

The Multiple Roles of the Protein in the Photoactivation of Orange Carotenoid Protein

Mattia Bondanza, Lorenzo Cupellini, Filippo Lipparini, and Benedetta Mennucci*

Università di Pisa, Dipartimento di Chimica e Chimica Industriale

E-mail: benedetta.mennucci@unipi.it

Abstract

Orange Carotenoid Protein (OCP) is a carotenoid-binding protein involved in photo-protection mechanisms of cyanobacteria. Upon exposure to high-light, OCP interconverts from an orange resting form (OCP^O) to a red active one (OCP^R); the mechanism of this interconversion, even if extensively studied, has still not been fully elucidated. The comparison of the spectroscopic properties of OCP^O and OCP^R also offers a very intriguing problem: upon interconversion, the UV-VIS absorption spectrum of the carotenoid not only undergoes a significant red-shift, but its line shape also changes, losing the vibronic structure and becoming very similar to the one recorded for the same carotenoid in solution. In this work we use multiscale atomistic models ranging from classical to polarizable QM/MM molecular dynamics not only to reproduce the experimental spectra but also to give a comprehensive molecular explanation of such drastic spectroscopic differences. The findings are finally used to formulate a new hypothesis on the role of the protein in the photoactivation mechanism.

Introduction

In oxygenic photosynthetic organisms, the harvest of sunlight is accompanied by photoprotective mechanisms which prevent potential damages under high light conditions. Plants, algae and bacteria have developed a strategy, known as nonphotochemical quenching (NPQ), which is used to dissipate excess energy and prevent oxidative damage¹⁻³. Due to the different living conditions, however, the various organisms have optimized their own mechanisms. For example, in plants and algae, the NPQ is performed by the pigment-protein complexes of the photosystems placed within the thylakoid membrane⁴⁻⁶. In cyanobacteria, instead, the photoprotective mechanism in action is realized by a small water-soluble carotenoid-containing protein known as Orange Carotenoid Protein (OCP)⁷⁻⁹. The presence of the carotenoid allows the complex to absorb 400-600 nm light and the absorbed energy is finally used to generate a dramatic structural change in the protein leading to the red active form (OCP^R)¹⁰⁻¹³. In particular, OCP contains two different domains: an all α -helical N-terminal domain and a mixed α -helical/ β -sheet C-terminal domain. In the inactive orange form, the domains are in contact and they encapsulate the non-covalently bound carotenoid¹⁴. Upon photoactivation, the protein undergoes a conformational change: the two domains separate and the carotenoid moves into the N-terminal domain (NTD). It is exactly this large (ca. 12 Å) translocation that allows the carotenoid to be-

come active as a quencher for the excited bilins of the phycobilisome (PBS) antenna complexes.

Notwithstanding the large amount of data collected so far on the OCP photoactivation and the following structural changes, many aspects of the whole process are still unsolved; in particular, there is not yet any decisive evidence about the real nature of the events that, after the electronic excitation of the carotenoid, activate the structural dynamics in the protein, the carotenoid translocation, and the domain separation^{13,15,16}. What is known is that OCP can bind various carotenoids, but its photoprotective function is active only in the presence of a keto-carotenoid, in particular, 3'-hydroxyechinenone, 3hECN; echinenone, ECN; and canthaxanthin, CAN. The main effect of the carbonyl substituents in the end-ring(s) is that hydrogen bonding interactions with specific residues in the C-terminal domain (CTD) can be formed (see Figure 1d-e), thus indirectly suggesting that these interactions are fundamental for the light-driven process to occur.

Very recently, a combination of time-resolved UV-visible and mid-infrared spectroscopy has been employed to assess the electronic and structural dynamics of OCP, from femtoseconds to 0.5 ms.^{17,18} What emerges from such a study is a model with a fast (ps) hydrogen-bond rupture at the carotenoid end-ring placed in the CTD domain (the β_1 -ring). This event, even if occurring at a low yield of <1%, initiates structural changes in the protein, finally allowing the carotenoid to translocate into the N-terminal domain in a μ s scale. These findings suggest

that one of the main contributors to the success of the whole process is the coupling between electronic and structural dynamics within the carotenoid, and short and long range interactions with the embedding composite system (the protein matrix and the water solvent). As a matter of fact, the inactive and active forms of OCP show a different UV-VIS absorption spectrum in the carotenoid region, in terms of both a significant shift in the position of the band and a change in its shape. In particular, the inactive form presents a lineshape with a clear vibronic structure centered at 480 nm while in the active form the same band is shifted at 520 nm and loses any vibronic structure. The Resonance Raman spectra of the two forms, despite some differences,^{9,19} show very similar features; an analogous spectral structure is observed also for the CAN in THF solvent (see Figure 1a-b). This should suggest that similar vibronic couplings are present for the two form of OCP and for the CAN itself. These two pieces of evidence are somehow contradictory and up to now a clear explanation of the molecular origins of such behavior has not been given.

While the atomistic structure of the inactive form of OCP (from now on OCP^O) was solved using X-ray diffraction techniques^{10,13,20–22}, the one of the light-activated form (OCP^R) was never reported in literature. Nonetheless, some authors have identified a protein called Red Carotenoid Protein (RCP) that, on the basis of spectroscopic and biological properties, is considered a faithful model for the NTD of OCP^R ; the crystallographic structure of RCP was reported¹⁰.

In the present study, we give a comprehensive explanation of the spectroscopic differences between OCP^O and RCP in terms of protein-induced effects both on the slow conformational dynamics of the carotenoid and the relative stabilization of its ground and excited states. To achieve such a picture we combine a multiscale strategy which integrates quantum-mechanical and classical descriptions for the carotenoid and the environment, respectively, with (enhanced) molecular dynamics. This strategy has the important characteristic of accounting for the electrostatic effects and mutual polarization be-

tween the two subsystems, while still keeping a fully atomistic detail of the protein and the solvent.^{23,24} After having successfully reproduced the experimental spectra of the two complexes, we use the obtained results to formulate a new hypothesis for the photoactivation mechanism. According to this hypothesis, the constraints of the binding pocket and the unfavorable electrostatic fields acting on CAN in the OCP^O are the driving force leading to the dissociation and separation of C-terminal and N-terminal domains accompanied by carotenoid translocation into the N-terminal domain.

Results and Discussion

Conformational effects

In the Introduction we have pointed out the contrasting behavior found in the UV-Vis and Resonance Raman spectra for OCP^O and OCP^R (and RCP). Interestingly, this behavior is very similar to what observed for homologous carotenoids of CAN (same chain length and possessing a cyclic terminal group) in solution as recently shown by Uragami et al.²⁵ According to the latter study, the differences in absorption lineshape in this class of carotenoids are due to different inhomogeneous broadening, as all the carotenoids presented very similar Resonance Raman spectra and thus very similar vibrational modes coupled to the excitation. Moreover, because all the spectra were recorded in the same low polar and non-protic solvent (THF), the variability in inhomogeneous broadening was related to a different conformational space explored by the carotenoids. In particular, the rotation of the terminal rings around the single C–C bonds (corresponding to dihedral angles β_1 and β_2 reported in Figure 1f) was intuitively selected as the fundamental degree of freedom whose slow conformational dynamics can affect the excitation energy.

Here, we hypothesize that the conformational freedom along this dihedral motion is not only the origin of the specific inhomogeneous broadening of each solvated carotenoid but it can also explain the differences in the absorption spectra

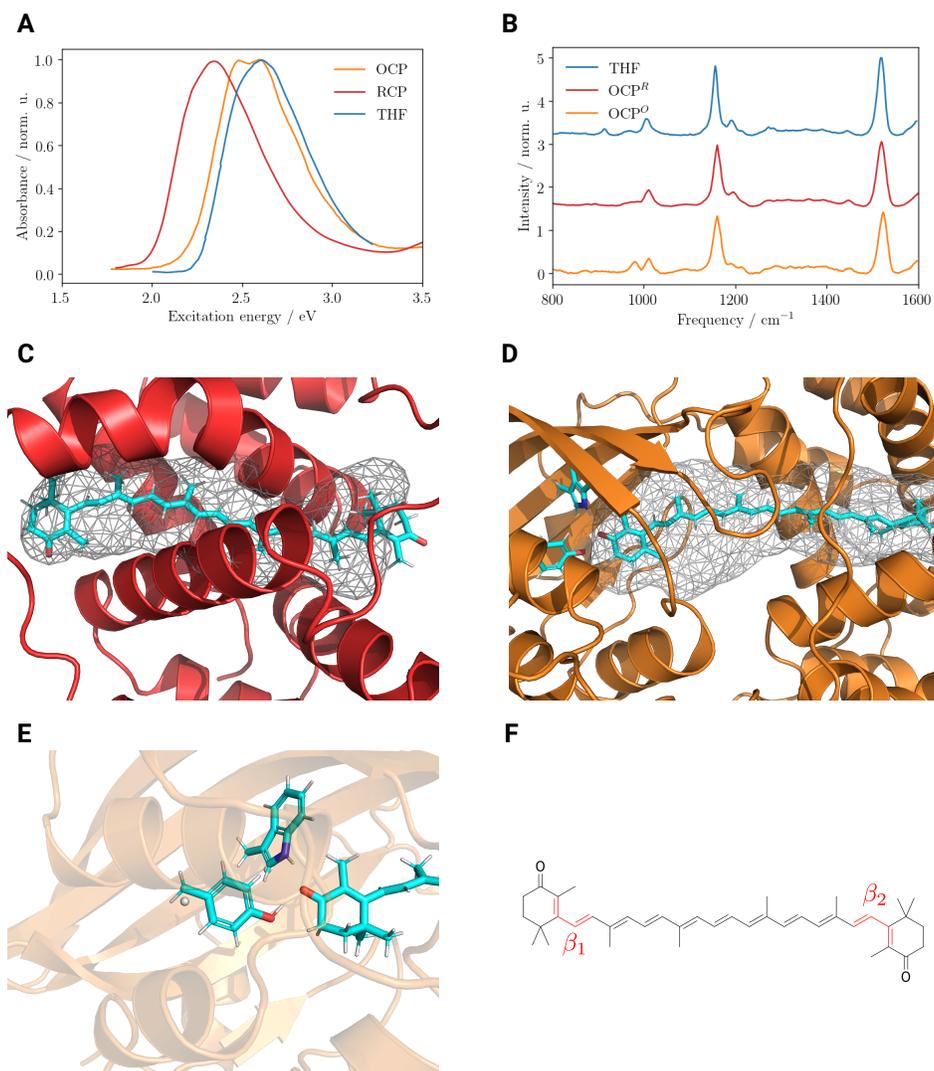


Figure 1: (a) Comparison of UV-VIS spectra of $\text{OCP}^{\text{O}10}$, RCP^{10} and CAN in THF^{25} . (b) Comparison of resonance Raman spectra of CAN in THF^{25} , RCP^{15} and $\text{OCP}^{\text{O}15}$. (c, d) Representation of CAN in the pocket of RCP and OCP^{O} respectively (the structures shown are $4\text{XB}4$ and $4\text{XB}5^{10}$): the mesh visualizes the internal cavity defined by the residues closest ($< 5 \text{ \AA}$) to the chromophore. In (d) the side chains of two hydrogen-bonding residues are also drawn. (e) Detail of the hydrogen bonds between $\text{Tyr}201$ and $\text{Trp}288$ of OCP^{O} and CAN . The β_1 ring is always on the left side, while β_2 is on the right. (f) Structure of CAN with the two dihedral angles defining the rotation of the β_1 and β_2 rings underlined in red. From now on these angles will be indicated with the same notation of the corresponding rings.

between OCP^{O} and RCP (see Figure 1a).

In order to validate this hypothesis, we have proceeded in two steps.

As a first step, we have investigated four carotenoids among the studied homologous compounds, selecting those showing very different absorption band shapes, namely, violaxanthin (VIO) for which the vibronic structure is very well-resolved, β -carotene (CAR), zeaxanthin (ZEA) for which the same structure is less

sharp and, finally, CAN , for which it is almost completely lost (in Figure S1 the structures of these carotenoids are shown; their UV-Vis spectra are reported in Figure S2). For these four carotenoids, we have performed a relaxed scan of the dihedral angle (due to the symmetric nature of all the selected carotenoids, only one scan is needed) and calculated the excitation energy along the scan. In all these calculations we have neglected the effect of the THF solvent

as it is expected to be negligible. The calculated data, reported in Figure S2, show that the energy curve along the dihedral angle is very similar for CAN, CAR, and ZEA. On the contrary, VIO presents a peculiar behavior due to the presence of epoxydic groups, which strongly reduce the dihedral flexibility. More in detail, the scans show that the first three carotenoids have two similar minima for the *s*-cis conformation at about $\pm 50^\circ$ (*s*-cis⁺ and *s*-cis⁻), as a perfectly planar conformation is prevented by steric clashes between methyl groups.

The other minimum is a *s*-trans conformation. Interestingly, we note that the position of this minimum along the dihedral angle coordinate is influenced by another conformational parameter. In fact the six-term rings at the ends of the conjugate chain of the carotenoid are flexible and present two possible *half-boat* conformations. The presence of only two conformations instead of the many that are normally allowed for a six-term ring is due to the three rigid sp²-carbons in the ring. In order to quantify this conformational freedom we used the generalized ring puckering coordinate originally introduced by Cremer and Pople²⁶. Since for CAN only two stable ring conformations are possible, only one of the angles used in the definition by Cremer and Pople is sufficient to fully describe this degree of freedom: we refer to this angle (corresponding to the θ of Cremer and Pople) as *puckering* and we identified the two conformations as p⁻ ($\theta = 75^\circ$) and p⁺ ($\theta = 105^\circ$). Due to steric effects, the puckering is coupled with the dihedral angle in the *s*-trans region (see Figure S3b). When the ring is in p⁻ conformation, the *s*-trans minimum is located at 150° (we call this minimum *s*-trans⁺), vice versa when the ring is in p⁺ the minimum shifts to -150° (*s*-trans⁻).

In CAN, however, the *s*-trans minima are much lower than in the two other carotenoids, and the barrier between the two *s*-cis minima is also considerably smaller. A simple explanation is that the *s*-trans minima are nearly planar and therefore take advantage of the conjugation of the system: the stability gained by CAN due to the planarity of the ring is much higher than for CAR or ZEA, thanks to the addition of a

further (C=O) double bond. When the scans are combined with calculations of excitation energies, we can conclude that the two quasi-symmetric *s*-cis isomers behave very similarly, while the planar *s*-trans conformer(s) have a much lower excitation energy.

From this analysis, it emerges that the inhomogeneous broadening of the absorption spectra can indeed be explained in terms of different conformers with different excitation energies: the single conformer of VIO gives the least broadened spectrum with a clear vibronic structure, while the two similar CAR and ZEA give more broadened spectra that, however, still show a partially recognizable vibronic structure. In CAN, which is the only one with four possible different conformers, the details of the vibronic structure are instead lost, and the spectrum appears as a broad band.

In order to generalize this interpretative model to OCP^O and RCP, we have to set up a proper computational protocol to investigate the conformational space of CAN in the two proteins and to quantify its impact on the spectra. Here, however, the problem is completely reversed: instead of having different carotenoids in the same apolar solvent, we have the same carotenoid in two different (and more complex) environments, namely the two protein matrices in water.

To properly account for the effects of the protein (and the solvent) we have run 1 μ s-long classical molecular dynamics (MD) in explicit water (see Methods). It is here important to note that while the protein was described with standard molecular mechanics (MM) force fields, the one used for CAN was reparameterized using DFT data as reference values (see the Supporting Information for details):²⁷ in this way we make sure that the degrees of freedom of CAN (and in particular the torsional ones) explore the correct potential energy surface.

From the two MD simulations, we have found that, in OCP^O, both terminal rings show some conformational flexibility, namely the same puckering observed in the previous scan of the isolated carotenoid (see Figure 2a and the RMSD-2D plot of Figure S11 in the Supporting Information). On the contrary, the β_1 and

β_2 dihedral angles remain fixed in one stable conformation (Figure 2b) along the entire (and long) trajectory, in spite of the relatively low barriers for their rotation, namely *s*-trans for β_1 and *s*-cis for β_2 . A rigidity was indeed expected for β_1 , due to the constraints imposed by hydrogen bonding with the residues Tyr201 and Trp288. However, also the apparently looser β_2 remains in the same conformation found in the crystal structure showing that the NTD pocket induces a rigidity in the carotenoid even in absence of specific H-bonds.

From the same MD trajectory, the free energy surface (FES) on the dihedral and puckering subspaces was calculated and the different configurations of the CAN were classified using the density peak clustering algorithm²⁸ (See the Supporting Information for details on the clustering). Four clusters (A-D) were identified based on the different ring puckering and dihedral angles (Figure 2a and 2b)

For RCP, instead, several conformations were observed along the MD trajectory, showing that CAN in RCP has a much larger conformational freedom with respect to OCP^O. However, the transitions between the different conformations are too rare to assess convergence (see Figure S9). To solve this problem and to achieve a quantitative sampling also for CAN in RCP, we used an enhanced sampling technique, namely well-tempered Metadynamics (WT-MetaD)²⁹. Unfortunately, the exploration of the four-dimensional subspace defined by the puckering and dihedral coordinates of the two rings is unfeasible because of the too high computational cost; therefore we explored separately the puckering/dihedral subspace for each ring (Figure 2c and 2f) and then we reconstructed the full conformational space under the hypothesis that the two rings' conformations are uncorrelated.

The FES along the two dihedral angles is presented in Figure 2d. From these results it is clear that, in RCP, both rings of CAN can rotate almost freely, and many different conformers are significantly populated at room temperature, similarly to CAN in solution. Configurations were finally individuated and classified using the same clustering algorithm as before and

five main clusters were obtained (A-E) based on the different ring conformation (Figure 2c-2f).

These findings seem to explain the apparent mismatch between the experimental results of Leverenz *et al.*¹⁰ and Konold *et al.*³⁰. In fact, the latter group has proposed a mechanism for the interconversion of OCP^O into OCP^R that does not involve the *s*-isomerization of the terminal bond, which should remain in the *s*-trans conformation; this mechanism is also supported by many previous experimental findings¹³. However, if RCP was a good model for OCP^R, as suggested by Leverenz *et al.*, the β_1 dihedral of CAN should isomerize to the *s*-cis conformation, as observed in the crystal structure of RCP.¹⁰ In our dynamical representation, RCP, and likely OCP^R, exists in multiple conformations, which are able to interconvert on a μ s timescale. One of these conformations (E) corresponds to the RCP crystal. In the other ones, β_1 has the same conformation as in OCP^O. This allows a comprehensive rationalization of the findings of the two groups, as it shows that an equilibrium between different conformations including the one crystallized by Leverenz *et al.* is possible in OCP^R.

Based on the conformational and clustering analysis here presented, it was finally possible to extract a set of representative configurations for both complexes and use them to calculate the excitation energies. Before that, however, a preliminary refinement of the geometry of CAN was necessary. In fact, in order to achieve accurate estimates of the excitation energies, an accurate geometry of the chromophore is needed. For highly conjugated systems like CAN, this accuracy cannot be achieved at the MM level, even if the FF is optimized on QM data as we have done for CAN. Therefore, for all the different configurations selected to properly represent the populations of the clusters of OCP^O and RCP, the geometry of CAN was optimized at the DFT/MM level, keeping the protein and solvent frozen (see the Supporting Information for details). The resulting relaxed configurations were finally used to calculate the excitation energies at the TD-DFT level. The calculations were performed for the isolated CAN (vac) and the CAN in the presence of the protein and

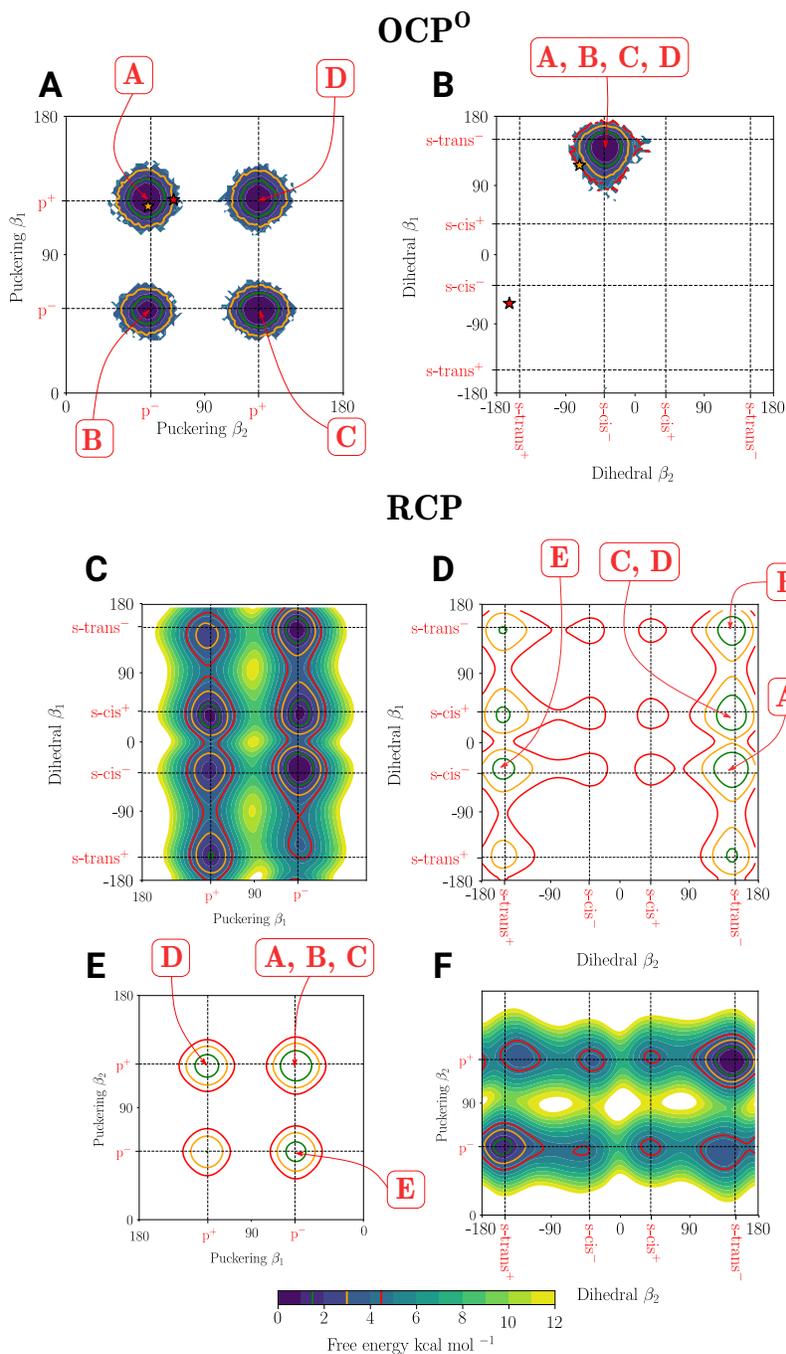


Figure 2: Free energy surfaces of CAN’s terminal ring conformational parameters in RCP and OCP. Puckering of the β_1 ring vs puckering of the β_2 ring (a) and dihedral β_1 vs dihedral β_2 (b) FESs in OCP^O , calculated from a 1 μs long MD trajectory; puckering β_1 vs puckering β_2 (e) and dihedral β_1 vs dihedral β_2 (d) FESs in RCP, reconstructed from the puckering vs dihedral FESs of β_1 (c) and β_2 (f) calculated with well-tempered metadynamics. The red labels indicate the free energy basin of each conformer.

the solvent (env) (see the Methods section for details). All the results are reported together with cluster populations in Table 1.

All four clusters of OCP^O show very similar excitation energies, with the same variance. Therefore, in this complex, the conformational

variability has only a minor effect on the energetic disorder. In RCP, the excitation energies are significantly red-shifted with respect to OCP^O ; furthermore, they are much more spread in the various clusters than in OCP^O . This latter observation suggests that the larger

Table 1: Populations (%) of the clusters of OCP^O and RCP. Excitation energies (EE in eV) (and standard deviations) calculated on the relaxed configurations of the different clusters: two sets of data are reported and they refer to calculations performed on the isolated CAN (vac) and the CAN in the presence of the protein and the solvent (env)

Cluster	Population	EE (env)	EE (vac)
OCP^O			
(A)	30	2.26 ± 0.06	2.45 ± 0.03
(B)	13	2.24 ± 0.07	2.44 ± 0.04
(C)	33	2.23 ± 0.05	2.42 ± 0.03
(D)	24	2.24 ± 0.08	2.44 ± 0.03
Average	–	2.24 ± 0.07	2.44 ± 0.03
RCP			
(A)	41	2.12 ± 0.08	2.39 ± 0.03
(B)	15	2.14 ± 0.07	2.36 ± 0.02
(C)	11	2.04 ± 0.12	2.37 ± 0.03
(D)	12	2.03 ± 0.10	2.36 ± 0.02
(E)	6	2.11 ± 0.05	2.37 ± 0.01
Average	–	2.10 ± 0.09	2.37 ± 0.03

conformational variability in RCP can indeed explain the observed band broadening.

Environment effect on excitation energies

Before moving to the simulation of the spectra, we present an analysis of the environment effects on the excitation of CAN.

It is known that the excitation to the lowest bright state of carotenoids (the one seen in one-photon absorption spectra) is strongly affected by the polarizability of the solvent³¹ and the observed red shifts strongly depend on the refractive index of the solvent.³² Following this evidence, we have checked whether this effect contributes to the overall shift between OCP^O and RCP. Since we are here using a polarizable MM embedding where each atom of the environment is described by fixed multipoles and an atomic polarizability, we can easily disentangle the induction from the electrostatic effect.²³ By doing this analysis on both OCP^O and RCP, we found that the effect of the polarizability is, on average, a substantial redshift (~ 0.1 eV) for both systems, but the difference between OCP^O

and RCP is quite modest.

On the basis of this result, we shifted the rest of the analysis on electrostatic effects, with a major focus on which residues have a greater influence in the two complexes.

For all the representative structures of CAN in OCP^O and RCP, we estimated the electrostatic contribution of each residue by computing the interaction of the MM charges of the residue with the difference in the density from ground to excited state. The average contributions over the different configurations, for the residues of OCP^O and RCP that contribute most, is given in Figure 3a,b. With the exception of some close residues, the largest contributions are given by charged residues in both complexes.

Interestingly, the residues that contribute the most to the redshift of the excitation energy are the positively charged ones (Arg, Lys) on the β_1 side, such as arginines 155 (the largest contribution), 9, 185 and 289 in the case of OCP^O , and the negatively charged residues (Asp, Glu) on the β_2 side, such as Glu34 and Asp35, which have a stronger effect on CAN in RCP being closer to the carotenoid backbone. On the

contrary, the Asp and Glu residues on the β_1 side (e.g. Asp304 and Glu244 in OCP, Glu146 in RCP) tend to blue shift the excitation energy. This observation suggests an asymmetry in the change of CAN’s electron density upon excitation, with a redistribution of charge from β_2 towards the β_1 ring. This conclusion is unexpected given the symmetry of canthaxanthin: clearly, such an asymmetry must be induced by the electrostatic field of the protein (See Figure 3e,f) on the ground-state of canthaxanthin. We confirmed this hypothesis by noting that the vacuum-to-environment solvatochromic shift of CAN was strongly correlated with the ground-state dipole moment induced by the environment (See Figure S15 in the Supporting Information).

Here, it is also interesting to recall that Arg155 forms, together with Glu244, the salt bridge that contributes to keep together the CTD and NTD in OCP^O. From our analysis, these residues have opposite effects on the solvatochromic shift of CAN in OCP^O. Upon dissociation of the CTD, the charge of Arg155 remains unbalanced, and, together with the translocation of CAN into the NTD, contributes to the further redshift in RCP. Translocation into the NTD also brings the β_2 end of the carotenoid closer to the negatively charged residues Glu34 and Asp35, which are located on the opposite face of the NTD.¹⁰

In order to complete our analysis of environment effects on the spectral differences between OCP^O and RCP, we consider the contribution of residues to the stabilization of the ground state. These data, reported in Figure 3c,d, show that in RCP the residues contributing the most to the solvatochromic shift are also the ones that stabilize the ground state (Arg155, Glu34, Asp35). On the contrary, in OCP^O, the hydrogen-bonding residues Tyr201 and Trp288 which are the main ones for the stabilization of the ground state, do not influence significantly the excitation energy. This can be explained by the partially broken conjugation of CAN in OCP^O, which makes the excitation less responsive to the electrostatic perturbations around the β_1 ring. On the other hand, the strongly interacting charged residues in RCP are closer

to the ends of the CAN backbone, which clearly enhances their contribution to the excitation energy.

Comparing the ground-state contributions in Figure 3c,d, it is evident that, at least regarding electrostatics, the positioning of CAN in the RCP cavity allows for stabilization of the ground state by several residues, whereas in OCP^O only the hydrogen-bonding Trp and Tyr give a substantial contribution to the stabilization. We therefore propose that, after breaking the two hydrogen bonds, electrostatic interactions will drive the translocation of CAN towards the NTD. Indeed, single-point mutagenesis showed that Glu34, which is one of the residues with the strongest electrostatic interaction in our model of RCP, was essential to stabilize the red form of OCP.¹⁰ Also Pro126 presents an analogous situation.

As a last note, we observe that OCP^O and RCP are also very different in terms of the exposition of CAN to water and this could also contribute to the overall redshift observed in RCP. From our calculations, instead, it comes out that the effect of the solvent is to reduce the solvatochromic shift given by the protein (see SI for details). This situation, which is somewhat counterintuitive standing the strong polarity of water, is explained by the reorganization and polarization of water molecules, which lead to partial screening of the protein electrostatic field.

Resonance Raman and absorption spectra

The previous analysis has been based on vertical excitation energies and environment-induced broadening, however, to simulate the absorption spectra, we need to include vibronic coupling as well.

Here, we calculate the vibronic coupling of CAN in OCP^O and RCP in terms of the “spectral density” (SD). The SD measures the frequency-dependent coupling of the nuclear motions to the excitation and, as such, it can be used to define a lineshape function (see SI for the details). From a computational point of view, it can be calculated from the Fourier

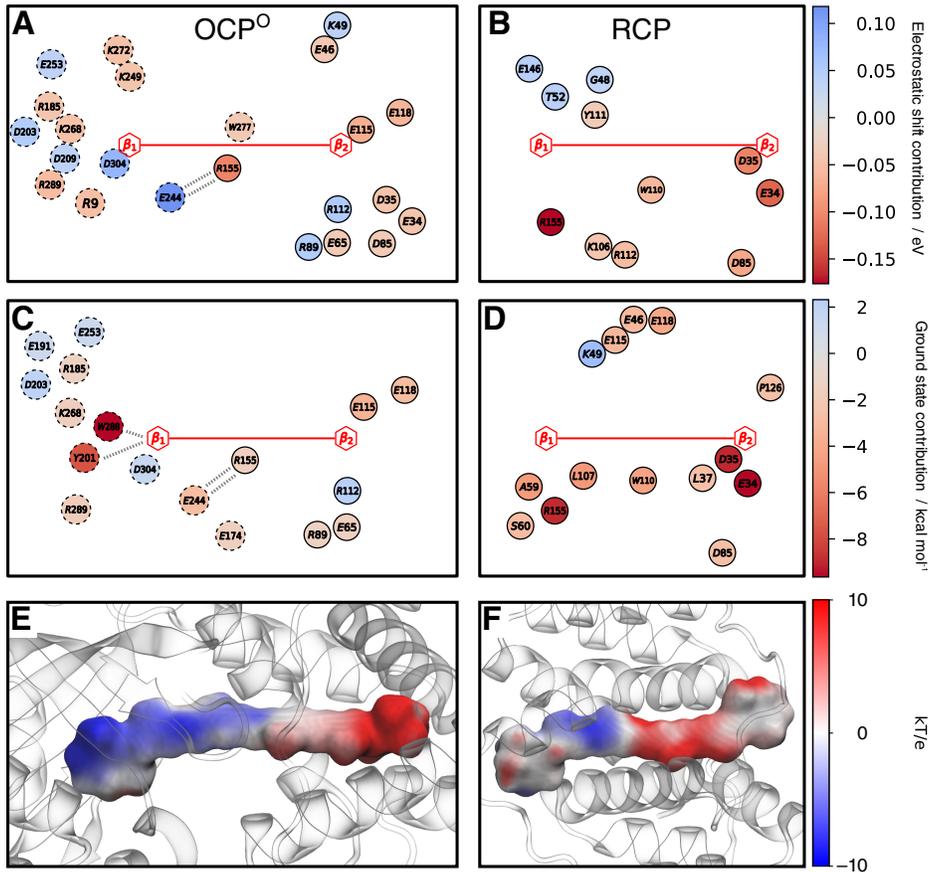


Figure 3: (a,b) Representation of the residues that contribute the most to the solvatochromic shift in (a) OCP^O and (b) RCP. Each circle represents a residue in the protein, and the residue name is given with the one letter code (R=Arg, L=Lys, D=Asp, E=Glu, Y=Tyr, W=Trp). Residue numbering follows the OCP sequence also for RCP for better comparison. The color scale represents the electrostatic contribution to the solvatochromic shift (blue, positive; red, negative). The circles are positioned according to the position of the residues around the CAN. (c,d) Same as (a,b), but here the color scale represents the polarizable QM/MM contribution to the ground-state energy. (e,f) Electrostatic potential generated by the protein charges on the CAN surface in (e) OCP^O and (f) RCP. The dotted lines indicate hydrogen-bonds between Tyr201 and Trp288 and CAN and the salt bridge between Arg155 and Glu244.

transform of the autocorrelation function of the excitation energies fluctuations. To obtain such fluctuations, we run Born-Oppenheimer MD trajectories of the CAN treated at DFT level within a polarizable MM environment for both the protein and the solvent (see Methods). Since the SD gives direct information on the displacement between ground and excited-state potential energy surfaces, and on the corresponding Huang-Rhys factors,³³ it can be directly related to Resonance Raman spectra³⁴⁻³⁶.

The calculated SDs (Figure 4a) reproduce all the four main groups of bands, termed ν_1 - ν_4 ,

of Raman spectra of carotenoids:³⁷ the two intense signals attributed to in-phase C=C (ν_1) stretching and C-C (ν_2) stretching coupled with C-H in-plane bending modes along the conjugate chain, respectively, the in-plane rocking vibrations of the methyl groups attached to the conjugated chain (ν_3) and the C-H out-of-plane wagging motions (HOOP) coupled with C=C torsional modes (ν_4).^{25,38} In particular, it is here worth noting that the calculations reproduce the change in the ν_4 peak observed moving from OCP^O to RCP. This change was attributed to the fact that the HOOP mode couples to the excitation only in the bent struc-

ture of CAN in OCP^O.³⁸ A normal-mode analysis (NMA), performed on two representative trajectories of CAN in the two systems, confirmed this attribution showing that the lower-frequency peak at $\sim 1000\text{ cm}^{-1}$ in OCP^O corresponds to normal modes involving HOOP motions of the conjugated backbone (See Supporting Information).

We note, however, that the spectral densities present a general blue-shift (probably due to the selected QM method) and do not accurately describe the ν_1 mode, which is broader than in the experiment. Given the necessarily limited sampling of our polarizable QM/MM MDs, it is possible that some extreme configurations are overrepresented in our calculations, giving rise to a broader band.

Integrating the spectral density, it is also possible to calculate the reorganization energy λ , which quantifies the magnitude of the total exciton-phonon coupling. As the absolute value of λ has a significant impact on the band-shape, we have performed an analysis of the sensitivity of the SD in general and λ in particular, on the QM method used for calculating the excitation energies (here TD-DFT). From this analysis (see Supporting Information) it was possible to define a scaling factor of 0.9 which has been used for the TD-DFT SD of both OCP^O and RCP. Moreover, because the low frequency part of the SD comes from very slow (inter)molecular motions that are poorly sampled in the QM/MM MDs, we replaced the SD below 600 cm^{-1} with an overdamped brownian oscillator³³ with a damping constant of 100 cm^{-1} , and with $\lambda_{slow} = 100\text{ cm}^{-1}$ and 500 cm^{-1} for OCP^O and RCP respectively. The different value of λ_{slow} used for the two complexes is justified by the larger variability of the excitation energies calculated in RCP, as shown in Table 1. It is interesting to note that, according to the calculations, the total reorganization energy of RCP (2413 cm^{-1}) is lower than the one of OCP^O (2670 cm^{-1}).

From the spectral densities we finally calculated homogeneous lineshapes reported in Figure 4b. Then, assuming that the homogeneous lineshape is not significantly dependent on the specific cluster, we reproduced the absorption

spectra of OCP^O and RCP by convoluting it with the inhomogeneous frequency distribution of excitation energies reported in Table 1. To account for the effect of the environment disorder, the contribution of each cluster was broadened with a Gaussian distribution whose standard deviation is obtained from the variance of excitation energies within the cluster. The same analysis was performed using the excitation energies calculated both in the environment and *in vacuo*. The comparison of these results allows us to disentangle the environment and the geometrical effects on the absorption spectrum.

As shown in Figure 4c,d both conformational and environmental disorder seem to significantly contribute to the broadening of RCP spectrum; in particular, the large conformational disorder smooths the vibronic structure, which almost disappears when the environmental disorder is also taken into account. In OCP^O, instead, a partial vibronic structure is retained because of a lower variability of the excitation energy in the different conformations.

It should be noted that, in our model, the residual vibronic peaks of OCP^O correspond to the 0-1 and 0-2 vibronic bands of CAN, whereas the 0-0 band is barely visible on the low energy side of the absorption band. In the previous literature, these bands were assigned to the 0-0 and 0-1 vibronic transitions^{39,40}, whereas the broad red tail of OCP^O absorption was suggested to arise from another specific conformation of the complex with red-shifted absorption. This “red-shifted” form should be characterized by larger disorder and likely a looser binding pocket⁴⁰. Our calculations cannot exclude the presence of such a conformation in OCP^O even if it was not observed within our long simulation time. This seems to indicate that, if this conformation exists, it is substantially different from the main one (the one here characterized) and enhanced sampling techniques are needed to allow its exploration.

Our assignment of vibronic bands is consistent with the stronger vibronic coupling of CAN in OCP^O, and the appearance of the additional HOOP signal in the Resonance Raman spectrum. In fact, the 0-0 band is already weaker

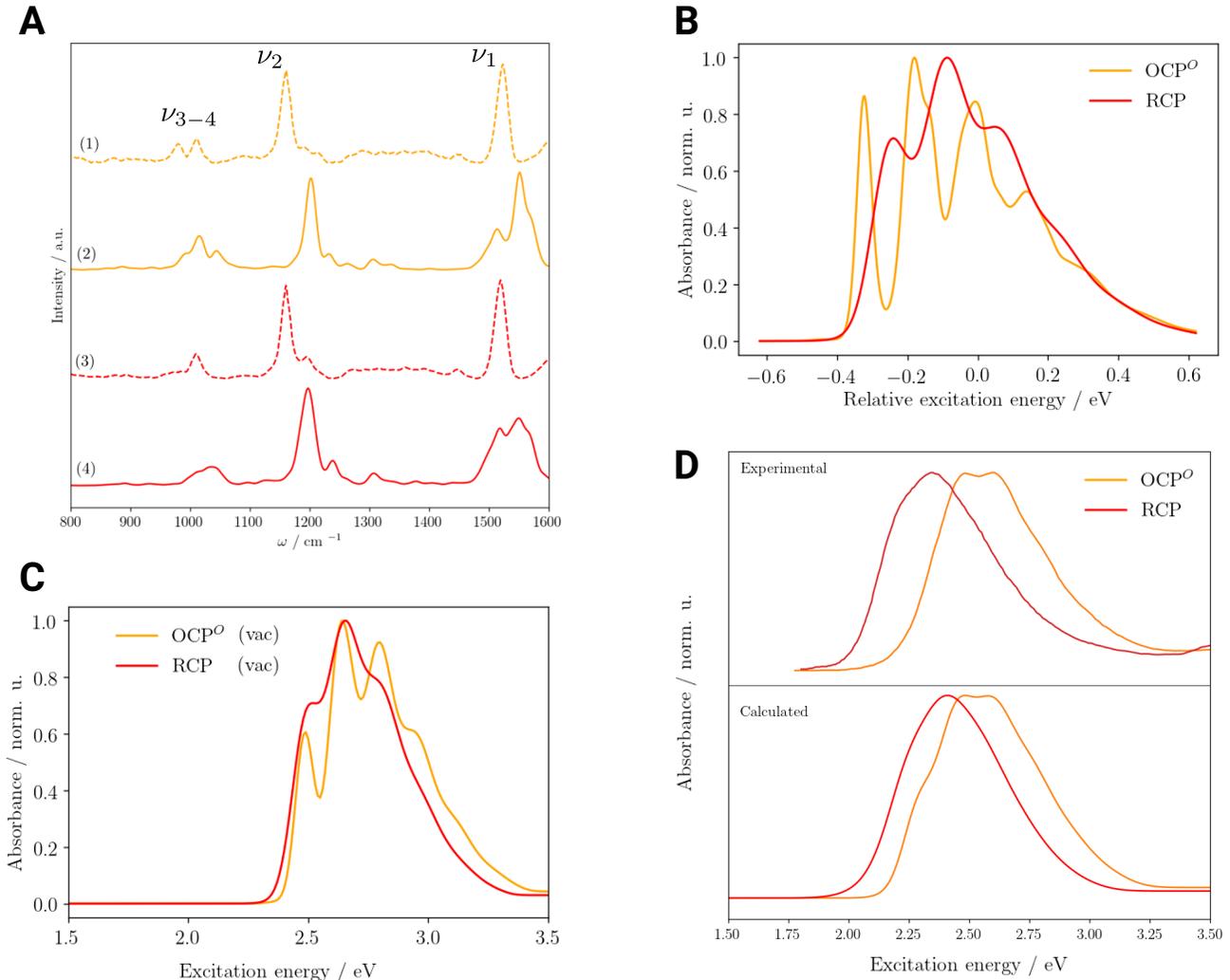


Figure 4: (a) Comparison of measured resonance Raman spectrum¹⁵ (1. OCP, 3. RCP) and calculated SD in the frequency range of 900-1600 cm^{-1} (2. OCP, 4. RCP). (b) Comparison of the homogeneous lineshapes of RCP and OCP. Spectra computed using vacuum (c) and QM/AMOEBA (d) excitation energies; the last ones are compared with experimental spectra.

than the 0-1 band in RCP. An increase in the vibronic coupling decreases the intensity of the 0-0 and 0-1 bands, making the 0-1 and 0-2 bands roughly equal in intensity. It is nevertheless possible that a “red” form of OCP^O, which gives rise to a small red-shifted and broadened spectral contribution, covers the 0-0 band of the major absorption band of OCP^O.

Conclusions

By combining classical and polarizable QM/MM dynamics with excited state calculations, we have explained the intriguing spectroscopic analogies and differences between OCP^O

and RCP. We have shown that they are due to a composite action of the protein.

Our results clearly indicate that RCP allows for a much larger conformational mobility of the β_1 and β_2 dihedral angles, while the binding pocket of OCP^O only allows minor fluctuations in the structure of the bound carotenoid. Our calculations also confirm that the conformation of canthaxanthin in RCP allows for a much larger conjugation than in OCP^O, because the looser structural constraints of the embedding cavity allow for a quasi-planar structure. This fact, already noted from Resonance Raman spectra⁹ and in the crystal structures¹⁰, holds true in the configurations extracted from

our classical and QM/MM MD trajectories. As a result of this geometrical effect, a significant red-shift is observed even without taking the electrostatic effect of the environment into account. Moreover, we verified that the differences in the lineshape can be fully explained by a combination of vibronic coupling and inhomogeneous disorder, due to the presence of different conformations of CAN in the two systems.

In addition to the different conformational constraints induced by the two embedding cavities, the atomistic nature of our model allowed us to clarify other important roles of the protein. From the analysis of the electrostatic and polarization effects of the various residues, it emerged that in the OCP^O cavity the carotenoid is in an unfavorable configuration for both electrostatic and steric reasons; however, this is energetically counterbalanced by the presence of the two strong hydrogen bonds. On the contrary, the different composition of the RCP cavity allows for a more stable configuration of CAN, due to the favorable interactions with many different (charged) residues and a nearly planar conformation.

Combining all these findings we can finally propose a new outline for the mechanism of photoactivation and back-conversion of OCP. We can infer that the orange form of the complex is the thermodynamically most stable one, mainly due to the presence of the hydrogen bonds. As many experiments suggest^{15,30,41}, upon photoexcitation these two strong interactions are lost (or largely weakened) and, according to our results, the other residues of the cavity can only provide a small stabilization for the carotenoid. The system should therefore either evolve through the formation of the red form, that offers a much more favorable binding pocket, or reform the two hydrogen bonds. The low quantum yield ($\sim 1\%$) observed for the process suggests that the latter case is by far the most frequent event. The electrostatic differences between RCP and OCP^O here revealed play a major role in tuning the relative stability of OCP^O and OCP^R that is crucial for the reversible conversion of the protein. A simple explanation that rationalizes both experimental

findings and the herein presented data is that OCP^O with hydrogen bonds formed is more stable than OCP^R (as the ability to back-convert in the dark suggests). Despite the dynamics of the photo-conversion process is still to be clarified, it is now evident that when the chromophore undergoes an electronic excitation, the resulting change in its electronic structure could be sufficient, even if with the observed low probability, to initiate a relaxation towards a stable structural and positional configuration finally leading to the translocation of the CAN into the NTD domain and the observed opening of the protein.

Materials and Methods

System Preparation and Molecular Dynamics Initial structures of OCP^O and RCP were taken from the PDB entries 4XB5 and 4XB4, respectively.¹⁰ For RCP, one of the two monomers present in the dimeric structure was used. For each alternative location of sidechains, the most populated was kept. Except for the CAN molecule, all other cofactors were deleted. Protonation was assigned using the H++ software⁴². The systems were inserted in a 30 Å wide TIP3P truncated octahedron water box with a NaCl concentration of 0.15 M using AmberTools 18⁴³. Classical MD was run using the ff14SB AMBER forcefield⁴⁴ for the protein and a DFT-fitted forcefield for the CAN (see Supporting Information section S2.1 for details). Plain MD runs were started with a minimization, a 50 ps long NVT heating up to 50 K and a second heating in NPT ensemble up to 300 K (250 ps) followed by a 750 ps long equilibration in NPT. Then the production (about 1 μ s for OCP^O and 2 μ s for RCP) was run in the NVT ensemble. Geometrical analyses were run using the cpptraj program from AmberTools 2018⁴³ and custom-made python scripts.

Well-tempered metadynamics simulations were run from equilibrated snapshots of the plain MD, following the same heating and equilibration process as the one described before. GROMACS v. 2018.4⁴⁵ patched with

PLUMED 2.4^{46,47} was used. The conversion between the AMBER and GROMACS file formats was handled with InterMol⁴⁸. Details of the bias applied are supplied in the Supporting Information section S2.3.

Polarizable QM/MM Born-Oppenheimer MD simulations were run starting from frames of the plain MD trajectory. Only the protein and the water molecules within 2.0 Å from the protein were kept and the simulation was run in the microcanonical ensemble, after a minimization and a 0.5 ps heating procedure (details in Supporting Information section S2.4). The dynamics was propagated with a modified version of Tinker^{49,50}. For describing the MM part of the system, we used the polarizable AMOEBA FF⁵¹ and the QM/AMOEBA forces were calculated using a modified version of Gaussian 16 A.03⁵² that implements the Born-Oppenheimer MD as reported by Loco *et al.*^{24,53} The QM part of the system was treated at the B3LYP/6-31G(d) level.

Electronic Structure Calculations Excited state energies were calculated on structures extracted from plain MD simulations after optimization in the ONIOM scheme⁵⁴ with electrostatic embedding. The protein and all water molecules within 10 Å from the protein were retained in the MM part and kept frozen, whereas the QM part comprising the carotenoid alone was treated at the DFT B3LYP/6-31G(d) level. The forcefield for the MM part was the same as for the MD simulations. Excited state energies were calculated on the optimized structures at the TD-DFT CAM-B3LYP/6-31+G(d) level treating the environment (protein, water molecules, and ions) at the AMOEBA level.²³ All calculations were performed with a modified version of Gaussian 16 A.03⁵².

Spectra Simulations Excitation energies were calculated at the same level as above directly on snapshots of the Born Oppenheimer MD extracted every 4 fs. Tests with different functionals are described in the Supporting Information section S6.3. Autocorrelation functions and their Fourier transforms were calculated using custom-made python scripts

based on SciPy and NumPy libraries.⁵⁵ Vibronic absorption lineshapes were computed in the second order cumulant expansion approach starting from the spectral densities (details in Supporting Information section S6.2).

Supporting Information Available

Analysis of dihedral angle conformation of canthaxanthin, zeaxanthin, β -carotene and violaxanthin and comparison of absorption spectra; analysis of puckering/dihedral angle correlation for canthaxanthin; details on simulation parameters used in classical MD; details on CAN forcefield reparametrization; analysis of a 1 μ s long classical MD simulation of CAN in vacuo; details on parameters used in wt-MetaMD simulations and convergence analysis; details on QM/AMOEBA Born-Oppenheimer MD simulations; analysis on plain classical MD trajectories; clustering of CAN conformations; details on spectral density calculations; normal mode analysis of HOOP and rock modes of CAN in RCP and OCP; analysis of correlation of solvatochromic shifts and ground state dipoles.

Acknowledgement The authors acknowledge funding by the European Research Council, under the grant ERC-AdG-786714 (LIFE-TimeS). FL gratefully acknowledges the support of NVIDIA Corporation with the donation of the Titan Xp used for this research.

Author Contributions

Conceptualization B.M.; Data curation M.B. and L.C.; Formal analysis B.M., M.B. and L.C.; Funding acquisition B.M. and F.L.; Project administration B.M.; Software F.L. L.C. and M.B.; Supervision B.M.; Visualization M.B. and L.C.; Writing – original draft M.B. and L.C.; Writing – review & editing B.M., F.L., L.C. and M.B.

References

- (1) Müller, P.; Li, X.-P.; Niyogi, K. K. *Plant Physiol.* **2001**, *125*, 1558–1566.
- (2) Fleming, G. R.; Schlau-Cohen, G. S.; Amarnath, K.; Zaks, J. *Faraday Discuss.* **2012**, *155*, 27–41.
- (3) Niyogi, K. K.; Truong, T. B. *Curr. Opin. Plant Biol.* **2013**, *16*, 307–314.
- (4) Croce, R.; van Amerongen, H. *Nat. Chem. Biol.* **2014**, *10*, 492–501.
- (5) Pinnola, A.; Bassi, R. *Biochim. Soc. Trans.* **2018**, *46*, 467–482.
- (6) Ruban, A. V. *FEBS Lett.* **2018**, *592*, 3030–3039.
- (7) Kirilovsky, D.; Kerfeld, C. A. *Nat. Plants* **2016**, *2*, 16180.
- (8) Kerfeld, C. A.; Melnicki, M. R.; Sutter, M.; Dominguez-Martin, M. A. *New Phytol.* **2017**, *215*, 937–951.
- (9) Wilson, A.; Punginelli, C.; Gall, A.; Bonetti, C.; Alexandre, M.; Routaboul, J.-M.; Kerfeld, C. A.; van Grondelle, R.; Robert, B.; Kennis, J. T. M.; Kirilovsky, D. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 12075–12080.
- (10) Leverenz, R. L.; Sutter, M.; Wilson, A.; Gupta, S.; Thurotte, A.; Bourcier de Carbon, C.; Petzold, C. J.; Ralston, C.; Perreau, F.; Kirilovsky, D.; Kerfeld, C. A. *Science* **2015**, *348*, 1463–1466.
- (11) Gupta, S.; Guttman, M.; Leverenz, R. L.; Zhumadilova, K.; Pawlowski, E. G.; Petzold, C. J.; Lee, K. K.; Ralston, C. Y.; Kerfeld, C. A. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5567–E5574.
- (12) Liu, H.; Zhang, H.; Orf, G. S.; Lu, Y.; Jiang, J.; King, J. D.; Wolf, N. R.; Gross, M. L.; Blankenship, R. E. *Biochemistry* **2016**, *55*, 1003–1009.
- (13) Bandara, S.; Ren, Z.; Lu, L.; Zeng, X.; Shin, H.; Zhao, K.-H.; Yang, X. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 6286–6291.
- (14) Bao, H.; Melnicki, M. R.; Kerfeld, C. A. *Curr. Opin. Plant Biol.* **2017**, *37*, 1–9.
- (15) Maksimov, E. G. *Biophys. J.* **2015**, 13.
- (16) Gurchiek, J. K.; Bao, H.; Domínguez-Martín, M. A.; McGovern, S. E.; Marquardt, C. E.; Roscioli, J. D.; Ghosh, S.; Kerfeld, C. A.; Beck, W. F. *J. Phys. Chem. B* **2018**, *122*, 1792–1800.
- (17) Konold, P. E.; van Stokkum, I. H. M.; Muzzopappa, F.; Wilson, A.; Groot, M.-L.; Kirilovsky, D.; Kennis, J. T. M. *J. Am. Chem. Soc.* **2018**, *141*, 520–530.
- (18) Mezzetti, A.; Alexandre, M.; Thurotte, A.; Wilson, A.; Gwizdala, M.; Kirilovsky, D. *J. Phys. Chem. B* **2019**, *123*, 3259–3266.
- (19) Kish, E.; Pinto, M. M. M.; Kirilovsky, D.; Spezia, R.; Robert, B. *Biochim. Biophys. Acta - Bioenergetics* **2015**, *1847*, 1044–1054.
- (20) López-Igual, R.; Wilson, A.; Leverenz, R. L.; Melnicki, M. R.; Bourcier de Carbon, C.; Sutter, M.; Turmo, A.; Perreau, F.; Kerfeld, C. A.; Kirilovsky, D. *Plant Physiol.* **2016**, *171*, 1852–1866.
- (21) Wilson, A.; Kinney, J. N.; Zwart, P. H.; Punginelli, C.; D’Haene, S.; Perreau, F.; Klein, M. G.; Kirilovsky, D.; Kerfeld, C. A. *J. Biol. Chem.* **2010**, *285*, 18364–18375.
- (22) Kerfeld, C. A.; Sawaya, M. R.; Brahmamandam, V.; Cascio, D.; Ho, K. K.; Trevithick-Sutton, C. C.; Krogmann, D. W.; Yeates, T. O. *Structure* **2003**, *11*, 55–65.
- (23) Loco, D.; Polack, E.; Caprasecca, S.; Lagardère, L.; Lipparini, F.; Piquemal, J.-P.; Mennucci, B. *J. Chem. Theory Comput.* **2016**, *12*, 3654–3661.

- (24) Loco, D.; Lagardère, L.; Caprasecca, S.; Lipparini, F.; Mennucci, B.; Piquemal, J.-P. *J. Chem. Theory Comput.* **2017**, *13*, 4025–4033.
- (25) Uragami, C.; Saito, K.; Yoshizawa, M.; Molnár, P.; Hashimoto, H. *Arch. Biochem. Biophys.* **2018**, *650*, 49–58.
- (26) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.
- (27) Prandi, I. G.; Viani, L.; Andreussi, O.; Mennucci, B. *J. Comput. Chem.* **2016**, *37*, 981–991.
- (28) Rodriguez, A.; Laio, A. *Science* **2014**, *344*, 1492–1496.
- (29) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*.
- (30) Konold, P. E.; van Stokkum, I. H. M.; Muzzopappa, F.; Wilson, A.; Groot, M.-L.; Kirilovsky, D.; Kennis, J. T. M. *J. Am. Chem. Soc.* **2019**, *141*, 520–530.
- (31) Andersson, P. O.; Gillbro, T.; Ferguson, L.; Cogdell, R. J. *Photochem. Photobiol.* **1991**, *54*, 353–360.
- (32) Renge, I.; Sild, E. *J. Photochem. Photobiol., A* **2011**, *218*, 156–161.
- (33) Mukamel, S. *Principles of Nonlinear Optical Spectroscopy*; Oxford Series in Optical and Imaging Sciences 6; Oxford Univ. Press: New York, 2009; OCLC: 549165695.
- (34) Page, J. B.; Tonks, D. L. *J. Chem. Phys.* **1981**, *75*, 5694–5708.
- (35) Myers, A. B. *Acc. Chem. Res.* **1997**, *30*, 519–527.
- (36) Scholz, R.; Gisslén, L.; Schuster, B.-E.; Casu, M. B.; Chassé, T.; Heinemeyer, U.; Schreiber, F. *J. Chem. Phys.* **2011**, *134*, 014504.
- (37) Llansola-Portoles, M. J.; Pascal, A. A.; Robert, B. *J. R. Soc. Interface* **2017**, *14*, 20170504.
- (38) Fujisawa, T.; Leverenz, R. L.; Nagamine, M.; Kerfeld, C. A.; Unno, M. *J. Am. Chem. Soc.* **2017**, *139*, 10456–10460.
- (39) Kish, E.; Pinto, M. M. M.; Kirilovsky, D.; Spezia, R.; Robert, B. *Biochim. Biophys. Acta - Bioenergetics* **2015**, *1847*, 1044–1054.
- (40) Šlouf, V.; Kuznetsova, V.; Fuciman, M.; de Carbon, C. B.; Wilson, A.; Kirilovsky, D.; Polívka, T. *Photosynth. Res.* **2017**, *131*, 105–117.
- (41) Maksimov, E. G.; Sluchanko, N. N.; Slonimskiy, Y. B.; Slutskaya, E. A.; Stepanov, A. V.; Argentova-Stevens, A. M.; Shirshin, E. A.; Tsoraev, G. V.; Klementiev, K. E.; Slatinskaya, O. V.; Lukashev, E. P.; Friedrich, T.; Paschenko, V. Z.; Rubin, A. B. *Sci. Rep.* **2017**, *7*, 1–12.
- (42) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. *Nucleic Acids Res.* **2012**, *40*, W537–W541.
- (43) Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; T.E. Cheatham, I. a. C.; Darden, T.; R.E. Duke, D. G.; Gilson, M.; Gohlke, H.; Goetz, A.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; F. Pan, R. Q.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; J. Smith, R. S.-F.; Swails, J.; Walker, R.; Wang, J.; Wei, H.; Wolf, R.; Wu, X.; L. Xiao, D. Y.; Kollman, P. AMBER 2018. 2018; University of California, San Francisco.
- (44) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

- (45) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, *1-2*, 19–25.
- (46) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D.; Capelli, R.; Carloni, P.; Ceriotti, M.; Cesari, A.; Chen, H.; Chen, W.; Colizzi, F.; De, S.; De La Pierre, M.; Donadio, D.; Drobot, V.; Ensing, B.; Ferguson, A. L.; Filizola, M.; Fraser, J. S.; Fu, H.; Gasparotto, P.; Gervasio, F. L.; Giberti, F.; Gil-Ley, A.; Giorgino, T.; Heller, G. T.; Hocky, G. M.; Iannuzzi, M.; Invernizzi, M.; Jelfs, K. E.; Jussupow, A.; Kirilin, E.; Laio, A.; Limongelli, V.; Lindorff-Larsen, K.; Löhr, T.; Marinelli, F.; Martin-Samos, L.; Masetti, M.; Meyer, R.; Michaelides, A.; Molteni, C.; Morishita, T.; Nava, M.; Paissoni, C.; Papaleo, E.; Parrinello, M.; Pfaendtner, J.; Piaggi, P.; Piccini, G.; Pietropaolo, A.; Pietrucci, F.; Pipolo, S.; Provasi, D.; Quigley, D.; Raiteri, P.; Raniolo, S.; Rydzewski, J.; Salvalaglio, M.; Sosso, G. C.; Spiwok, V.; Šponer, J.; Swenson, D. W. H.; Tiwary, P.; Valsson, O.; Vendruscolo, M.; Voth, G. A.; White, A.; The PLUMED consortium, *Nat. Meth.* **2019**, *16*, 670–673.
- (47) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput. Phys. Comm.* **2014**, *185*, 604–613.
- (48) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. *J. Comput. Aided Mol. Des.* **2016**, *31*, 147–161.
- (49) Ponder, J. W. TINKER, Software Tools for Molecular Design. <http://dasher.wustl.edu/tinker>.
- (50) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289.
- (51) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- (52) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 16 Revision A.03. 2016; Gaussian Inc. Wallingford CT.
- (53) Lipparini, F. *J. Chem. Theory Comput.* **2019**, *15*, 4312–4317.
- (54) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. *Journal of Computational Chemistry* **2003**, *24*, 760–769.
- (55) Jones, E.; Oliphant, T.; Peterson, P. SciPy: Open source scientific tools for Python. 2001–; <http://www.scipy.org/>.