001–6019.
- 56 A. J. Laos, J. C. Dean, Z. S. D. Toa, K. E. Wilk, G. D. Scholes, P. M. G. Curmi and P. Thordarson, *Angew. Chemie Int. Ed.*, 2017, **56**, 8384–8388.
- 57 D. B. Turner, R. Dinshaw, K.-K. Lee, M. S. Belsley, K. E. Wilk, P. M. G. Curmi and G. D. Scholes, *Phys. Chem. Chem. Phys.*, 2012, **14**, 4857–4874.
- 58 V. I. Novoderezhkin, A. B. Doust, C. Curutchet, G. D. Scholes and R. van Grondelle, *Biophys. J.*, 2010, **99**, 344–352.