Journal Name

Elucidating the role of structural fluctuations, intermolecular and vibronic interactions in the spectroscopic response of a bacteriophytochrome.[†]

Veronica Macaluso,[*a] Lorenzo Cupellini,[a] Giacomo Salvadori,[a] Filippo Lipparini,[a] and Benedetta Mennucci,[*a]

We present the first comprehensive multiscale computational investigation of Resonance Raman (RR), absorption and Circular Dichroism (CD) spectra of the resting state of the dimeric phytochrome of *Deinococcus radiodurans*. The spectra are simulated in all their components, namely the energy position and the lineshapes of both the far-red and the blue bands. To achieve such a goal, we have combined a $4.5 \mu s$ MD simulation of the solvated dimeric phytochrome with a hybrid quantum mechanics/molecular mechanics (QM/MM) model, which accounts for both electrostatic and mutual polarization effects between the QM and the MM subsystems. From simulations, we find a transient H-bond network within the binding pocket of the chromophore that, unexpectedly, do not significantly affect the absorption and CD spectra. In parallel, we characterize the vibrations which are more strongly coupled to the excitation confirming the important role of the hydrogen-out-of-plane mode of the vinyl C-H in between C and D rings together with the expected C=C stretching of the double bond involved in the photoisomerization.

1 Introduction

In the class of photosensory proteins, light is used to activate a downstream signal transduction event. While using different chemistries and signaling modes, common features are shared among the different systems to achieve this goal. In particular, they can all be seen as molecular machines composed of sensory domains which contain the absorbing chromophore(s) and effector modules. Conformational changes induced by photon absorption are transmitted from the chromophore binding pocket to the exterior of the protein matrix.[1–4] In phytochromes, a major family of red-light-sensing proteins that control diverse cellular functions in plants, bacteria and fungi, the protein binds a tetrapyrrole chromophore (a bilin) which absorbs light. When this happens, the bilin undergoes a reversible conformational change, which also involves the surrounding protein, thus acting as a switch between resting (or inactive) and active states.[5–8] As a matter of fact, the family of phytochromes is large and diverse in terms of domain organization, photochemistry of different bilin chromophores, and the type and position of the protein residues near the chromophore; as a result, their response to light can signifi-

cantly vary. In most cases, however, phytochromes share the characteristic that red light irradiation converts the red-absorbing (Pr) state into the metastable, far-red light absorbing (Pfr) state. This photoconversion is reversible and Pfr returns to Pr either upon absorption of a far-red photon or via a thermal process.[6,9]

Among the large family of phytochromes, those from bacteria represent a very interesting subcase for biochemical, spectroscopic, and crystallographic analyses.[10–12] Here, in particular, we focus on the phytochrome from the extremophile bacterium *Deinococcus radiodurans* (DrBph). This is a homodimeric protein where the photosensory module contains a Per/Arnt/Sim (PAS) domain, a cyclic guanosine monophosphate phosphodiesterase/adenylyl cyclase/FhlA (GAF) domain, and a phytochrome-specific (PHY) domain (see Figure 1 a). The bilin chromophore (a biliverdin IXα, from now on BV) is nested in a binding pocket formed by the PAS/GAF domains (Figure 1 b and c).

In DrBph, the PHY domain is connected to the GAF domain through a long helical spine and through the so-called PHY "arm" or "tongue". The chromophore is covalently linked to a cysteine residue of the N-terminal extension of the PAS domain but it is also non-covalently linked (through a network of hydrogen bonds involving the propionate groups) to the GAF domain and, more indirectly through water molecules, to the tongue of the PHY domain (see figure 1d). It is now generally accepted that the light absorption by the BV induces a Z/E isomerization of the
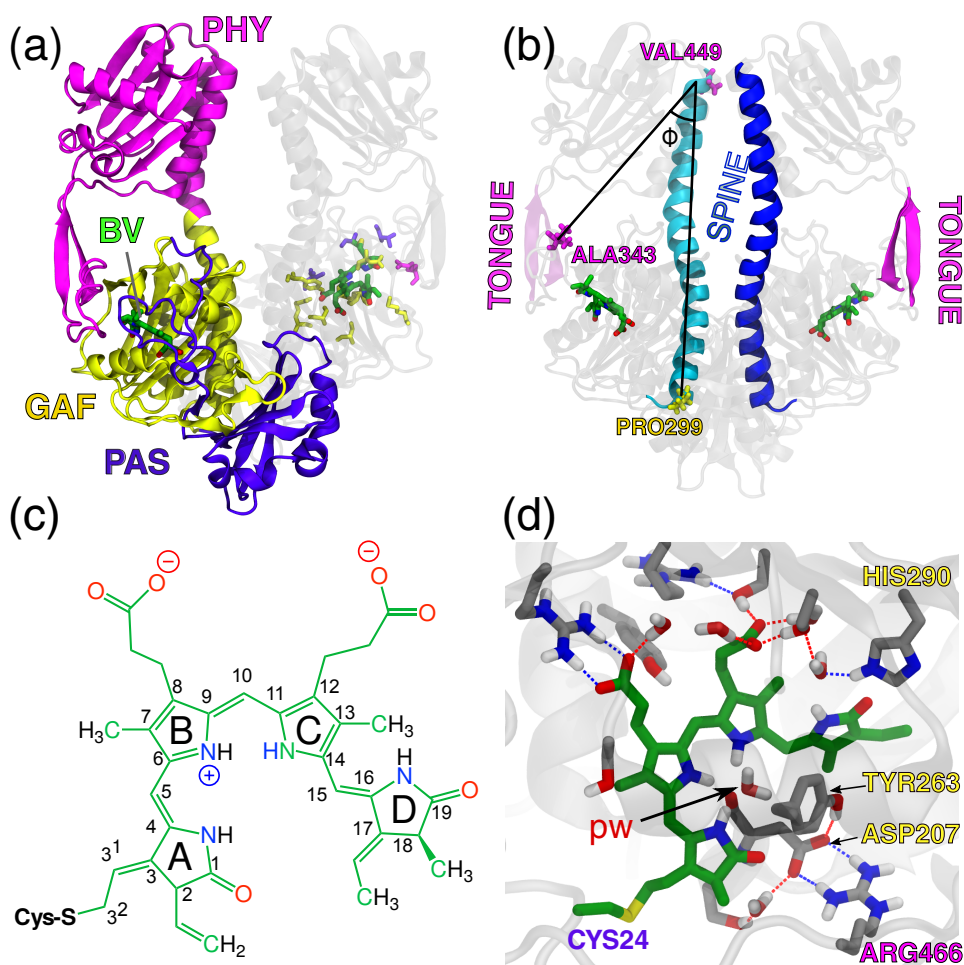
Fig. 1 (a) DrBph crystal structure in the Pr form highlighting the PAS (blue), GAF (yellow) and PHY (purple) domains. The BV chromophore is represented in green. (b) Representation of the "tongue" and spine elements of the phytochrome structure, and of the monomer opening angle PRO299-VAL449-ALA343 (ϕ). (c) Chemical structure of the BV chromophore and carbon atom numbering. (d) Zoom-in of the chromophore binding pocket with indication of the most important amino acids and crystal water molecules surrounding the BV. Dotted lines represent H-bond interactions. In particular, with *pw* we indicate the pyrrole water. Residue labels in panels (b,d) are colored as the domains in panel (a).

C15=C16 double bond between the C- and D-pyrrole rings (see Fig. 1c) [13–15] but the details of the multiscale processes which finally lead to the change of the folding of the PHY tongue from β sheet in Pr to α helix in Pfr and the large scale rearrangement of the whole phytochrome are still to be clarified. In particular, what needs to be understood is how the localized structural change in BV (e.g. the twisting around C15=C16 double bond) propagates through the surrounding residues up to the PHY tongue.

In this context, computational studies can be the key to connect the experiments with detailed atomistic descriptions. [16] Due to the enormous complexity of describing the many time and space scales involved, simulations of the whole light-induced process have not appeared so far whereas few attempts have been presented for the short-time scale (the photoisomerization) and the structural changes that photoisomerization could induce on the residues more strongly interacting with the chromophore [17,18]. Instead, most of the computational work have been focused on the simulation of absoprtion spectra of Pr and Pfr states. [19–25] These studies, however, have either used simplified models of

the system (in particular of the protein matrix and the environment) or have neglected the vibronic interactions determining the spectral bandshape. Here, for the first time, absorption (Abs), circular dichroism (CD) and Resonance Raman (RR) spectra are simulated together for the entire dimeric structure of the bacteriophytochrome in solution. Moreover, these simulations are not limited to the prediction of the transition energies and transition dipoles (or rotational strengths) but they include the reproduction of the band shapes by calculating both vibronic effects and static disorder. To achieve such a goal, we combine *microsecond* molecular dynamics (MD) simulations of the entire dimeric phytochrome structure in aqueous solution with a hybrid quantum mechanics/molecular mechanics (QM/MM) model, which accounts for both electrostatic and mutual polarization effects between the QM and the MM subsystems. The mutual polarization is obtained by using an induced dipole formulation through the addition of atomic polarizabilities to the MM atoms (from now on MMPol) [26–28].

From our MD simulation, we find a transient or absent H-bond

network for the carbonyl and amino group of the BV D-ring, respectively. On the other hand, we find a connection between the dynamics of the tyrosine in proximity of the D ring (TYR263) and the water molecule interacting with the pyrrole hydrogens (from now indicated as pyrrole water, *pw* in Figure 1d). However, these transient behavior within the binding pocket seems not to significantly affect the absorption and CD spectra as calculations performed using either the crystal structure or different snapshots extracted from the MD simulation give almost identical results. On the other hand, the very good agreement found for all the three simulated spectra (RR, Abs and CD) with the measured ones (both in the position and the shape of all the peaks), indicate that a proper selection of the QM/MM approach can indeed give a complete and detailed description of the structure and the spectroscopic properties of the phytochrome in its resting state.

2 Methods and computational details

2.1 Crystal structure preparation

Initial atomic coordinates of the phytochrome were extracted from the 4Q0J[14] entry of the protein data bank (PDB), which was chosen for the relatively high resolution of the PAS-GAF-PHY monomer and the presence of crystallized waters. The crystal was compared with biological sequences using the ALIGN program, while MODELLER[29] was used to add the missing residues in the crystal structure, namely, ALA107, ALA108, TRP132, ASP133, SER134, THR135, GLY431, GLU432. The H++ program[30] was used to determine the protonation at pH 6.5. The AMBER ff14SB force-field[31] was used for the protein, while the GAFF force-field[32] was used for biliverdin IXα (BV) and for the bonding interactions between BV and the covalently bonded cysteine residue CYS24. In particular, the GAFF atom types were firstly assigned using antechamber[33,34] and then suitably modified to reflect the conjugation pattern. Partial charges of the BV were computed using the RESP protocol[35] on top of a HF/6-31G(d) electrostatic potential calculation. The entire dimer was then built by using the transformation matrices contained in the PDB. All hydrogens and added residues were then minimized with AMBER18.[36,37]

2.2 Molecular dynamics setup

The crystal structure, refined as described in the previous section, was immersed in a water box, ensuring a minimum 40 Å separation between periodic images of the protein, and neutralized with sodium ions. The solvated system was subjected to 800 steps of energy minimization using steepest descent and other 3200 steps using conjugate gradient. The system was then gradually heated to 300 K in a NVT ensemble in 800 ps, restraining the movement of the chromophores, the protein backbone, and the crystal water oxygens with a 4 kcal mol^{-1}Å^{-2} harmonic potential. Three different NPT equilibration runs followed to assure a slow release of the restraints: (i) 2000 ps with restraints on chromophores and all the protein backbone; (ii) 1600 ps in which only the added residues and some of the adjacent residues were released from the previous restraint; (iii) 2000 ps in which all restraints were released except the phytochrome tongues. The production run was run without any restraint for 4.5 μs in the NPT ensemble.

All simulations were performed applying the particle mesh Ewald (PME) truncation method (with a short-range cut-off of 10 Å), an integration step of 2 fs, the SHAKE algorithm, a Langevin thermostat with a friction coefficient of 1 ps^{-1}, the Monte Carlo thermostat for NPT simulations. All molecular dynamics simulations were run with AMBER18.[36,37]

2.3 Geometry optimizations

To avoid possible inaccuracies due to the crystal resolution (in the static model) or the MM force field used in the MD (in the dynamic model), all excited-state calculations were performed on partially optimized structures. All such geometry optimizations were performed using a ONIOM(QM:MM) scheme,[38] as implemented in the Gaussian 16 suite of programs.[39] The flexible QM subsystem was represented by the BV chromophore and described at the B3LYP-D3/6-31G(d) level of theory,[40] whereas the protein was kept frozen and described by the ff14SB force field[31]. The S and C$_A$ atoms of CYS24 were kept flexible and included in the QM part, by cutting the C$_A$–C$_B$ covalent bond.

The partial optimizations were repeated for the crystal structure and for all the configurations extracted from the MD that will be successively used for the simulation of electronic spectra. In the case of the configurations from the MD all the water molecules within 18 Å of the BV chromophore and all residues of the respective monomer were included in the frozen MM subsystem.

For calculations in chloroform, the solvent was included as a continuum dielectric with the IEFPCM model[41]. The propionate groups of the chromophore were cut to prevent charge transfer contamination to the excited states, which is more pronounced without explicit stabilizing interactions with the solvent. It should be noted that experimental absorption in chloroform was obtained on the BV with esterified propionates.[42]

2.4 Simulation of Resonance Raman and electronic spectra

The spectroscopic properties were computed applying the QM/MMPol scheme where the QM subsystem (BV) was treated at the (TD)DFT/6-31+G(d) level of theory using several DFT functionals (*vide infra*). The MM part included the protein matrix and, in the case of the dynamic model, the solvent within 40 Å of the BV. The QM part of the chromophore was cut at the C3^1-C3^2 covalent bond (See Figure 1c). For the static model, the protein matrix consisted only of one monomer. In the dynamic model, 40 snapshots extracted from the MD from 2μs to 4μs of the production run with an interval of 50 ns were used. For the same snapshots, calculations were performed on each of the two chromophores.

The homogeneous lineshape of the biliverdin was described within the second-order cumulant expansion formalism,[43] in which all the information on the vibronic coupling is described in terms of a *spectral density* (SD) function $J_{eg}(\omega)$. The spectral density describes the linear coupling of the excitation $g \to e$ to an infinite set of harmonic oscillators, with frequencies ω_k,

$$J_{eg}(\omega) = \pi \sum_k S_k^{(eg)} \omega_k^2 \big(\delta(\omega - \omega_k) - \delta(\omega + \omega_k)\big) \quad (1)$$

where $S_k^{(eg)}$ is the dimensionless displacement, or Huang-Rhys factor, for excited state e along the normal mode k. These quantities were calculated by projecting the excited-state gradients onto the normal modes. The details of the homogeneous lineshape calculation and inhomogeneous broadening estimation are given in the ESI[†].

The spectral density $J(\omega)$ can also be directly compared to the experimental Resonance Raman spectrum, which was obtained in pre-resonance conditions.[44] To the first order in the vibronic coupling, the Resonance Raman intensity for a mode k is proportional to the Huang-Rhys factors $S_k^{(eg)}$, with a prefactor that depends on the excitation frequency.[45,46] However, for pre-resonance Raman spectra and in the case of broad lineshapes the Resonance Raman cross-section reduces to $\propto \omega_k^2 S_k$.[46]

To obtain the spectral density, normal mode calculations for the ground state embedded BV were carried out excluding the propionate groups from the QM part and using either the CAM-B3LYP[47] or ωB97XD[48] functional, in combination with the 6-31G(d) basis set, after reoptimizing the structure at the same level of theory. The excited-state gradients were calculated, at the ground-state optimized geometry, using the same QM and MM subsystems in combination with the same functionals, and employing the 6-31+G(d) basis set to better describe the excited state.

3 Results and discussion

To reveal the role of the structure (and its thermal fluctuations), the coupling of vibrations with the excitation process, and the intermolecular interactions in determining the spectroscopic response of the phytochrome, we have compared two different models: a static one that is based on the crystal structure (see section 3.2) and a dynamic one which uses configurations of the solvated phytochrome coming from a 4.5 μs MD simulation (see section 3.3). Before considering the two models, however, we performed some test calculations to select the best DFT functional: these are presented and discussed in the following section.

3.1 Selection of the DFT functional

As a preliminary test, we compared the calculated and experimental[42] absorption spectra of BV in chloroform using four different DFT functionals, previously used for BV[19,22,23,25], i.e. BLYP[49], B3LYP,[50] its long-range corrected extension (CAM-B3LYP[47]) and another long-range corrected functional (ωB97XD[48]) which recovers the correct 100% exact exchange at infinite separation.

All the investigated functionals describe the far-red region of the absorption spectrum (the so-called Q band) with a single and bright electronic transition (see Table S1 and Figure S1a of the ESI[†]). However, the ωB97XD and CAM-B3LYP functionals are the ones that better describe both the higher-energy transitions (namely those determining the Soret band) and the relative position of the two bands.

An identical analysis was repeated for BV in protein using the crystal structure and the obtained results (reported in table S1 and Figure S1b of the ESI[†]) are in line with those obtained in chloroform: the BLYP and B3LYP functionals perform signif-

icantly worse for the Q-Soret band gap, confirming the better performance of the long-range corrected functionals for this chromophore. We note that these findings agree with previous computational studies;[19,22–24] however, here, we have used a different approach for the analysis of the functional. In fact, even though CAM-B3LYP and ωB97XD are not the functionals that best reproduce the Q band position in absolute terms, they can accurately reproduce Q-Soret band gap both in chloroform and in the protein environment. In addition, Figure S1 shows that B3LYP and especially BLYP predict a number of spurious states between the Q and Soret bands, both in chloroform and in protein. We finally note that CAM-B3LYP and ωB97XD are also able to correctly reproduce the relative band intensities both in chloroform solution and in the protein.

To gain more insight into the description of excited states given by the two best performing functionals, namely CAM-B3LYP and ωB97XD, a Natural Transition Orbital (NTO) analysis[51] was performed for the three lowest energy excitations (see Figure S2). From the resulting NTOs, it is evident that the two functionals give a very similar description of the first two excitations, with the second one only showing some minor differences. However, they differ in the description of the third excitation. In particular, ωB97XD describes a transition strongly localized on the C and D rings, while CAM-B3LYP shows a significant contribution of all the rings.

On the basis of these results, the following analyses on RR, absorption and CD spectra will be presented only for CAM-B3LYP whereas the analogue results obtained with ωB97XD will be reported in the ESI[†] and discussed on the basis of the experiments.

3.2 The static model

As detailed in the Methods section, a refined crystal structure was used as a starting point for our spectroscopic investigation of the phytochrome.

As a first analysis, we computed the spectral density and compared it with the measured Resonance Raman spectrum:[52] this comparison is reported in Figure 2 for CAM-B3LYP (the ωB97XD results are reported in the figure S3 of the ESI[†]). In the same figure we also show the normal modes that correspond to the RR main signals (a more complete description of the different RR peaks is reported in Fig. S4 and Table S2 of the ESI[†]).

As it can be seen from Figure 2(b), the peak calculated at 804 cm^{-1} (mode 79) is here explained as the hydrogen-out-of-plane (HOOP) mode of the vinyl C-H in between C and D rings . Because the intensity and frequency of this mode are correlated with out-of-plane distortions caused by steric interactions between the C and D rings, its changes upon illumination have been used as a probe of the Z to E photoisomerization dynamics.[54] At $\sim$1309 (mode 127) and at $\sim$1567 cm^{-1} (mode 159) we find the rocking and the in-plane bending of N-H of the B and C rings, respectively (Figure 2(c) and (d)). Finally, the most intense peaks at 1615 and $\sim$1656 cm^{-1} correspond to the C15=C16 and C9=C10 stretching modes, respectively (modes 162 and 163 in Figure 2(e)). These results are consistent with previously reported signal assignments[44,52,55].
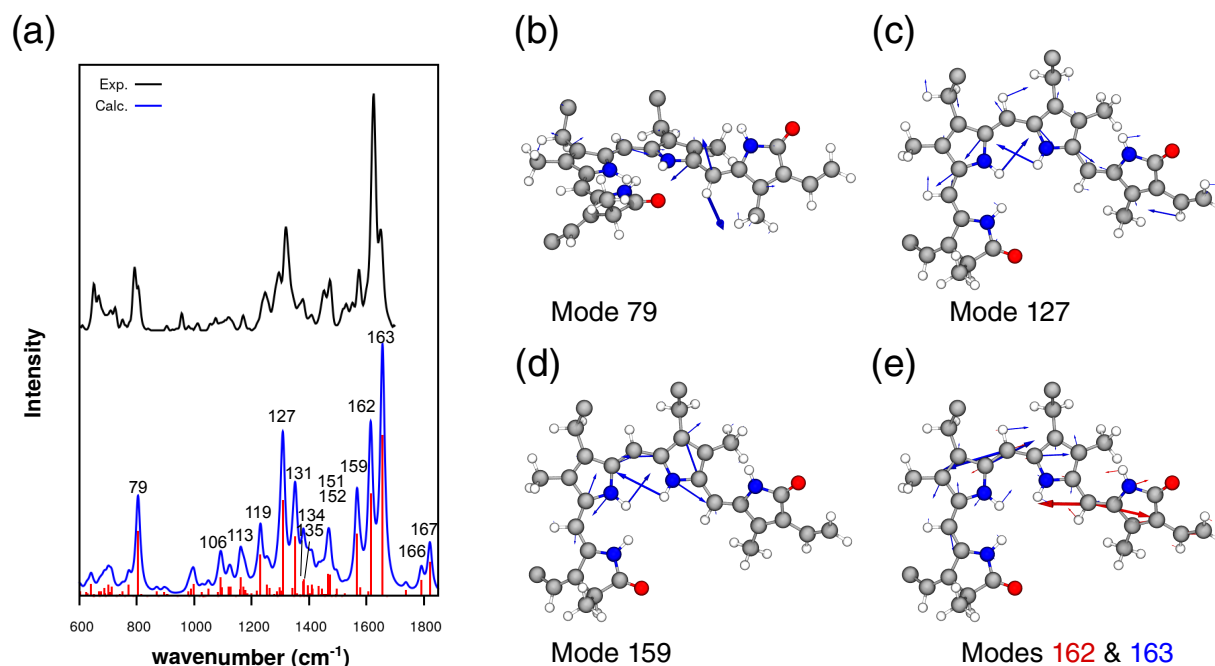
Fig. 2 (a) Comparison of experimental Resonance Raman spectrum [52] and CAM-B3LYP calculated spectral density for the first excited state on the crystal structure. The calculated SD frequencies were scaled by 0.95 for this comparison [53]. The modes with the strongest coupling to the excitation are labeled. (b-e) Arrow representation of the five modes with strongest coupling. (b) HOOP mode of the vinyl C_{15}-H in between C and D rings; (c) combined mode comprising in-plane N-H bending and C-N stretching of the B and C rings . (d) in-plane N-H bending of the B and C rings . (e) In blue the $C_9=C_{10}$ and in red the $C_{15}=C_{16}$ stretching modes.

The spectral densities were finally used to obtain the homogeneous lineshapes for the simulation of absorption and CD spectra (see Methods section and ESI† for details).

The CAM-B3LYP results, reported in Figure 3(a,c), show a very good agreement with experiments [15] for the Q band, whose lineshape is perfectly reproduced by the calculations. Some discrepancies are instead found in the Soret region, for which the shape, and, in the case of CD, the position are not well reproduced. Both the second and third excited states contribute to this band in the calculated absorption spectrum, but their contributions cannot be separated from the experimental absorption alone. However, in the experiment, the CD maximum for this band is significantly blue shifted compared to the corresponding absorption maximum, indicating that at least two transitions contribute, with different intensities, to the absorption and CD intensities in this region.

Inspecting the oscillator and rotational strengths, and comparing the Soret absorption and CD bands with experiment [15], we can conclude that CAM-B3LYP gives a wrong ordering of the higher-energy excitations: in fact, we obtain a much better agreement with the experimental CD (and the relative absorption-CD shift) by switching the second and third electronic transitions.

We note that for ωB97XD, (see Figure S5 in the ESI†), the shape of the Q band is very similar to what obtained for CAM-B3LYP, with only a slightly stronger vibronic coupling. For the other two excitations, the differences between the functionals are somewhat larger but still moderate, even for the third excited state, whose electronic structure is described differently by the two functionals. However, for ωB97XD, it was not possible to improve the description of the Soret band even by switching the excitation energies as done with the CAM-B3LYP transitions.

From this analysis we can conclude that a good agreement between the experimental and calculated electronic spectra (both absorption and CD) can be obtained by relaxing the internal geometry of the chromophore within a protein frozen in its crystal structure. Moreover, the very good agreement found for the spectral line shape of the Q transition strongly suggests that the shoulder at higher energies has vibronic origin, and the presence of different protonation states, recently invoked in the literature [23], is not needed to accurately reproduce the spectrum. The combined comparison of absorption and CD results indicates that (TD)DFT does not accurately reproduce the energy ordering of the excitations in the Soret region. We note that this discrepancy cannot be pinpointed only by looking at the absorption spectra, thus highlighting the importance of computing simultaneously absorption and CD spectra, to better clarify the nature and composition of the bands.

3.3 The dynamic model

To investigate the role of structural fluctuations and of the solvent, the analyses previously performed on a single (crystal) structure were repeated for different configurations of the solvated phytochrome sampled along the MD simulation.

The motion of the phytochrome as a whole was first assessed by a root mean square displacement (RMSD) analysis on the dimer,
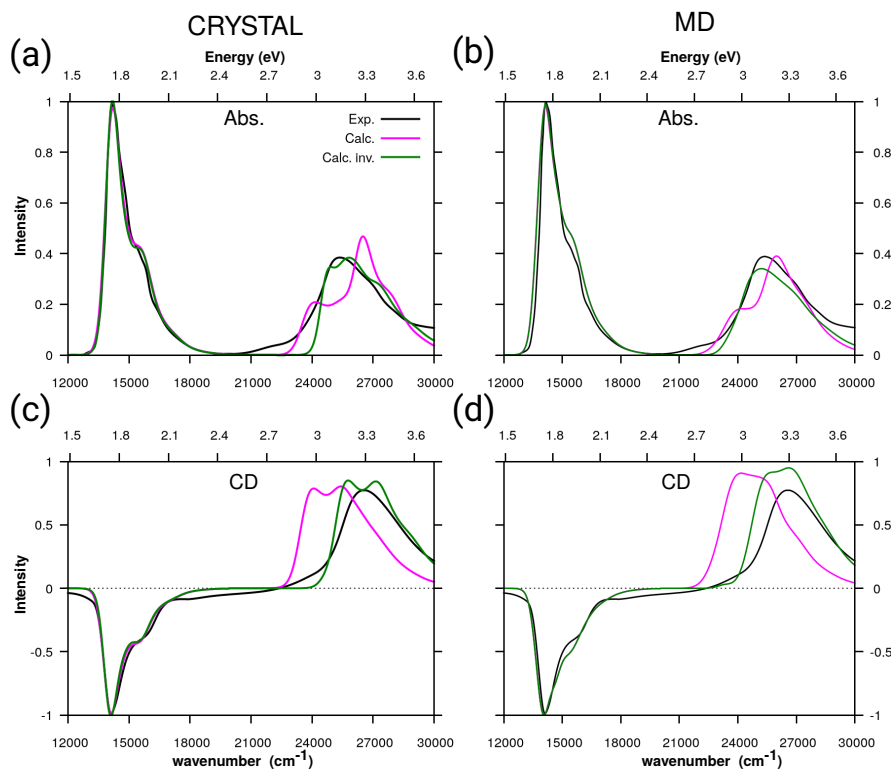
Fig. 3 (TD)DFT absorption and CD spectra computed with CAM-B3LYP (a,c) on the crystal structure and (b,d) along the MD trajectory. In the latter case the spectra have been obtained as an average over 80 MD configurations (See the ESI† for details). Green spectra have been obtained by switching the energies of the second and third excitations. All the calculated spectra have been homogeneously shifted by −0.16 eV for the crystal and −0.19 eV for the MD, to match the experimental[15] lowest energy Q band. All spectra are normalized to the Q-band maximum (minimum in the case of CD).

the two monomers and the two embedded BV chromophores. This analysis (reported in Figure S5 of the ESI†) shows a significant displacement from the crystal structure for both monomers, which find an equilibrium at a value of RMSD with respect to the crystal of about ∼2.5 Å in the first μs. The RMSD of the protein backbone shows only some small oscillations (differing by at most for ∼1.5 Å) in both monomers and also the relative motion of the two monomers is limited. A similar behavior is also found in the RMSD of the two domains (PAS and GAF) most strongly interacting with the chromophore. The results show that after ∼ 2 μs both PAS and GAF domains are equilibrated in both monomers.

Let us now analyze the binding pocket of the BV chromophore starting with the hydrogen-bond network of the D-ring carbonyl group (C19=O), which might play a central role in the Z to E isomerization (Table S3 in the ESI†). This group is hydrogen bonded to a neighboring water molecule for ∼20% of the simulation time, and only for a small fraction with HIS290. These results indicate a dynamic H-bond involvement of the carbonyl oxygen of the D ring. Notably, HIS290 is not hydrogen bonded to the D-ring carbonyl in the crystal structure (Figure 1d).

In contrast with other simulations[23], we do not find any H-bond between another histidine (HIS260) and BV for monomer 1 along the trajectory, while we observe a transient H-bond with a 4.4% probability for monomer 2 (mainly with the pyrrole hydrogen of ring A).

As a matter of fact, we found significant differences between

the two monomers in the dynamics of the pyrrole water (*pw* in Figure 1). In fact, the pyrrole water of monomer 2 exchanges with bulk water molecules, or, more rarely, it even leaves its binding site (see Figure 4 b). On the contrary, the pyrrole water of monomer 1 is present along the whole trajectory and never exchanges with any other molecule. We recall that recently it was suggested that one of the key-steps initiating the Pr-Pfr phytochrome conversion after the chromophore photo-isomerization is the pyrrole-water detachment from its binding site.[56] Analyzing the crystal structure, we observed that the pyrrole water is kept in place by a dynamic "cage" formed by a hydrogen bonding network among ARG466, ASP207 TYR263 and HIS260 (see Figure 1d). Looking at the MD simulation, we can see that whenever this cage is closed, the pyrrole water cannot escape but when another water molecule enters in the binding pocket, a temporary cleavage of the ASP207-O⋯HO-TYR263 interaction is observed (figure 4 (b)), thus opening the cage with the final exit of the pyrrole-water. On the other hand, the cleavage of the ASP207-O⋯HO-TYR263 interaction does not affect the ARG466-ASP207 H-bond interaction.

To have a more detailed picture, we analyzed the interactions involved in the pyrrole water cage along the trajectory in the 2 − 4.5 μs range . First of all, we note the formation of a H-bond between HIS260 and the pyrrole water with a probability of the 89% and 65% in the binding pocket of monomer 1 and monomer 2, respectively. H-bonds interactions are also present between
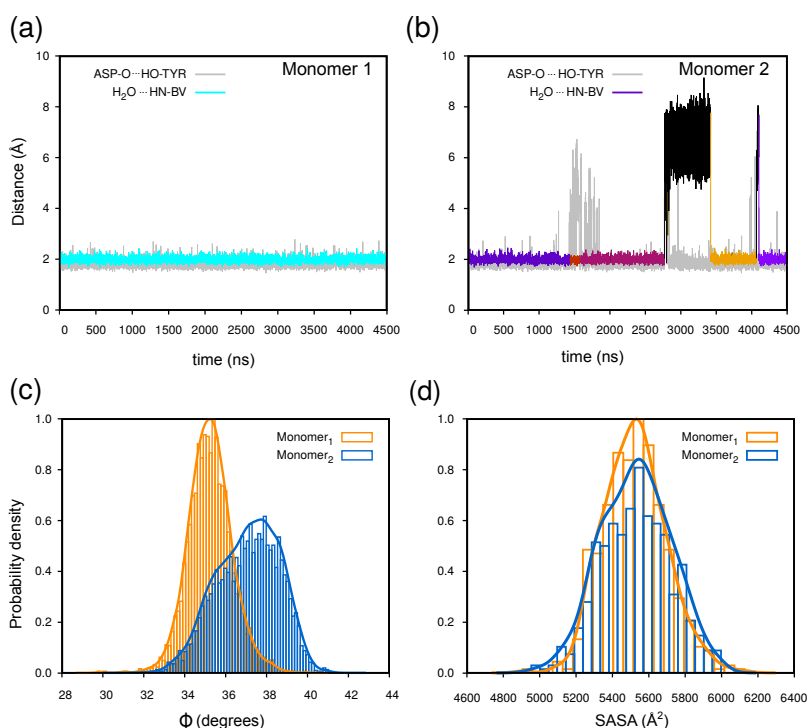
Fig. 4 (a, b) Minimum ASP-O···HO-TYR distance and minimum distance between the pyrrole water oxygen and any aminic hydrogen (H_2O·HN-BV) of the pyrrole ring. For the H_2O·HN-BV distance, different colours indicate distinct water molecules; (c) Distribution of the monomer opening angle (PRO299-VAL449-ALA343) for the two monomers; (d) Solvent accessible surface area (SASA) distribution. The distributions are calculated on the whole production run.

ASP207 and ARG466 with a probability higher than 95% for both monomers, and TYR263 with a probability of 93% and 89% in monomers 1 and 2 respectively.

Secondary structure analyses and dynamic cross correlation (DCC) analyses, which were carried out on the two monomers to understand the origin of the two binding pockets difference (for example regarding the different pyrrole water stability), did not show any important dissimilarity (Figure S8 of the ESI[†]). In order to understand the structural differences between the two monomers, we analyzed the PAS-PHY-Tongue angle ϕ, shown in Figure 1(b), which describes the opening of the monomer. However, looking at the monomers bending distribution along the simulation (see Figure 4 (c)), a larger angle distribution is found for monomer 2. This may indirectly lead to the observed differences, for example by causing a different exposure of the binding pocket to the solvent. Indeed, we observe a broader distribution of the solvent accessible surface area (SASA) in monomer 2, as illustrated in Figure 4 (d).

Once identified and characterized the main fluctuations in the BV binding pocket, we can investigate if and how these fluctuations really affect the spectra. To do so we computed the absorption and CD spectra on 80 snapshots extracted from the MD. On each snapshot the same protocol used for the static model was applied and the final spectra were obtained by combining the homogeneously broadened spectra obtained for each snapshot with the inhomogeneous broadening due to the distribution of excitation energies and transition dipoles, as described in the Methods

Section and the ESI[†]. The resulting spectra are reported in Figure 3(b,d).

A very good agreement for the band shape of the Q transition and the Q-Soret band gap in both absorption and CD spectra is found, while the agreement for the Soret band shape and its absorption-CD shift is not well reproduced. In order to check if some differences could be induced by the different mobility of the pyrrole water observed in the binding pocket of the two monomers, we have also recalculated the absorption and CD spectra separately per each monomer. The differences found (see Figure S5 in the ESI[†]) are very small thus indicating a minor effect of this possible release of the water. Following the same analysis done in the previous static model, we have recalculated the spectra by introducing a switch in the order of the second an third excitation, and the results confirm the same conclusions, namely that this switch largely improves both the band shape of the Soret band and the Q-Soret band gap.

From the comparison of the spectra obtained using the crystal structure or the different configurations extracted from the MD, it is evident that the observed fluctuations in the network of intermolecular interactions keeping the BV within its binding pocket do not significantly alter the spectroscopic response. In particular the Q transition which is the one initiating he photoinduced isomerization is extremely robust with respect to small changes in the BV structure and fluctuations both in the position and the H-bonding interaction of the surrounding residues.

4 Conclusions

We have presented a comprehensive computational simulation of Resonance Raman, absorption, and CD spectra of the resting state of the dimeric phytochrome of *Deinococcus radiodurans*. To achieve such a goal, we have combined a $4.5\mu s$ MD simulation of the solvated phytochrome with a hybrid quantum mechanics/molecular mechanics (QM/MM) model, which accounts for both electrostatic and mutual polarization effects between the QM and the MM subsystems. This integrated approach has allowed to to clarify how structural fluctuations of the binding pocket of the biliverdin, the corresponding changes in its interactions with specific protein residues and water molecules, as well as its vibrations, are coupled to the excitation process.

In particular, a "pyrrole-water cage" has been found, formed by the mutual interactions between ARG466 (which is part of the PHY domain and only few residues away from the "tongue" region), ASP207 (which strongly H-bonds ARG466), HIS261 (which H-bonds the pyrrole water) and TYR263 (which H-bonds ASP207 and the pyrrole water). Whenever the cage is broken, the pyrrole water is destabilized and tends to exchange with other molecules or leave the binding site. In this context, it was recently suggested that one of the key-steps initiating the Pr-Pfr phytochrome conversion after the chromophore photo-isomerization is the pyrrole-water detachment from its binding site.[56] However, beside the important pyrrole-water cage connections, we do not find any significant change in the binding pocket or in the ASP207-O$\cdots$HN-ARG466 interaction, which could thus be directly responsible for the β-sheet to α-helix conversion. This finding does not exclude a connection between the pyrrole-water detachment and the phyctochrome conversion *after* the photo-isomerization has taken place. Another possibility is that the detachment occurs concurrently with the photo-isomerization, when the BV is in the excited state, as also suggested by the very short timescale observed for this detachment.[56]

Comparing the spectra simulated at the crystal structure and along the MD, we show that the chromophore is very stable in its binding pocket and that the fluctuations in the interactions (mostly H-bonds) with the surrounding residues and water molecules (including the pyrrole one) do not significantly affect the absorption and CD spectra. By performing a spectral density analysis we have also found that the lineshape of the Q band can be fully explained in terms of the vibronic coupling and it does not require to invoke the contemporary presence of other protonation states of BV, as it was recently proposed.[23] Our result is supported by the finding of the same broadening observed for other bilins,[57] and explained by vibronic coupling through the computation of the same spectral density approach here used. Finally, by comparing the calculated spectral density with the measured RR spectrum we could characterize the vibrations which are more strongly coupled to the excitation confirming the important role of the hydrogen-out-of-plane (HOOP) mode of the vinyl C-H in between C and D rings together with the expected C15=C16 stretching.

The completeness and accuracy achieved in this study show that the selected integrated approach which combines a long MD simulation with accurate QM/MMpol description represents a valid strategy to describe photo-induced multiscale processes in complex biosystems.[58] The natural following step, which in progress at the moment, is the extension of the same approach to investigate the downstream cascade of photochemical and conformational processes determining the photobiological activity of the phytochrome.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 M. A. van der Horst and K. J. Hellingwerf, *Accounts of Chemical Research*, 2004, **37**, 13–20.

2 A. Möglich, X. Yang, R. A. Ayers and K. Moffat, *Annual Review of Plant Biology*, 2010, **61**, 21–47.

3 J. T. M. Kennis and T. Mathes, *Interface Focus*, 2013, **3**, 20130005–20130005.

4 T. Kottke, A. Xie, D. S. Larsen and W. D. Hoff, *Annu. Rev. Biophys.*, 2018, **47**, 291–313.

5 N. C. Rockwell and J. C. Lagarias, *The Plant Cell*, 2006, **18**, 4–14.

6 N. C. Rockwell, Y.-S. Su and J. C. Lagarias, *Annu. Rev. Plant Biol.*, 2006, **57**, 837–858.

7 E. S. Burgie and R. D. Vierstra, *The Plant Cell*, 2015, **26**, 4568–4583.

8 K. Anders and L.-O. Essen, *Current Opinion in Structural Biology*, 2015, **35**, 7–16.

9 N. C. Rockwell, L. Shang, S. S. Martin and J. C. Lagarias, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 6123–6127.

10 M. E. Auldridge and K. T. Forest, *Critical Reviews in Biochemistry and Molecular Biology*, 2011, **46**, 67–88.

11 A. Björling, O. Berntsson, H. Takala, K. D. Gallagher, H. Patel, E. Gustavsson, R. St Peter, P. Duong, A. Nugent, F. Zhang, P. Berntsen, R. Appio, I. Rajkovic, H. Lehtivuori, M. R. Panman, M. Hoernke, S. Niebling, R. Harimoorthy, T. Lamparter, E. A. Stojković, J. A. Ihalainen and S. Westenhoff, *The Journal of Physical Chemistry Letters*, 2015, **6**, 3379–3383.

12 G. Gourinchas, S. Etzl and A. Winkler, *Current Opinion in Structural Biology*, 2019, **57**, 72–83.

13 H. Takala, A. Björling, O. Berntsson, H. Lehtivuori, S. Niebling, M. Hoernke, I. Kosheleva, R. Henning, A. Menzel, J. A. Ihalainen and S. Westenhoff, *Nature*, 2014, **509**, 245–248.

14 E. S. Burgie, T. Wang, A. N. Bussell, J. M. Walker, H. Li and R. D. Vierstra, *J. Biol. Chem.*, 2014, **289**, 24573–24587.

15 E. S. Burgie, J. Zhang and R. D. Vierstra, *Structure*, 2016, **24**, 448–457.

16 A. V. Nemukhin, B. L. Grigorenko, M. G. Khrenova and A. I. Krylov, *The Journal of Physical Chemistry B*, 2019, **123**, 6133–6149.

17 O. Falklöf and B. Durbeej, *ChemPhysChem*, 2016, **17**, 954–957.

18 J. A. Ihalainen, E. Gustavsson, L. Schroeder, S. Donnini, H. Lehtivuori, L. Isaksson, C. Thöing, V. Modi, O. Berntsson, B. Stucki-Buchli, A. Liukkonen, H. Häkkänen, E. Kalenius, S. Westenhoff and T. Kottke, *Journal of the American Chemical Society*, 2018, **140**, 12396–12404.

19 R. A. Matute, R. Contreras, G. P. Rez-Hernández and L. González, *J. Phys. Chem. B*, 2008, **112**, 16253–16256.

20 R. A. Matute, R. Contreras and L. González, *J. Phys. Chem. Lett.*, 2010, **1**, 796–801.

21 O. Falklöf and B. Durbeej, *Journal of Computational Chemistry*, 2013, **34**, 1363–1374.

22 I. V. Polyakov, B. L. Grigorenko, V. A. Mironov and A. V. Nemukhin, *Chem. Phys. Lett.*, 2018, **710**, 59–63.

23 V. Modi, S. Donnini, G. Groenhof and D. Morozov, *J. Phys. Chem. B*, 2019, **123**, 2325–2334.

24 C. Wiebeler, A. G. Rao, W. Gärtner and I. Schapiro, *Angew. Chemie - Int. Ed.*, 2019, **58**, 1934–1938.

25 C. Wiebeler and I. Schapiro, *Molecules*, 2019, **24**, 1720.

26 C. Curutchet, A. Muñoz-Losa, S. Monti, J. Kongsted, G. D. Scholes and B. Mennucci, *Journal of Chemical Theory and Computation*, 2009, **5**, 1838–1848.

27 S. Caprasecca, S. Jurinovich, L. Lagardère, B. Stamm and F. Lipparini, *J. Chem. Theory Comput.*, 2015, **11**, 694–704.

28 F. Lipparini, *J. Chem. Theory Comput.*, 2019, **15**, 4312–4317.

29 R. Sánchez and A. Šali, *Proteins Struct. Funct. Genet.*, 1997, **29**, 50–58.

30 R. Anandakrishnan, B. Aguilar and A. V. Onufriev, *Nucleic Acids Res.*, 2012, **40**, 537–541.

31 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

32 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **56531**, 1157–1174.

33 J. Wang, W. Wang, P. a. Kollman and D. a. Case, *J. Am. Chem. Soc*, 2001, **222**, U403.

34 R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2013, **3**, 198–210.

35 C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.

36 D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, V. W. D. Cruzeiro, T. A. Darden, R. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *AMBER 2018*, 2018, University of California, San Francisco.

37 T.-S. Lee, D. S. Cerutti, D. Mermelstein, C. Lin, S. LeGrand, T. J. Giese, A. Roitberg, D. A. Case, R. C. Walker and D. M. York, *J. Chem. Inf. Model.*, 2018, **58**, 2043–2050.

38 L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H.-B. Li, L. Ding and K. Morokuma, *Chemical Reviews*, 2015, **115**, 5678–5796.

39 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16 Revision A.03*, 2016, Gaussian Inc. Wallingford CT.

40 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

41 J. Tomasi, B. Mennucci and R. Cammi, *Chemical Reviews*, 2005, **105**, 2999–3094.

42 D. Krois and H. Lehner, *Journal of the Chemical Society, Perkin Transactions 2*, 1993, 1351–1360.

43 S. Mukamel, *Principles of Nonlinear Optical Spectroscopy*, Oxford University Press, New York, 1995.

44 D. Von Stetten, S. Seibeck, N. Michael, P. Scheerer, M. A. Mroginski, D. H. Murgida, N. Krauss, M. P. Heyn, P. Hildebrandt, B. Borucki and T. Lamparter, *J. Biol. Chem.*, 2007, **282**, 2116–2123.

45 J. B. Page and D. L. Tonks, *The Journal of Chemical Physics*, 1981, **75**, 5694–5708.

46 R. Scholz, L. Gisslén, B.-E. Schuster, M. B. Casu, T. Chassé, U. Heinemeyer and F. Schreiber, *The Journal of Chemical Physics*, 2011, **134**, 014504.

47 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.

48 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615.

49 A. D. Becke, *Phys.Rev. A*, 1988, **38**, 3098–3100.

50 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

51 R. L. Martin, *J. Chem. Phys.*, 2003, **118**, 4775–4777.

52 J. R. Wagner, J. Zhang, D. von Stetten, M. Günther, D. H. Murgida, M. A. Mroginski, J. M. Walker, K. T. Forest, P. Hildebrandt and R. D. Vierstra, *Journal of Biological Chemistry*, 2008, **283**, 12212–12226.

53 D. O. Kashinski, G. M. Chase, R. G. Nelson, O. E. Di Nallo, A. N. Scales, D. L. Vanderley and E. F. Byrd, *J. Phys. Chem. A*, 2017, **121**, 2265–2273.

54 J. Dasgupta, R. R. Frontiera, K. C. Taylor, J. C. Lagarias and R. A. Mathies, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 1784–1789.

55 C. Kneip, P. Hildebrandt, W. Schlamann, S. E. Braslavsky, F. Mark and K. Schaffner, *Biochemistry*, 1999, **38**, 15185–15192.

56 C. E. et all, *biorXiv*, 2019, 1–41.

57 M. Corbella, L. Cupellini, F. Lipparini, G. D. Scholes and C. Curutchet, *ChemPhotoChem*, 2019, **3**, 945–956.

58 B. Mennucci and S. Corni, *Nature Reviews Chemistry*, 2019, **2**, 1.