Excited States of Xanthophylls Revisited: Towards the Simulation of Biologically Relevant Systems

Mattia Bondanza,^{*,†} Denis Jacquemin,[‡] and Benedetta Mennucci^{*,†}

[†]*Dipartimento di Chimica e Chimica Industriale, University of Pisa, via G. Moruzzi 13,
56124, Pisa, Italy*

[‡]*Université de Nantes, CNRS, CEISAM UMR 6230, F-44000 Nantes, France*

E-mail: mattia.bondanza@phd.unipi.it; benedetta.mennucci@unipi.it

Abstract

Xanthophylls are an important class of oxygen containing carotenoids, which play a fundamental role both in light harvesting pigment-protein complexes and in many other photo-responsive proteins. The complexity of the manifold of the electronic states and the large sensitivity to the environment still prevent a clear and coherent interpretation of their photophysics and photochemistry. In this Letter, we first apply cutting-edge *ab-initio* methods (CC3, DMRG/NEVPT2) on model C_{2h} keto-carotenoids to critically assess the performances of single and multi-reference methods. We then lift symmetry restrictions and extend the tested methods to time dependent DFT and Semiempirical CI (SECI). From this analysis, we demonstrate that SECI can indeed represent an optimal method for describing light-induced processes of xanthophylls in biologically relevant systems. As an example, we investigate canthaxanthin in the Orange Carotenoid Protein showing that the conical intersections between excited states and excited-ground states are mostly determined by the effective bond length alternation coordinate, which is significantly tuned by the protein through geometrical constraints and electrostatic effects.

Carotenoids are ubiquitous compounds in living organisms where they contribute to the functions of different pigment-protein complexes.¹⁻³ This large diffusion and the multiple roles they play stem from their unique characteristics of being both “extensible” and “flexible” structures, with electronic and optical signatures which are strongly interlinked with their length and conformation. These characteristics render the properties of carotenoids highly-tunable depending on the specific embedding environment.

Many of the biological functions involving carotenoids exploit the specific photophysics brought by their low-lying $\pi \rightarrow \pi^*$ excited state manifold. Illustratively, in an idealized perfectly symmetric C_{2h} polyene, the two lowest-lying excited states that are typically investigated are the optically dark A_g and bright B_u . However, another B_u state is present amongst the low-lying $\pi \rightarrow \pi^*$ states but with a dark character. To distinguish between the two, the notation B_u^+ (bright) and B_u^- (dark) derived from particle-hole symmetry⁴ is often used. Moreover, in xanthophylls, e.g. carotenoids containing carbonyl or hydroxy groups, additional states of $n \rightarrow \pi^*$ character are present. In C_{2h} symmetric cases, two (low-lying) $n \rightarrow \pi^*$ states are present and they belong to the A_u and B_g representations.

The strong absorption of carotenoids arises from a transition from the ground to the B_u^+ state, which, by internal conversion, rapidly de-

cays (in less than 200 fs) to A_g state. The latter next decays to the ground state by internal conversion in several picoseconds and fluorescence consequently occurs with an extremely low quantum yield only. The presence of another “dark S^* ” state in the vicinity of A_g and B_u^+ has been also suggested.^{5,6} However, the nature and energetic position of this S^* state remains unclear, and there are debates regarding its nature: is it a real, separate electronic excited state, or, alternatively a twisted A_g state or even a vibrationally hot ground state.^{7,8} Finally, in xanthophylls, a further complexity has been suggested, namely that the excited state relaxation may proceed through the coupling with an intramolecular charge transfer (ICT) state depending on the solvent polarity.⁹

From this brief summary, it is clear that, despite the large interest in the photophysics of carotenoids, a clear and established theoretical characterization of their excited state manifold is still missing. The main reason is that applying accurate *ab-initio* QM methods to the study of carotenoids remains challenging. Indeed, the delicate coupling between structural and electronic degrees of freedom, combined with the very different nature of the key excited states render the complete characterization of these systems feasible for “short” carotenoids only, which are not those of real biological interest. In particular, the exact (i.e., FCI) positions of the A_g and B_u states has only been resolved for small compounds, namely butadiene and

hexatriene.^{10,11} We note, however, that a very recent paper by Khokhlov *et al.*¹² has shown that a balanced picture of the low-lying excited states of large polyenes can be achieved using Driven Similarity Renormalization Group-Multi Reference Perturbation Theory (DSRG-MRPT2).^{13,14} In any case, the realistic simulation of the photophysics of carotenoids within their biological environment requires the inclusion of dynamic effects thus making the application of highly-accurate *ab-initio* methods even more difficult if not impossible.

The most obvious computational choice to explore excited states is Time-Dependent Density Functional Theory (TD-DFT) which is the gold standard in terms of cost-effectiveness. TD-DFT, however, has the strong limit of being a single-reference method and, as such, it cannot properly describe states with a significant multireference character and/or states having a significant double excitation character. This is indeed the case for the A_g state in which the doubly-excited configurations play a significant role, meaning that conventional TD-DFT overshoots its transition energy, in contrast to the B_u^+ state that is reasonably well captured by TD-DFT. A promising strategy to go beyond this single-reference description is to select semiempirical (SE) Hamiltonians, and combine them with configuration interaction descriptions (SECI).^{15–17} In this case, however, another limit applies, namely the intrinsic accuracy of the SE Hamiltonian.

While many previous theoretical studies have investigated polyenes of increasing length,^{12,18–27} much less has appeared so far for their analogs containing carbonyl (or hydroxy) groups.^{28–31} Hence, the first part of the present study is an effort to fill this gap by establishing reliable reference *ab-initio* data for a set of model keto-carotenoids of C_{2h} symmetry and $C_{2n-2}H_{4n-2}O_2$ stoichiometry, where $n=2-13$ (from now on, CO_n). These reference data were used to assess the accuracy and reliability of TD-DFT and SECI approaches. In the second part of the study we have selected a representative example of the CO_n systems, and analyzed the changes in nature and relative energy for its manifold of states due to structural

deformations breaking the C_{2h} symmetry. Once again, we have compared *ab-initio* methods with TD-DFT and SECI approaches. Finally, in the third part of the study, we have applied the SECI approach tested and established in the first two steps, to investigate the Orange Carotenoid Protein (OCP), a protein involved in cyanobacteria photoprotective mechanisms. Upon absorption of blue light, the xanthophyll embedded within the two domains of the protein (Canthaxantin, CAN, see Scheme 1), translocates inside one domain and the protein opens up, thus allowing the interaction with the cyanobacterial antenna, the phycobilisome.^{32,33} In the last few years, many aspects of the opening process have been clarified,^{34–36} but the molecular details of the initial photochemical process, which takes place within few ps after the excitation of CAN, remain elusive.

Since at present almost no routinely available methods are able to accurately and simultaneously describe static and dynamic correlation on systems of moderate size, we built our first part of the analysis on the comparison of a high quality single reference method and a multi-reference approach. For the former we selected the third-order Coupled-Cluster (CC3) and its Equation of Motion (EOM) extension to excited states,^{37,38} (from now on we simply use CC3 for both ground and excited, EOM-CC3, calculations). As a multi-reference approach we selected CAS-SCF and we constructed the “natural active space” of each CO_n characterized by n double bonds, in terms of n occupied and n virtual π orbitals, and the two occupied non-bonding lone-pair orbitals. Since the CAS-SCF method is strongly limited by the exponential scaling in the size of the active space, we decided to overcome this limitation by exploiting the advantages offered by Density Matrix Renormalization Group (DMRG).^{39,40} This method is able to provide approximate results for larger active spaces and it was shown to be particularly well-suited for describing linear polyenes and carotenoids.^{27,31,41} To account also for the effects of dynamic correlation, we have applied a perturbative correction (NEVPT2) to DMRG-CASSCF energies.⁴² We note that, DMRG/NEVPT2 has been success-

fully used to capture the effects of nuclear relaxation on the manifold of excited states for five extended carotenoids.³¹

Scheme 1: Top: chemical structure of model keto-carotenoids (CO_n). Bottom: structure of Canthaxanthin with the CO_{13} core highlighted in orange.

Our analysis starts in a classical Frank-Condon (FC) framework by studying the excitation energies of low lying electronic states of CO_n systems at their ground-state (GS) equilibrium geometries. However, obtaining reliable GS geometries for large π -conjugated systems is not trivial since a treatment of electronic correlation is needed to recover the correct degree of delocalization along the conjugated chain. Specifically, it is well known that the Bond Length Alternation (BLA), one of the most important geometrical parameters quantifying the degree of localization of the obtained electronic structure, is significantly dependent on the level of theory used.^{22,43} Due to its diagnostic importance in these systems we choose the BLA as our main reference parameter in selecting a GS optimization method. Since the GS of polyenes is a single reference state, we have taken as gold standard CC3, computing optimized geometries for systems with 2 to 6 double bonds (CO_2 – CO_6), and comparing them to methods that allow treatment of larger systems, namely MP2, DFT/B3LYP,^{44,45} and DFT/CAM-B3LYP⁴⁶).

To test the quality of the MP2 structures, we compared them with the CC3 ones for the shorter CO_n chains. The results, reported in Figure S1, show that the excitation energies calculated with the two sets of structure are extremely similar. The results reported in Figure S2, clearly show that CAM-B3LYP gives a too overlocalized picture of the conjugated system

as compared to CC3, while B3LYP exaggerates the system conjugation. For longer systems instead, B3LYP behaves similarly to MP2; such behaviours are consistent with previous studies on other conjugated systems.^{22,43} On the basis of this analysis, we selected MP2 as the method of choice for the optimization of the ground state geometries of CO_n compounds with up to 13 double bonds.

To characterize the excited states, we calculated the electron density changes upon electronic excitation for each different state of the CO_{13} system at DMRG-CASSCF level. The results are shown in Figure 1. It is clear that the B_u^+ , B_u^- , and A_g states share a similar density displacement in which π density is reduced on the double bonds and increased on the single ones. We also note, from the apparently different volumes embraced by the iso-density surfaces, that the density reorganization is larger for the dark states (B_u^- and A_g) than for the bright one (B_u^+), consistent with the better description of the latter with single-reference methods. In contrast, the density differences of both the A_u and B_g states confirm their $n \rightarrow \pi^*$ nature, in which the density is displaced from the terminal regions corresponding to lone pairs of the oxygen atoms to the π^* orbitals vicinal to the carbonyl group. The fact that $n \rightarrow \pi^*$ transitions, which are nearly degenerate in longer systems such as CO_{13} , involve both carbonyl groups is a consequence of the imposed symmetry, and they can be more conveniently represented as localized transitions on each of the two carbonyl groups.

More detailed analyses of the DMRG wavefunctions were performed by using the Mutual Information Diagrams^{47,48} (see Figure S3) and by estimating the contribution of single Slater determinants in the wavefunction⁴⁹ (see Table S3). The latter scheme allows for a more straightforward interpretation: one can easily recognize that the GS has a largely dominant HF character, while the two $n \rightarrow \pi^*$ transitions are a combination of single excitations from one non-bonding orbital to π^* orbitals. In contrast, the dark $\pi \rightarrow \pi^*$ transitions (A_g and B_u^-) show large contributions of both double and single excited determinants while the bright

Figure 1: DMRG-CASSCF density difference between excited and ground state of CO_{13} represented as $\Delta\rho$ ($\Delta\rho = \rho_n - \rho_{\text{GS}}$) isosurfaces at +0.001 (red) and -0.001 (blue).

B_u^+ state has a predominant HOMO $\rightarrow$ LUMO character without any significant contribution of double excited determinants. This confirms what expected, namely that the A_g and B_u^- states are the ones with the most important multi-reference character.

Let us proceed with the analysis of the transition energies by comparing CC3 with DMRG-CASSCF (Figure 2A) and DMRG-CASSCF/NEVPT2 (Figure 2B) results. Since the computational demand of CC3 is much larger than the one of DMRG, we limited the calculations performed with the former method to CO_8 , which is already a computational challenge.

The comparison between CC3 and DMRG-CASSCF highlights advantages and drawbacks of these two methods, which were only partially predictable. In particular, we note a good agreement between the two for the dark B_u^- state, but also for the dark A_g state, both being characterized by significant contributions from double excitations. At this stage, we recall that CC3 was shown to slightly overestimate the excitation energies of the dark A_g state of butadiene and hexatriene, but to be rather competitive with multi-reference methods,¹¹ for these short systems in which the contributions of the doubles in the A_g state remain limited to ca. 25%. As we expect the double excitation character to increase with chain length, the CC3 error should become larger as longer systems are considered. The $n \rightarrow \pi^*$ transitions are some-

how blue-shifted in DMRG-CASSCF with respect to CC3 but it is the bright B_u^+ state which, quite surprisingly, shows the most significant difference. As the B_u^+ state is mainly characterized by single excitations, the corresponding excitation energy is expected to be accurately described by CC3: this implies that it is largely overestimated by DMRG-CASSCF, likely due to the lack of dynamic correlation.

In order to improve these results we applied NEVPT2 on top of the DMRG-CASSCF. Looking at the NEVPT2-corrected results we note only minor changes for both dark $\pi \rightarrow \pi^*$ states. In contrast, the transition energies are now in very good agreement with the CC3 ones for the two $n \rightarrow \pi^*$ transitions. Two anomalies appear in the data: both the $\pi \rightarrow \pi^*$ states of CO_4 and the $n \rightarrow \pi^*$ of CO_8 - CO_{12} correspond to a quasi degenerate situation (of B_u^+/B_u^- and $B_u^-/A_u/B_g$ respectively) in the mutireference wavefunction (see Figure 2A). In these cases, NEVPT2 is expected to be unreliable. However, when these near-degeneracy effects become smaller (CO_{13}), a smooth trend is recovered. The largest effect of the NEVPT2 correction is found for the bright B_u^+ state which is significantly red-shifted with respect to DMRG-CASSCF, resulting in excitation energies which are 0.5–0.7 eV lower than the ones predicted with CC3. We interpret this outcome has an over-correction typical of second-order perturbative models when the starting point is far from the spot.

Combining all these considerations we can conclude that DMRG/NEVPT2 is a good reference for the dark A_g and B_u^- states; the same holding for the $n \rightarrow \pi^*$ states as far as they are well separated (energetically) from other states. In contrast, one has likely to be cautious in considering this method as a reliable reference for the bright B_u^+ state as it is here shown that a significant underestimation of its excitation energy is systematically observed for the treated systems.

Based on this analysis, we have built a dataset of benchmark excitation energies for the CO_n systems by mixing the most accurate results obtained with different methods. In this dataset, we used (i) for A_u and B_g : CC3 for short sys-

Figure 2: Excitation energies of CO_n systems at the DMRG-CASSCF level without (left panel) and with (right panel) NEVPT2 correction. CC3 data (stars) are also reported for the shorter systems. The color code indicates the electronic state: A_g in blue, B_u^+ in orange, B_u^- in red, B_g in green, and A_u in violet. All calculations have been performed within a C_{2h} symmetry.

tems and NEVPT2 for longer ones; (ii) for A_g : NEVPT2; (iii) for the bright B_u^+ : CC3 for shorter systems and their extrapolation for longer systems. Details on the extrapolation procedure and numerical data of extrapolated/interpolated excitation energies (Table S2) are reported in the SI.

This data set can now be used to assess the performances of both TD-DFT (using B3LYP and CAM-B3LYP functionals) and SECI methods. The results for the FC excitation energies of the CO_n systems are reported in Figure 3. We note that SECI results are reported for CO_n systems with $n > 4$, due to the low accuracy of SECI to describe the shortest systems.

Both TD-DFT and SECI give qualitatively good results, but quantitatively they significantly differ. Focusing on TD-DFT, one notes that both the tested functionals give the same ordering of the states with the $n \rightarrow \pi^*$ transitions being the lowest ones for the short systems ($n=4-6(7)$) and the bright B_u^+ state becoming the lowest for longer chains. However, the two functionals behave quite differently when quantitatively comparing with the benchmark data. In more details, CAM-B3LYP gives very accurate results for the $n \rightarrow \pi^*$ energies while

it significantly overestimates the dark $\pi \rightarrow \pi^*$ energies (A_g and B_u^-) and underestimates the transition energy to the bright B_u^+ state. Unexpectedly, B3LYP seems to give a much more faithful picture than CAM-B3LYP even for the A_g state – despite its significant double excitation character – as well as for the other dark B_u^- state. For the other states, generally underestimated energies (by about 0.5 eV) are found with this functional. Comparisons of the transition densities computed at TD-B3LYP and DMRG-CASSCF levels for CO_{13} indeed show that such an agreement on excitation energies correspond to a qualitative agreement on the description of the underlying electronic states (see Figure S5).

The picture offered by SECI is somehow more balanced than the one obtained with TD-DFT. Interestingly, the energies computed with this method are within 0.5 eV from the reference values for all the states of all the systems. It should also be recalled that the SECI calculations on a medium sized system (CO_{11}) are about three order of magnitude cheaper (computationally) than their TD-DFT counterpart which are in turn two order of magnitude cheaper than the DMRG-CASSCF calculations.

Figure 3: Comparison of TD-CAM-B3LYP (left panel), TD-B3LYP (central panel) and SECI (right panel) excitation energies of CO_n systems with the benchmark values obtained from CC3 and DMRG-CASSCF/NEVPT2 (dotted lines). The same color code is used as in Figure 2 (A_g in blue, B_u^+ in orange, B_u^- in red, B_g in green, and A_u in violet).

The next step in our investigation is to study the lifting of the C_{2h} symmetry constrain. To this end, we focused on a representative system (CO_{11}) and selected a few degrees of freedom that, from chemical intuition, should play a role in the relaxation of the excited states after the Frank-Condon excitation: the BLA, and torsions around a single and a double bond in the terminal part of the chain (quantified by the ϕ_2 and θ_2 dihedral angles, respectively – see Figure 4).

For each of these degrees of freedom we performed relaxed MP2 scans, and computed excitation energies with DMRG-CASSCF, SECI, and TD-B3LYP. The results of these calculations are reported in Figure 4. Since we have already noted that the main pitfall of DMRG-CASSCF is to severely overestimate the energy of the bright B_u^+ state, we reported both its original position and the one shifted by the magnitude of the NEVPT2 correction determined at the GS equilibrium geometry.

Let us start by considering the distortion of the ϕ_2 and θ_2 dihedral angles. All the three methods considered (DMRG-CASSCF, SECI, and TD-B3LYP) yield qualitatively similar results. Even though they provide a different ordering of the electronic states, they all predict that the distortion of either ϕ_2 or θ_2 increases the energy of all the electronic states. Moreover, all methods agree that the rotation

around the single bond (ϕ_2) lifts the quasi-degeneracy of the two $n \rightarrow \pi^*$ transitions after ca. 20° , while twisting the double bond (θ_2) leaves the two transitions almost degenerate even at quite large dihedral angles. All these predictions fit chemical intuition: the distortion of a dihedral angle is expected to reduce the “effective conjugation length” causing an up-shift of all the electronic levels in the molecule. Moreover, while the rotation around the single bond is expected to generate two independent π systems, each of them with its own set of orbitals and therefore with different energies for the $n \rightarrow \pi^*$ transition, the rotation of the double bond should induce a deformation of the π orbitals without isolating the two sides of the system, thus leaving the two $n \rightarrow \pi^*$ nearly degenerate.

When considering the dependence on the BLA, a much more complex profile is found, and qualitative differences are noticed between the different methods. At the DMRG-CASSCF level, the $\pi \rightarrow \pi^*$ states calculated at the BLA coordinate in the FC region are far away from their minima. Indeed, if the BLA in the GS minimum is $0.125 \text{ \AA}$, its A_g and B_u^- counterparts are ca. $0.025 \text{ \AA}$ whereas the B_u^+ state is most stable for a BLA of $0.075 \text{ \AA}$. All these states show roughly harmonic surfaces in the BLA space. In contrast, the GS presents a large anharmonicity for BLA smaller than 0.025 with

Figure 4: Potential energy surfaces of CO_{11} along key degrees of freedom computed with different methods (from top to bottom: DMRG, SECI, and TD-DFT/B3LYP). From left to right: BLA, ϕ_2 (distortion of the second C-C single bond, 0 = *s*-trans, 180 = *s*-cis), θ_2 (distortion of the second double bond – C=O being the first one – 0 = trans, 180 = cis). The point corresponding to the fully optimized MP2 geometry is indicated by a vertical dashed line, and colored circles indicate the electronic state according to the same color code used in the previous figures. The state with the largest transition dipole is always in orange; for DMRG, and additional curve has been reported (dotted line), obtained by shifting the energies by the magnitude of the NEVPT2 correction determined at the GS equilibrium geometry.

a plateau behaviour. Unfortunately, we were not able to obtain a fully converged DMRG-CASSCF wavefunction for any structure with BLA lower than $-0.009 \text{ \AA}$ making an *ab-initio* analysis of this part of the PES impossible. Considering now the interplay between the different states, once the energetic position of the B_u^+ is corrected, it seems that upon excitation, this state should relax along the BLA coor-

dinate through a conical intersection with the A_g state. We note that this conversion should be rather fast as, according to our results, it does not involve large motions. Such description is consistent with the experimental observations.⁵⁰ We also note that the $n \rightarrow \pi^*$ transitions are weakly influenced by the BLA and do not significantly couple with the other electronic states.

Once more, SECI is surprisingly good in reproducing all the relevant features of the DMRG-CASSCF PESs. The largest differences are seen for the $n \rightarrow \pi^*$ states for which the energy slightly increases when reducing the BLA. Furthermore, SECI seems to be more numerically stable and data can be obtained even for BLA smaller than $-0.025 \text{ \AA}$. On the other hand, TD-DFT yields a very different picture with respect to DMRG-CASSCF, underlying the difficulties of this method in describing photochemically important features in carotenoids. In particular, it gives a completely harmonic picture of the GS energy which does not fit the trend produced by multi-reference methods. Indeed, this could be seen as an artificial increase of the GS energy at low BLA due to the lack of multi-reference effects in DFT. Noteworthy, a similar trend is obtained with semiempirical HF (see Figure S6A). Moreover, all the excited states computed with TD-DFT show a very similar harmonic behaviour with minima placed roughly at BLA of $0.050 \text{ \AA}$.

To better understand the changes in the nature of the electronic states during the evolution on the BLA coordinate, we relied on the SECI description and compared the contributions of the different determinants in the wavefunction (see Figure S7). From this analysis, it is clear that, at the GS equilibrium geometry, the description provided by SECI is consistent with the one provided by DMRG-CASSCF for all the $\pi \rightarrow \pi^*$ transitions (see Table S3 in SI). However, moving to smaller BLAs, a dramatic change in the character of the states takes place. In particular, we observe that the GS acquires a significant multi-reference character gaining an important contribution from the $\text{HOMO}^2 \rightarrow \text{LUMO}^2$ determinant while simultaneously the A_g state gains an important contribution from the HF determinant suggesting a mixing of those two states in proximity of their avoided crossing ($\text{BLA} = 0.02 \text{ \AA}$ see Figure 4). The character of the B_u^- state, instead, is only slightly influenced by the BLA, excluding the point close to the conical intersection with the B_u^+ state ($\text{BLA} = 0.025 \text{ \AA}$). Finally, the B_u^+ state shows a significant increase in the double excitation contribution at low BLA to-

gether with a loss of its “pure” $\text{HOMO} \rightarrow \text{LUMO}$ character. Since at low BLA values, the GS presents a non-negligible multireference character, it is expected that the quality of MP2 geometries deteriorates moving away from the minimum. Therefore, we have compared SECI energies computed on BLA relaxed scans performed at MP2 and SECI levels (Figure S6B). From such a comparison, it is clear that the two methods show a good agreement in the BLA region between $0.025 \text{ \AA}$ and $0.130 \text{ \AA}$ while, at lower BLA, the geometries obtained with two methods differ significantly, making the results in this region less trustworthy.

As a final step in our analysis, we investigated Canthaxantin (CAN, see Scheme 1) in vacuum and within a biological matrix, namely the photoresponsive Orange Carotenoid Protein (OCP).

As above, we have performed a relaxed scan along the BLA coordinate of CAN at MP2 level, but this time the optimizations have been repeated both in vacuum and in the protein using the ONIOM(QM:MM)⁵¹ model (see details in the SI). The resulting geometries were finally used to compute SECI excitation energies. To disentangle the effects induced on the geometry of CAN by the constraints of the protein binding pocket from the direct electrostatic effects on the excitation energies, the geometries obtained in protein were also used to compute vacuum excitation energies. In Figure 5 we report the energy profiles calculated on the geometries optimized in the protein while the corresponding ones for the vacuum geometries can be found in the SI (Figure S9).

Unsurprisingly, the energy profiles obtained for CAN outside its biological environment closely resemble the already commented ones of CO_{11} , the only key difference being that the GS equilibrium minimum presents a higher BLA. This was expected as CAN presents methyl groups on the two terminal rings which prevent a full planarization of the conjugated backbone. When moving to the geometries optimized within the protein but still neglecting the effect of the protein electric field on the excitation energies, one notices that the GS basin is much larger than the one obtained on vacuum

Figure 5: SECI energies of different electronic states of CAN calculated for different BLA values in OCP (QM/MM) and in vacuum: GS in black, A_g in blue, B_u^+ in orange, B_u^- in red, unidentified states are in gray. In both cases the geometries are those optimized in OCP. Bottom panel: BLA value mapped on the color of CAN which is represented as licorice, together with the two H-bonding residues Y²⁰¹ and W²⁸⁸.

geometries, and larger decreases of the BLA can be reached at a smaller energetic cost. This behavior is most likely a consequence of the constraints imposed by the protein residues in the binding pocket as confirmed by the fact that the most significant changes in the CAN structure along the BLA scan take place where it is hydrogen bonded with a tyrosin (Y²⁰¹) and tryptophan (W²⁸⁸) residue (see Figure 5). Here, when BLA increases, the backbone tends to become more co-planar with the terminal ring. These findings show that, while in isolated symmetric model systems, the BLA can be considered as a degree of freedom independent of other important structural parameters such as dihedral

angles, this assumption does not hold for CAN in OCP. In the latter case, the excited state relaxation likely happens on a space with a higher dimensionality which cannot be easily identified using chemical intuition only.

Moving now to the electrostatic embedding SECI/MM calculations of CAN in protein, we observe less steep profiles with respect to the ones calculated neglecting the electrostatic effects of the protein (vac). As a result, the position of the conical intersection between the A_g and B_u^+ states is shifted to significantly smaller BLAs. Moreover, in the region of very low BLA (below 0.10 Å), the protein electric field strongly affects the behaviours of both the GS

and A_g states, bringing them at least 0.25 eV closer to each other than in vacuum.

All these observations together suggest that the excited state dynamics of CAN in OCP is significantly different than in vacuum as the protein plays an important role at two levels. First, the constraints induced by the protein binding pocket break the symmetry of CAN, thus enabling new relaxation pathways. Secondly, the protein electric field induces a further fine-tuning of the electronic structure of the carotenoid affecting the relaxation pathways. From these results, we propose some insights on the photoactivation mechanism of the OCP complex. First, it is clear that after the vertical excitation to the B_u^+ state, CAN relaxes with a significant reduction of its BLA. As already commented, it is likely that this major structural change is accompanied by changes in dihedral angles and other internal coordinates. Along this relaxation pathway, CAN rapidly reaches the conical intersection between the B_u^+ and A_g states allowing a population transfer. The final step is expected to be the relaxation from A_g to the GS. From the present comparison of the QM/MM and vacuum energy profiles, it appears that the protein strongly reduces the gap between the two states. Even if this effect should facilitate the relaxation to an high energy region of the GS PES (coherently with experimental observations⁵²), unfortunately from our data we cannot determine whether the population of a metastable “hot” GS, which could resemble the S^* state, is possible. The confirmation of such an hypothesis in fact requires to consider the coupling of the BLA with other structural degrees of freedom, a type of analysis that can only be done with non-adiabatic dynamics techniques.

In the present Letter, for the first time we have compared state of the art *ab initio* methods and computationally cheaper ones on the manifold of electronic states of keto-carotenoids of increasing conjugation length, giving a comprehensive view of advantages and drawbacks of each method. From such a comparison we found that SECI is by far the most reasonable choice to study the photochemistry of such systems as it gives a balanced pic-

ture of all the low lying states both in proximity of the FC point and along relevant relaxation coordinates. On the other hand, we showed that DMRG-CASSCF tends to overestimate the excitation energies in particular for the bright B_u^+ state. Moreover, such an overestimation cannot be satisfactorily corrected using DMRG/NEVPT2. We also confirmed that the widely used TD-DFT reasonably behaves near the FC point, but it dramatically fails when the multi-reference character of the states increase (eg. at low BLA values) making it useless to study the photochemistry. We finally applied SECI to investigate cantaxanthin in the orange carotenoid protein showing the strong impact that the protein matrix has on the electronic states through geometry constraints imposed by the binding pocket on the embedded chromophore and by electric field effects.

Combining the results obtained for the isolated keto-carotenoids with those of CAN in OCP, we can say that the excited state dynamics of xanthophylls embedded in biological matrices is expected to be significantly different from what one can predict for isolated molecules. This clearly poses an important challenge for the modeling, however, we have shown that an effective computational strategy exists and we expect that, when applied to xanthophyll containing proteins, it will allow to definitely clarify the many open questions which still characterize important photoactivated biological functions. The only missing piece is the integration of such a strategy with a nonadiabatic dynamics approach: work in this direction is already in progress in our group.

Supporting Information Available

Methodological details on computational methods used; Numerical values of CO_n excitation energies at different levels of theory; Details on the fitting procedure performed on B_u^+ CC3 excitation energies. Numerical values of interpolated and fitted excitation energies of CO_n ; MP2 optimized structures of C_{2h} CO_n ($n = 2 - 13$), Comparison of excitation en-

ergies computed on CC3 and MP2 optimized structures. Comparison of BLA computed with different methods. Decomposition of DMRG-CASSCF wavefunctions of CO₁₃ in Slater determinants. Mutual information diagrams for DMRG-CASSCF wavefunctions of CO₁₃. Converged DMRG-CASSCF active space molecular orbitals of CO₁₃. Transition densities of CO₁₃ computed at DMRG-SCF level and at TD-B3LYP level. Comparison of GS of CO₁₁ at different values of BLA computed at SECI level and at semiempirical HF level. Comparison of SECI excitation energies of CO₁₁ at different values of BLA computed on MP2 and SECI optimized geometries. Decomposition in Slater Determinants of SECI wavefunctions of CO₁₁ at different values of BLA. SECI excitation energies for CAN at different values of BLA.

Acknowledgement M.B. and B.M. acknowledge funding by the European Research Council, under the grant ERC-AdG-786714 (LIFETimeS). M.B. acknowledge the CINECA award under the ISCRA initiative, for the availability of high performance computing resources and support (ESCS1). D.J. is indebted to the CCIPL computational center installed in Nantes for generous allocation of computational time. The authors thank Axel Koslowski and Frank Neese for providing a copy of the MNDO code.

References

- (1) Britton, G. Structure and Properties of Carotenoids in Relation to Function. *FASEB J.* **1995**, *9*, 1551–1558.
- (2) Polívka, T.; Frank, H. A. Molecular Factors Controlling Photosynthetic Light Harvesting by Carotenoids. *Acc. Chem. Res.* **2010**, *43*, 1125–1134.
- (3) Llansola-Portolés, M. J.; Pascal, A. A.; Robert, B. Electronic and Vibrational Properties of Carotenoids: From in vitro to in vivo. *J. R. Soc. Interface* **2017**, *14*, 20170504.
- (4) Pariser, R. Theory of the Electronic Spectra and Structure of the Polyacenes and of Alternant Hydrocarbons. *J. Chem. Phys.* **1956**, *24*, 250–268.
- (5) Cerullo, G.; Polli, D.; Lanzani, G.; De Silvestri, S.; Hashimoto, H.; Cogdell, R. J. Photosynthetic Light Harvesting by Carotenoids: Detection of an Intermediate Excited State. *Science* **2002**, *298*, 2395–2398.
- (6) Ostroumov, E. E.; Mulvaney, R. M.; Cogdell, R. J.; Scholes, G. D. Broadband 2D Electronic Spectroscopy Reveals a Carotenoid Dark State in Purple Bacteria. *Science* **2013**, *340*, 52–56.
- (7) Polívka, T.; Sundström, V. Dark Excited States of Carotenoids: Consensus and Controversy. *Chem. Phys. Lett.* **2009**, *477*, 1–11.
- (8) Balevičius Jr., V.; Abramavicius, D.; Polívka, T.; Galestian Pour, A.; Hauer, J. A Unified Picture of S* in Carotenoids. *J. Phys. Chem. Lett.* **2016**, *7*, 3347–3352.
- (9) Enriquez, M. M.; Fuciman, M.; LaFountain, A. M.; Wagner, N. L.; Birge, R. R.; Frank, H. A. The Intramolecular Charge Transfer State in Carbonyl-Containing Polyenes and Carotenoids. *J. Phys. Chem. B* **2010**, *114*, 12416–12426.
- (10) Chien, A. D.; Holmes, A. A.; Otten, M.; Umrigar, C. J.; Sharma, S.; Zimmerman, P. M. Excited States of Methylene, Polyenes, and Ozone from Heat-Bath Configuration Interaction. *J. Phys. Chem. A* **2018**, *122*, 2714–2722.
- (11) Véril, M.; Scemama, A.; Caffarel, M.; Lipparini, F.; Boggio-Pasqua, M.; Jacquemin, D.; Loos, P.-F. QUESTDB: a Database of Highly-Accurate Excitation Energies for the Electronic Structure Community. *WIREs Comput. Mol. Sci.* **2021**, *11*, e1517.
- (12) Khokhlov, D.; Belov, A. Toward an Accurate Ab Initio Description of Low-Lying

- Singlet Excited States of Polyenes. *Journal of Chemical Theory and Computation* **2021**,
- (13) Evangelista, F. A. A driven similarity renormalization group approach to quantum many-body problems. *The Journal of Chemical Physics* **2014**, *141*, 054109.
 - (14) Li, C.; Evangelista, F. A. Multireference Driven Similarity Renormalization Group: A Second-Order Perturbative Analysis. *Journal of Chemical Theory and Computation* **2015**, *11*, 2097–2108.
 - (15) Silva-Junior, M. R.; Thiel, W. Benchmark of Electronically Excited States for Semiempirical Methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2. *J. Chem. Theory Comput.* **2010**, *6*, 1546–1564.
 - (16) Thiel, W. Semiempirical quantum-chemical methods. *WIREs Comput. Mol. Sci.* **2013**, *4*, 145–157.
 - (17) Dral, P. O.; Wu, X.; Spörkel, L.; Koslowski, A.; Weber, W.; Steiger, R.; Scholten, M.; Thiel, W. Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Theory, Implementation, and Parameters. *J. Chem. Theory Comput.* **2016**, *12*, 1082–1096.
 - (18) Schulten, K.; Karplus, M. On the Origin of a Low-Lying Forbidden Transition in Polyenes and Related Molecules. *Chem. Phys. Lett.* **1972**, *14*, 305.
 - (19) Tavan, P.; Schulten, K. The Low-Lying Electronic Excitations in Long Polyenes: A PPP-MRD-CI Study. *J. Chem. Phys.* **1986**, *85*, 6602–6609.
 - (20) Hsu, C.-P.; Hirata, S.; Head-Gordon, M. Excitation Energies from Time-Dependent Density Functional Theory for Linear Polyene Oligomers: Butadiene to Decapentaene. *J. Phys. Chem. A* **2001**, *105*, 451–458.
 - (21) Starcke, J. H.; Wormit, M.; Schirmer, J.; Dreuw, A. How Much Double Excitation Character do the Lowest Excited States of Linear Polyenes Have? *Chem. Phys.* **2006**, *329*, 39–49.
 - (22) Peach, M. J. G.; Tellgren, E. I.; Salek, P.; Helgaker, T.; Tozer, D. J. Structural and Electronic Properties of Polyacetylene and Polyyne from Hybrid and Coulomb-Attenuated Density Functionals. *J. Phys. Chem. A* **2007**, *111*, 11930–11935.
 - (23) Marian, C. M.; Gilka, N. Performance of the Density Functional Theory/Multireference Configuration Interaction Method on Electronic Excitation of Extended π -Systems. *J. Chem. Theory Comput.* **2008**, *4*, 1501–1515.
 - (24) Schmidt, M.; Tavan, P. Electronic Excitations in Long Polyenes Revisited. *J. Chem. Phys.* **2012**, *136*, 124309–13.
 - (25) Hu, W.; Chan, G. K.-L. Excited-State Geometry Optimization with the Density Matrix Renormalization Group, as Applied to Polyenes. *J. Chem. Theory Comput.* **2015**, *11*, 3000–3009.
 - (26) Taffet, E. J.; Scholes, G. D. The state falls below at carotenoid-relevant conjugation lengths. *Chem. Phys.* **2018**, *515*, 757–767.
 - (27) Khokhlov, D.; Belov, A. Ab Initio Study of Low-Lying Excited States of Carotenoid-Derived Polyenes. *J. Phys. Chem. A* **2020**, *124*, 5790–5803.
 - (28) Götze, J. P.; Thiel, W. TD-DFT and DFT/MRCI study of electronic excitations in Violaxanthin and Zeaxanthin. *Chem. Phys.* **2013**, *415*, 247–255.
 - (29) Andreussi, O.; Knecht, S.; Marian, C. M.; Kongsted, J.; Mennucci, B. Carotenoids and Light-Harvesting: From DFT/MRCI to the Tamm–Dancoff Approximation. *J. Chem. Theory Comput.* **2015**, *11*, 655–666.

- (30) Spezia, R.; Knecht, S.; Mennucci, B. Excited State Characterization of Carbonyl Containing Carotenoids: A Comparison Between Single and Multireference Descriptions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 17156–17166.
- (31) Taffet, E. J.; Lee, B. G.; Toa, Z. S. D.; Pace, N.; Rumbles, G.; Southall, J.; Cogdell, R. J.; Scholes, G. D. Carotenoid Nuclear Reorganization and Interplay of Bright and Dark Excited States. *J. Phys. Chem. B* **2019**, *123*, 8628–8643.
- (32) 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. A 12 Å Carotenoid Translocation in a Photoswitch Associated with Cyanobacterial Photoprotection. *Science* **2015**, *348*, 1463–1466.
- (33) Kirilovsky, D.; Kerfeld, C. A. Cyanobacterial Photoprotection by the Orange Carotenoid Protein. *Nature Plants* **2016**, *2*, 16180.
- (34) Konold, P. E.; van Stokkum, I. H. M.; Muzzopappa, F.; Wilson, A.; Groot, M. L.; Kirilovsky, D.; Kennis, J. T. M. Photoactivation Mechanism, Timing of Protein Secondary Structure Dynamics and Carotenoid Translocation in the Orange Carotenoid Protein. *J. Am. Chem. Soc.* **2019**, *141*, 520–530.
- (35) Maksimov, E. G.; Protasova, E. A.; Tsoraev, G. V.; Yaroshevich, I. A.; Maydykovskiy, A. I.; Shirshin, E. A.; Gostev, T. S.; Jelzow, A.; Moldenhauer, M.; Slonimskiy, Y. B.; Sluchanko, N. N.; Friedrich, T. Probing of carotenoid-tryptophan hydrogen bonding dynamics in the single-tryptophan photoactive Orange Carotenoid Protein. *Sc. Rep.* **2020**, *10*, 1–12.
- (36) Bondanza, M.; Cupellini, L.; Faccioli, P.; Mennucci, B. Molecular Mechanisms of Activation in the Orange Carotenoid Protein Revealed by Molecular Dynamics. *J. Am. Chem. Soc.* **2020**, *142*, 21829–21841.
- (37) Christiansen, O.; Koch, H.; Jørgensen, P. Response Functions in the CC3 Iterative Triple Excitation Model. *J. Chem. Phys.* **1995**, *103*, 7429–7441.
- (38) Koch, H.; Christiansen, O.; Jørgensen, P.; Sanchez de Merás, A. M.; Helgaker, T. The CC3 Model: An Iterative Coupled Cluster Approach Including Connected Triples. *J. Chem. Phys.* **1997**, *106*, 1808–1818.
- (39) White, S. R.; Martin, R. L. Ab initio Quantum Chemistry using the Density Matrix Renormalization Group. *J. Chem. Phys.* **1999**, *110*, 4127–4130.
- (40) Daul, S.; Ciofini, I.; Daul, C.; White, S. R. Full-CI Quantum Chemistry using the Density Matrix Renormalization Group. *Int. J. Quantum Chem.* **2000**, *79*, 331–342.
- (41) Ghosh, D.; Hachmann, J.; Yanai, T.; Chan, G. K.-L. Orbital Optimization in the Density Matrix Renormalization Group, with Applications to Polyenes and β -Carotene. *J. Chem. Phys.* **2008**, *128*, 144117–15.
- (42) Guo, S.; Watson, M. A.; Hu, W.; Sun, Q.; Chan, G. K.-L. N-Electron Valence State Perturbation Theory Based on a Density Matrix Renormalization Group Reference Function, with Applications to the Chromium Dimer and a Trimer Model of Poly(p-Phenylenevinylene). *J. Chem. Theory Comput.* **2016**, *12*, 1583–1591.
- (43) Jacquemin, D.; Adamo, C. Bond Length Alternation of Conjugated Oligomers: Wave Function and DFT Benchmarks. *J. Chem. Theory Comput.* **2010**, *7*, 369–376.
- (44) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–6.

- (45) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (46) Yanai, T.; Tew, D. P.; Handy, N. C. A new Hybrid Exchange–Correlation Functional using the Coulomb-Attenuating Method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- (47) Rissler, J.; Noack, R. M.; White, S. R. Measuring Orbital Interaction using Quantum Information Theory. *Chem. Phys.* **2006**, *323*, 519–531.
- (48) Boguslawski, K.; Tecmer, P.; Örs Leg-eza,; Reiher, M. Entanglement Measures for Single- and Multireference Correlation Effects. *J. Phys. Chem. Lett.* **2012**, *3*, 3129–3135.
- (49) Moritz, G.; Reiher, M. Decomposition of Density Matrix Renormalization Group States Into a Slater Determinant Basis. *J. Chem. Phys.* **2007**, *126*, 244109.
- (50) Polívka, T.; Sundström, V. Ultrafast Dynamics of Carotenoid Excited States-From Solution to Natural and Artificial Systems. *Chem. Rev.* **2004**, *104*, 2021–2072.
- (51) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678–5796.
- (52) Polívka, T.; Kerfeld, C. A.; Pascher, T.; Sundström, V. Spectroscopic Properties of the Carotenoid 3'-Hydroxyechinenone in the Orange Carotenoid Protein from the Cyanobacterium *Arthrospira maxima*†. *Biochemistry* **2005**, *44*, 3994–4003.