

Theoretical $S_1 \leftarrow S_0$ absorption energies of the anionic forms of oxyluciferin by Variational Monte Carlo and Many Body Green's Function Theory

Emanuele Coccia,^{*,†} Daniele Varsano,^{*,†} and Leonardo Guidoni^{*,‡}

[†]*S3 Center, CNR Institute of Nanoscience, Via Campi 213/A, 41125 Modena, Italy*

[‡]*Dipartimento di Scienze Fisiche e Chimiche, Università degli Studi dell'Aquila, via Vetoio, 67100, L'Aquila, Italy*

E-mail: emanuele.coccia@nano.cnr.it; daniele.varsano@nano.cnr.it; leonardo.guidoni@univaq.it

Abstract

The structures of three negatively charged forms (anionic keto-1 and enol-1, dianionic enol-2) of oxyluciferin (OxyLuc), which are the most probable emitters responsible for the firefly bioluminescence, have been fully relaxed at the variational Monte Carlo (VMC) level. Absorption energies of the $S_1 \leftarrow S_0$ vertical transition have been computed using different levels of theory, such as TDDFT, CC2 and many body Green's function Theory (MBGFT). The use of MBGFT, by means of the Bethe-Salpeter (BS) formalism, on VMC structures provides results in excellent agreement with the value (2.26(8) eV) obtained by action spectroscopy experiments for the keto-1 form (2.32 eV). To unravel the role of the quality of the optimized ground state geometry, BS excitation energies have also been computed on CASSCF geometries, inducing a non negligible blue shift (0.08 and 0.07 eV for keto-1 and enol-1 forms, respectively) with

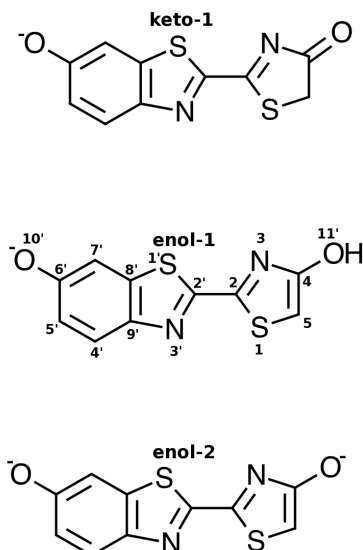
respect to the VMC ones. Structural effects have been analyzed in terms of over- or under-correlation along the conjugated bonds of OxyLuc by using different methods for the ground-state optimization. The relative stability of the S_1 state for the keto-1 and enol-1 forms depends on the method chosen for the excited state calculation, thus representing a fundamental caveat for any theoretical study on these systems. Finally, Kohn-Sham HOMO and LUMO orbitals of enol-2 are (nearly) bound only when the dianion is embedded into a solvent (water and toluene in the present work); excited state calculations are therefore meaningful only in presence of a dielectric medium which localizes the electronic density. The combination of VMC for the ground state geometry and BS formalism for the absorption spectra clearly outperforms standard TDDFT and quantum chemistry approaches.

1 Introduction

Bioluminescence¹⁻⁶ occurs in several living organisms, like fish, insects, algae and bacteria. Animals use bioluminescence for a variety of purposes, such as communication, camouflage, and self-defense. Light emission is the product of a reaction catalyzed by the luciferase in presence of ATP, Mg^{2+} and O_2 , in which D-luciferin is transformed into oxyluciferin (OxyLuc, Figure 1).⁷ The light emitter of the firefly is the OxyLuc in the excited state.

A wide debate on the emitting form of OxyLuc produced a large number of computational and experimental studies on the optical properties of keto and enol tautomers and their respective anions.⁸⁻¹⁴ The structure of OxyLuc is the same for various luciferases,^{15,16} but the emitted light naturally ranges from green (530 nm) to red (635 nm).^{17,18} It has been argued that several factors affect the OxyLuc bioluminescence color: the pH of the environment,^{19,20} according to the keto-enol tautomeric and dissociation equilibria of OxyLuc,⁸⁻¹³ the distortion of OxyLuc in its excited state,²¹ the polarization of the micro-environment,²² the resonance structures of the emitting form^{23,24} and the rigidity of the luciferase binding pocket.²⁵ Important advances in the comprehension of the mechanisms regulating the biolu-

Figure 1: Lewis representation of the keto-1, enol-1, and enol-2 forms of oxyluciferin studied in this work.



minescence mechanism have been achieved in the last ten years also thanks to computational studies, mainly based on multi-scale QM/MM techniques.^{26–31}

Despite the large amount of studies in the literature, there is still no general consensus on the relative stability of the tautomers (also with different protonation states) and how the interaction with the external environment modifies the stability of the various forms of Oxy-Luc. Several experimental^{23,24} and theoretical^{13,20,28,32,33} studies indicate the keto-1 form (Figure 1) as the light emitter. Nevertheless, Naumov *et al.*, reporting the crystal structure of OxyLuc in the enol-1 form (Figure 1), proposed that the enol-1 is the emitter.^{34–36}

In order to understand the photophysics of OxyLuc in complex environments, first a clear understanding of the ground state and absorption properties in the gas phase is needed. In this context, gas phase experiments are very useful to capture the intrinsic electronic properties of OxyLuc without any perturbation given by the surrounding environment and to produce a reference analysis for the gas phase absorption.¹⁴ Action spectroscopy on the bare OxyLuc singly-charged anion produces a quite broad absorption band at 548 ± 10 nm (2.26 ± 0.08 eV).¹⁴ Computational studies¹⁴ carried out so far were only partially in agreement with the experimental findings, showing a large blue shift with respect to the experiments: about 70

nm¹⁴ for EOM-CCSD and 58 and 84 nm¹⁴ for TDDFT/B3LYP and TDDFT/CAM-B3LYP respectively.

Following the indications of Refs^{13,33} we study here the absorption properties of the gas phase keto-1, enol-1 and enol-2 forms of OxyLuc, focusing the attention on the first (bright) excitation, $S_1 \leftarrow S_0$. In this work we aim: i) to study the possible interplay between the optimized ground state geometric parameters and the absorption properties of the OxyLuc forms; ii) to define a computational procedure able to accurately compute the relative shift in the absorption for the three forms; iii) to understand the reliability of gas phase calculations of the dianion enol-2.

When comparing different methods for the evaluation of the excited state properties of a conjugated molecule it is crucial to start from reference ground state structures, since excitation energies can be highly affected by the fine structure of the bond length pattern. Recently, we have shown how Quantum Monte Carlo (QMC)³⁷ methods can provide an accurate ground state geometry for polyenes and conjugated biochromophores.³⁸⁻⁴¹ Following these works we have optimized the ground state geometry of the three forms by means of variational Monte Carlo (VMC) using the Jastrow Antisymmetrized Geminal Power ansatz (JAGP).^{42,43} The combined use of QMC and JAGP has proven to be successful in several applications of physical and (bio)chemical interest, from small molecules (like methylene) to cobalt-based catalysts for the water splitting reaction.^{38-41,44-53} The JAGP ansatz has been recently extended to the calculation of excitation energies⁵⁴⁻⁵⁶ of small molecules and model systems. The VMC/JAGP ground state structures obtained for the keto-1, enol-1, and enol-2 forms have been then compared with those obtained within the DFT framework using different classes of functionals and with CASSCF(18,15) in Ref.¹²

The excited state calculations on keto-1 and enol-1 forms have been carried out using the many body Green's function theory (MBGFT),⁵⁷⁻⁵⁹ CC2 and TDDFT with several functionals. Within MBGFT, the GW approximation and the Bethe-Salpeter (BS) formalism^{60,61} have been recently applied to the study of optical properties of gas-phase^{40,41,62-68} and embed-

ded (in a protein complex environment, for instance see Ref⁴¹) molecular systems, showing results in good agreement with the experimental findings. For neutral excitations, the accuracy is comparable to that of high-level wave function methods,⁶⁹⁻⁷¹ but reliable results are also obtained for charge-transfer excitations.^{72,73} In a recent work Noguchi *et al.*⁷⁴ applied the GW/BS approach to the ionic form of the luciferin in order to describe Rydberg and resonance excitations.

In this work we have also carried out calculations on the dianionic enol-2 form, keeping in mind that, due to the unbound character of the density, ground- and excited-state calculations would not be significant unless solvent effects are considered to stabilize the frontier orbitals.^{33,75,76} For this reason, we have also performed TDDFT calculations for enol-2 and the other forms coupled to a continuum polarizable model (PCM), describing water and toluene solvents.

The present work is organized as follows: in Section 2 we briefly describe the theoretical methodology and discuss the computational details. In section 3 we first report the geometrical properties of the studied forms with different approaches and next we discuss the first optical vertical excitations of the keto-1 and enol-1 forms: our results are systematically compared with the experimental absorption of the bare ion¹⁴ and with previous theoretical studies at TDDFT¹⁴ and MS-CASPT2¹³ levels of theory. Our analysis of the electronic properties of enol-2 is illustrated in a dedicated subsection. Finally in Sec.4 we summarize our findings, underlining the good performance of VMC/BS for the study of the $S_1 \leftarrow S_0$ excitation of the anionic forms of OxyLuc.

2 Theoretical methods and computational details

2.1 Geometry optimization

Ground state geometry optimization of the three forms of OxyLuc has been performed at VMC and DFT level. Semilocal (BLYP), hybrid (B3LYP), hybrid range-separated (CAM-

B3LYP) and meta-GGA (M062X) functionals have been chosen for the DFT optimizations. The DFT structure calculations have been carried out using Gaussian 09⁷⁷ with the diffuse-augmented polarized double- ζ basis set Def2-SVPD.⁷⁸

The VMC structural optimization has been obtained through the minimization of the energy with respect to both the variational parameters $\boldsymbol{\alpha}$ and the nuclear coordinates \mathbf{R} of the trial wave function Ψ_T :

$$E_{VMC}^{OPT} = \min_{\boldsymbol{\alpha}, \mathbf{R}} E[\mathbf{R}; \Psi_T(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{R})]. \quad (1)$$

The energy E at a fixed configuration is evaluated as the expectation value of the electronic Hamiltonian \hat{H} , and the corresponding integral over the electronic coordinates \mathbf{x} (spatial \mathbf{r} and spin $\boldsymbol{\sigma}$) is written introducing the local energy $E_L(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{R}) = \hat{H}\Psi_T(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{R})/\Psi_T(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{R})$, and the probability density $\Pi(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{R}) = \frac{\Psi_T^2(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{R})}{\int \Psi_T^2(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{R}) d\mathbf{x}}$:

$$E[\mathbf{R}; \Psi_T(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{R})] = \int \Pi(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{R}) E_L(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{R}) d\mathbf{x} = \langle \hat{H} \rangle_{\Pi}. \quad (2)$$

The value of the integral in Eq. 2 is estimated as a sum over a finite set of points in the \mathbf{x} space, generated by the Metropolis-Hasting algorithm according to the probability density $\Pi(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{R})$.⁷⁹

We have applied the linear method⁸⁰ with the inclusion of a partial Hessian to accelerate convergence⁸¹ for the optimization of the wave function variational parameters $\boldsymbol{\alpha}$, while the steepest descent algorithm has been used for the structure optimization.

Within the VMC framework, ionic forces for the nucleus a are given by the following expression:⁸²

$$\mathbf{F}_a(\mathbf{R}) = - \left\langle \frac{dE_L}{d\mathbf{R}_a} \right\rangle_{\Pi} + 2 \left\{ \langle E_L \rangle_{\Pi} \left\langle \frac{d \ln [\Psi_T]}{d\mathbf{R}_a} \right\rangle_{\Pi} - \left\langle E_L \frac{d \ln [\Psi_T]}{d\mathbf{R}_a} \right\rangle_{\Pi} \right\}. \quad (3)$$

The variance on the calculated forces is drastically reduced by using the Space Warp Coordinate Transformation (SWCT).^{83,84} Analytical derivatives of the function resulting by the

combined use of pseudopotentials and SWCT, also taking into account the differentiation of the electronic coordinates with respect to the nuclear ones (beyond the expression reported in Eq. 3), have been obtained using the adjoint algorithmic differentiation scheme.⁸⁵

VMC geometry optimization can be considered an accurate reference for the determination of structures of chromophores involved in biological processes,^{39,40} thanks to the explicit presence of the Jastrow factor.⁵⁰ Another important advantage of the method is the extremely favorable parallelism of the algorithms, which is linear with the number of cores, allowing to efficiently exploit high-performing computing facilities and, consequently, to massively apply QMC to the study of medium- and large-size chemical systems.

The trial wave function adopted in this work is the Jastrow Antisymmetrized Geminal Power (JAGP),^{42,43,50} built as the product of an Antisymmetrized Geminal Power (AGP), defining the nodal surface, and a Jastrow factor. The latter is a positive function including many-particle terms for the correct description of the electron-electron and electron-nucleus cusp conditions⁸⁶ in the case of all-electron calculations, and of the dynamical correlation:

$$\Psi_T(\mathbf{x}) = \Psi_{AGP}(\mathbf{x}) \times J(\mathbf{r}). \quad (4)$$

Note that the dependence of Ψ_T from $\boldsymbol{\alpha}$ and \mathbf{R} will be omitted from here on. For closed-shell molecular systems of M atoms and N_e electrons in a spin singlet state (as is the present case), where $N_e/2 = N_e^\uparrow = N_e^\downarrow$, the AGP part is given by the antisymmetrized product:

$$\Psi_{AGP}(\mathbf{x}) = \hat{A} \prod_{i=1}^{N_e/2} \Phi_G(\mathbf{x}_i^\uparrow; \mathbf{x}_i^\downarrow), \quad (5)$$

where Φ_G are geminal functions of two electrons with opposite spin, defined as the linear combination of products of two one-electron atomic orbitals centered on the different nuclei pairs. \hat{A} is the antisymmetrization operator.

The Jastrow term J is split into a product of different terms $J = J_1 J_2 J_3 J_4$,^{44,87,88} accounting for one- (J_1), two- (J_2), three- (J_3) and four-body (J_4) contributions. The J_3 term refers to

the een contribution, while the J_4 one corresponds to $eenm$, with n and m being different nuclei. A detailed discussion on the role played by each contribution and on the functional form of the various Jastrow terms can be found in Refs.⁸⁸ and.⁵⁰ As shown in Eq. 4, the Jastrow factor is spin independent.

The TurboRVB package⁸⁹ has been used for the VMC calculations, following the computational protocol reported elsewhere.^{39,40} The basis sets used for the Ψ_{AGP} and J_3 and J_4 terms of the trial wave function in the VMC geometry optimization are reported in Table 1. All linear and nonlinear parameters belonging to the AGP and the Jastrow terms have been optimized by minimizing the ground-state energy E . Pseudopotentials for C, N, O and S atoms have been employed.^{90,91} The effect of the pseudopotentials on the geometric parameters is negligible when compared with all-electron results, as shown by previous calculations⁴⁴ for the C=C and C-H bonds of the ethylene molecule.

Table 1: Ψ_{AGP} , J_3 and J_4 basis sets used for the VMC geometry optimization of keto-1, enol-1 and enol-2 forms of oxyluciferin.

	Carbon	Nitrogen	Oxygen	Sulfur	Hydrogen
AGP	(4s4p)/[2s2p]	(4s4p)/[2s2p]	(4s4p)/[2s2p]	(4s4p)/[4s2p]	(3s1p)/[2s1p]
J_3, J_4	(3s2p)/[2s1p]	(3s2p)/[2s1p]	(3s2p)/[2s1p]	(3s2p)/[2s1p]	(2s1p)/[1s1p]

2.2 Electronic and Optical properties

Electronic and optical properties of the keto-1 and enol-1 forms have been calculated at DFT/TDDFT, CC2 and GW/BS levels. For enol-2, only DFT and TDDFT calculations have been performed. Kohn-Sham frontier orbitals have been calculated using Def2-SVPD, cc-pVTZ and aug-cc-pVTZ Dunning’s basis sets. TDDFT calculations have been carried out using BLYP, B3LYP, CAM-B3LYP and M062X functionals with Gaussian basis sets. For VMC and CAM-B3LYP structures we have employed basis sets up to the cc-pVQZ.

MBGFT excited-state calculations based on the GW/BS approach have been performed for the keto-1 and enol-1 forms. First we computed the single-particle Kohn-Sham states ϕ_i^{KS}

and corresponding energies ϵ_i^{KS} , needed as starting-point in GW/BS procedure, then quasi-particle energies have been computed by considering the GW self energy in the quasi-particle equation. GW quasi-particle energies ϵ_i^{GW} are given by:

$$\epsilon_i^{GW} = \epsilon_i^{KS} + \langle \phi_i^{KS} | \Sigma^{GW}(\epsilon_i^{GW}) - V_{xc} | \phi_i^{KS} \rangle \quad (6)$$

where Σ^{GW} is the GW self energy, which is the product between the Kohn-Sham Green's function G^{KS} and the screened Coulomb interaction W :

$$\Sigma^{GW}(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{2\pi} \int d\omega' e^{i\eta\omega'} G^{KS}(\mathbf{r}, \mathbf{r}', \omega + \omega') W(\mathbf{r}, \mathbf{r}', \omega'), \quad (7)$$

where the term $e^{i\eta\omega'}$ enforces the correct time ordering of the self energy. The screened potential W is obtained within the random phase approximation (RPA) and V_{xc} is the usual DFT exchange-correlation potential.

In order to remove the dependency of the final quasi-particle states on the specific functional used for the DFT starting point, a partially self-consistent scheme (evGW) has been applied, where the ϵ_i^{GW} obtained from Eq. 6 are then reinserted in the construction of the Green's function and the polarizability in Eq. 7 and Eq. 6 is iterated until self consistency is reached. In this procedure the Kohn-Sham wave functions are kept frozen. It has been shown by several authors that this partial self-consistent method strongly improves HOMO-LUMO gaps and thus BS excitation energies.^{71,92-95} We noticed that, in the evGW procedure, the dependence of the results on the initial conditions is strongly reduced for the keto-1 form, while it is only slightly reduced for the enol-1 form (see below). Therefore, as starting eigenvalues and eigenfunctions, we used those obtained from the LDA and CAM-B3LYP approximations for the exchange and correlation functional. Excitation energies were then computed by solving the BS equation, i.e. including the electron-hole interaction (e-h). An

electron-hole state can be described as:

$$\Phi^s(\mathbf{r}_e, \mathbf{r}_h) = \sum_i^{\text{occupied}} \sum_a^{\text{virtual}} A_{ia}^s \phi_a(\mathbf{r}_e) \phi_j^*(\mathbf{r}_h) + B_{ia}^s \phi_j(\mathbf{r}_e) \phi_a^*(\mathbf{r}_h) \quad (8)$$

where A_{ia}^s and B_{ia}^s are the resonant and anti-resonant amplitudes. In this basis the BS equation can be mapped onto a non-Hermitian eigenvalue problem^{58,96} and the A_{ia}^s and B_{ia}^s are then obtained as solutions of an excitonic Hamiltonian that reads:

$$\begin{pmatrix} R & C \\ -C^* & -R^* \end{pmatrix} \begin{pmatrix} A^s \\ B^s \end{pmatrix} = E^s \begin{pmatrix} A^s \\ B^s \end{pmatrix} \quad (9)$$

In the excitonic Hamiltonian R is the Hermitian resonant part:

$$R = (E_a^{GW} - E_i^{GW}) \delta_{i,j} \delta_{a,b} + \langle ai|K|jb\rangle \quad (10)$$

and C is the coupling symmetric part:

$$C = \langle ai|K|\overline{jb}\rangle, \quad (11)$$

where $K = W - 2V$ is the BS excitonic kernel, with W and V the screened and bare Coulomb interaction respectively, and \overline{jb} indicates the electron-hole anti-pairs. MBGFT calculations have been performed as implemented in the all-electron Gaussian basis set MolGW⁹⁷ package using the cc-pvQZ basis set and resolution-of-the-identity (RI) technique.⁹⁸ The RI technique expresses four-center integrals in terms of two and three-center integrals by using auxiliary basis sets. All virtual states are included in the construction of both the polarizability and the self-energy, while core states are neglected. Quasi-particle energies of Eq. 6 have been computed via graphical solution without resorting to linearization around Kohn-Sham energy for the correlation part of the self-energy. The BS equation has been solved by considering the full matrix of Eq. 9, beyond the Tamm-Dancoff approximation. For the keto-1 form

we have also calculated excitations at evGW/BS level using the plane-wave code Yambo⁵⁹ starting from the ground-state electronic structure calculated by the Quantum ESPRESSO package,⁹⁹ obtained by Troullier-Martins norm-conserving pseudo-potentials with a plane-wave cutoff of 60 Ry for the wave functions. The spurious interactions between images coming from periodic boundary conditions have been avoided by truncating the long-range Coulomb interaction using the cutoff technique described in Ref.¹⁰⁰ For the GW correction 1000 states have been included to calculate the dynamical dielectric matrix (8 Ha cutoff) and the Green's function. The dynamical screening is treated within the plasmon-pole approximation.¹⁰¹ For the BS spectra, we have included electron-hole pairs considering 33 occupied and 95 unoccupied states. 1500 states have instead been included in the static polarization function. We have used a cutoff of 10 Ha for the screened interaction and of 20 Ha for the exchange component.

The Molpro package¹⁰² has been used for the CC2 excited-state calculations with the cc-pVDZ and cc-pVTZ Dunning's basis set, while, unfortunately, cc-pVQZ basis set produces memory and instability issues.

We have also performed DFT geometry optimizations and TDDFT excited-state calculations in implicit solvent with the polarizable continuum model (PCM)¹⁰³ using the Gaussian 09 code. In detail, keto-1, enol-1 and enol-2 forms have been optimized in water and toluene using the Def2-SVPD basis set at DFT/CAM-B3LYP level of theory. The same functional has been used for the TDDFT calculations. In order to assess the stability of the dianionic enol-2 in gas phase, frontier orbitals and HOMO-LUMO gap have been computed with BLYP, B3LYP, CAM-B3LYP and M062X functionals and cc-pVTZ, aug-cc-pVTZ and Def2-SVPD basis sets.

3 Results and discussion

Starting from the consensus about the possible role as main emitters of keto-1, enol-1 and enol-2 forms, we report our results on the structural properties in Section 3.1. The vertical $S_1 \leftarrow S_0$ excitations for keto-1 and enol-1 are presented in Section 3.2, and in Section 3.3 we discuss the excitations of all three forms in the presence of a solvent, emphasizing the intrinsic difficulty to treat the enol-2 dianion in the gas phase.

3.1 S_0 gas phase structures

In this section we compare a selection of bond lengths obtained relaxing the ground-state structures at different levels of theory. These geometric parameters are key quantities to assess the quality of the ground state and to properly compute the vertical absorption energies. The comparison aims at assessing the role of the computational method in determining geometric parameters for a given form, and to show the differences in the geometrical parameters among the three forms, when the structures are optimized using the same approach.

To simplify the discussion, we have numbered the involved atoms as reported in Figure 1. In detail, we have focused the attention on the $C_{2'}$ - C_2 central bridge bond, which is characterized by an hybrid nature of single and double bond, the C_2 - N_3 and N_3 - C_4 bonds, the C_5 - S_1 bond and the C_4 - C_5 bond on the five-ring moiety, together with the two C-O bonds. All the values are reported in Table 2. The bond lengths between two atoms m and n will be indicated as $R_{m,n}$. Following the analysis reported in Ref.,³³ we preferred to analyze those bonds that display a resonance character, and which are therefore involved in the π conjugation of the OxyLuc forms. In Table S1 of the Supporting Information (SI) we also report the 3-2-2' and 1-2-2' angles and the 2-2-2'-3' dihedral for the optimized structures of the three forms.

Starting from the keto-1 form, we observe that the CASSCF(18,15)¹² bond length for $C_{2'}$ - C_2 is the shortest (1.401 Å), corresponding to a (slightly) smaller flexibility with respect to

the structures obtained with the other methods. This is a quite expected result, since it is well known that a proper description of the dynamic correlation (reducing the difference between single and double bonds in a conjugated moiety) is lacking in the CASSCF representation of the wave function. On the other hand, VMC (1.408(3) Å), CAM-B3LYP (1.412 Å), M062X (1.415 Å) and B3LYP¹⁴ (1.414 Å) values are very close to each other, with the exception of BLYP (1.420 Å).

For what concerns the two C-N bonds in the (formally) -N=C-C=N- moiety, a small but significant asymmetry in the bond length is found, with $R_{2',3'}$ longer than $R_{2,3}$ in all the structures considered in this work. Interestingly, the difference between $R_{2',3'}$ and $R_{2,3}$ increases when reducing the amount of dynamic correlation, moving from BLYP to CASSCF(18,15): the calculated differences are 9 mÅ for BLYP, 12 mÅ for B3LYP, 14 mÅ for CAM-B3LYP and M062X, 16(3) mÅ for VMC and 25 mÅ for CASSCF(18,15). CAM-B3LYP calculations produce bond distances in very good agreement with the VMC findings.

Besides the CASSCF(18,15) results, the VMC geometry shows the shortest double bond $R_{4,5}$, 1.538(2) Å. This finding is in fair agreement with what reported in the previous study of the VMC/JAGP computational protocol on linear conjugated chromophores as the retinal protonated Schiff base,³⁹ the peridinin carotenoid⁴⁰ and polyacetylene fragments:³⁸ BLYP and B3LYP functionals overcorrelate the polyenic chain, resulting in a small bond alternation length (i.e., too long double carbon bonds, as in this case), while VMC/JAGP produces a more balanced description of the electronic π -delocalization.

The VMC estimate of $R_{5,1}$, involving the sulfur atom, gives the shortest bond length (1.796(1) Å), around 10 mÅ smaller than the CAM-B3LYP value (1.807 Å). $R_{4,11'}$ and $R_{6',10'}$ oxygen-carbon bonds are not dramatically affected by the method chosen for the geometry optimization, with CAM-B3LYP (and M062X) and VMC producing very similar bond lengths; the only exceptions are given by the CASSCF(18,15) and BLYP, with respectively shorter and longer bond lengths. For $R_{3,4}$, on the other hand, all the methods show bond distances in a range of 10 mÅ.

An accurate description of the bond length pattern is essential for the computation of vertical absorption energies in biochromophores.³⁹⁻⁴¹ Therefore a comparison of the obtained structures with the one obtained with highly accurate quantum chemistry methods as CCSD(T) would be highly desirable. To the best of our knowledge, unfortunately, no CCSD(T) result is present in the literature on geometry optimization of (nonsymmetric) biological chromophores. VMC structures on polyenic chains have been shown to be as accurate as the CCSD(T) ones for the bond length alternation,³⁸ therefore we are confident that the VMC structure can be taken as the theoretical reference for assessing the quality of various DFT approaches. Here, we clearly see that the CAM-B3LYP functional provides structures in quantitative agreement with the VMC ones, with a maximum discrepancy of 11 mÅ, thus enforcing the idea that the use of CAM-B3LYP for the geometry optimization of chromophores is a fully reliable choice as already observed in previous studies.^{39,40} This conclusion can be also extended to M062X.

Qualitatively, the various methods adopted for the geometry optimization of keto-1 behave similarly when applied to enol-1 and enol-2. Looking at the specific features of enol-1 and enol-2 forms, in the VMC enol-1 structure we found a bridge bond $R_{2,2'}$ of 1.434(1) Å, longer than in keto-1 (1.408(3) Å). The large differences found for $R_{4,5}$ and $R_{4',11'}$ between keto-1 and enol-1 are obviously due to the keto-enol tautomerization. A rather large difference is also seen for the bond involving the sulfur atom, with a shortening of 56(2) mÅ. The $C_{2'}-C_2$ is even longer in the enol-2 form (1.457 Å).

Generally, the comparative analysis of keto-1, enol-1 and enol-2 forms allows us to state that the level of theory used for the geometry optimization of the three species does not affect much the relative structural properties.

Trans isomers with respect to the bridge $R_{2'-2}$ bond have been reported to be more stable than the cis ones in gas phase, as shown by simple calculations using the Boltzmann distribution and molecular dynamics simulations.¹⁴ At room temperature, the isomerization from trans to cis of the keto-1 form has substantially zero probability to occur.

Table 2: Optimized bond lengths (in Å) for the VMC and DFT (BLYP, B3LYP, M062X and CAM-B3LYP) structures of keto-1, enol-1 and enol-2. Def2-SVPD basis set has been used for DFT calculations. ^aGeometry from Ref. 14. ^bCASSCF(18,15) structures, with the the ANO-RCC-VDZP basis set, from Ref. 12.

	R _{2',2}	R _{2,3}	R _{2',3'}	R _{3,4}	R _{4,5}	R _{5,1}	R _{4,11'}	R _{6',10'}
keto-1								
BLYP	1.420	1.326	1.335	1.380	1.565	1.841	1.235	1.267
B3LYP ^a	1.414	1.313	1.325	1.371	1.550	1.821	1.221	1.251
M062X	1.415	1.304	1.318	1.375	1.545	1.806	1.210	1.241
CAM-B3LYP	1.412	1.305	1.319	1.370	1.541	1.807	1.214	1.245
VMC	1.408(3)	1.307(2)	1.323(2)	1.376(1)	1.538(2)	1.796(1)	1.211(1)	1.243(3)
CASSCF(18,15) ^b	1.401	1.292	1.317	1.375	1.523	1.803	1.203	1.230
enol-1								
BLYP	1.432	1.335	1.325	1.365	1.382	1.767	1.378	1.271
B3LYP	1.430	1.321	1.312	1.356	1.370	1.750	1.361	1.257
M062X	1.437	1.311	1.302	1.357	1.365	1.736	1.354	1.248
CAM-B3LYP	1.436	1.311	1.302	1.356	1.362	1.739	1.354	1.252
VMC	1.434(1)	1.312(2)	1.303(1)	1.361(2)	1.354(1)	1.730(2)	1.357(2)	1.252(1)
CASSCF(18,15) ^b	1.436	1.294	1.292	1.361	1.349	1.738	1.350	1.238
enol-2								
BLYP	1.452	1.315	1.316	1.434	1.434	1.760	1.273	1.288
B3LYP	1.452	1.302	1.301	1.422	1.420	1.745	1.261	1.273
M062X	1.462	1.292	1.290	1.424	1.413	1.735	1.254	1.264
CAM-B3LYP	1.460	1.293	1.290	1.421	1.409	1.738	1.257	1.269
VMC	1.457(2)	1.293(2)	1.292(2)	1.430(3)	1.404(4)	1.728(4)	1.258(2)	1.268(1)
CASSCF(18,15) ^b	1.460	1.278	1.281	1.434	1.383	1.752	1.246	1.253

Moreover, at MS-CASPT2¹³ and DFT¹⁴ levels of theory, the keto-1 form is significantly more stable than the enol-1, with trans-cis energy differences of 13.4 and 11.3 kcal/mol, respectively. Thus, assuming that in the gas-phase experiment only the keto-1 form exists, we are allowed to compute the isomerization energy of the keto-1 species. On the other hand, the enol-1 form was found to be slightly more stable than keto-1 in water clusters.⁷⁶ Using the Def2-SVPD basis set we have estimated the trans-cis energy difference at DFT level for the keto-1 form, obtaining 12.7 kcal/mol with BLYP, 9.5 kcal/mol with CAM-B3LYP and 7.3 kcal/mol with M062X. The CAM-B3LYP energy difference (taken as a representative example) diminishes when the two forms are embedded in a solvent: 7.1 kcal/mol in water and 8.2 kcal/mol in toluene. These findings confirm the fact that the experimental absorption energy corresponds to the trans isomer of the keto-1 form: indeed, the fraction of the enol-1 form, even using the smallest energy gap, is less than 10⁻⁵.

In Ref.,¹⁴ the tautomer identity of the OxyLuc moiety in gas phase was ascertained by the absorption experiment done using the 5,5-dimethylated OxyLuc anion, where the molecule is

locked as the keto-1 form by substitution. The red-shift observed in the absorption maximum matches well with the substitution effects measured in solution and computed at EOM-CCSD and TDDFT level.

The results of this analysis, of course, can not easily be extended to OxyLuc embedded in the various protein environments, where, as explained in the introduction, many factors occur in defining the role of the emitter species.

Even though literature^{13,14} and present results agree in identifying the trans isomer of the keto-1 form as the most stable moiety in gas phase at ambient conditions, the systematic comparison carried out between the bond lengths of keto-1, enol-1 and enol-2 represents an useful investigation on the correlation between geometric and optical properties, as reported in the next Subsection.

3.2 $S_1 \leftarrow S_0$ vertical absorption for keto-1 and enol-1

In gas-phase the only existing OxyLuc form is keto-1, as inferred from the stability arguments reported above. From a theoretical standpoint, therefore, we now wish to compare the experimental absorption of 2.26(8) eV (548(10) nm) obtained using the action spectroscopy technique¹⁴ with the vertical excitation energy computed on the keto-1 form. In order to achieve this goal, we computed the $S_1 \leftarrow S_0$ excitation energy (ΔE) using TDDFT with several functionals, evGW/BS (with Gaussian basis sets and plane waves), and CC2 on structures optimized at different levels of theory (VMC, DFT and CASSCF). Large variations exist in the ΔE value according to the chosen computational protocol. Moreover, the same systematic study on the optical properties of enol-1 revealed that the absorption energy shift of enol-1 with respect keto-1 can change sign when the theoretical approach and the convergence degree with respect to the basis set are changed.

The $S_1 \leftarrow S_0$ transition studied in this work is characterized by a predominant single-excitation nature.¹⁴ TDDFT excitation energies calculated on the different structures of keto-1 form obtained as described above are shown in Figure 2. It is evident that TDDFT

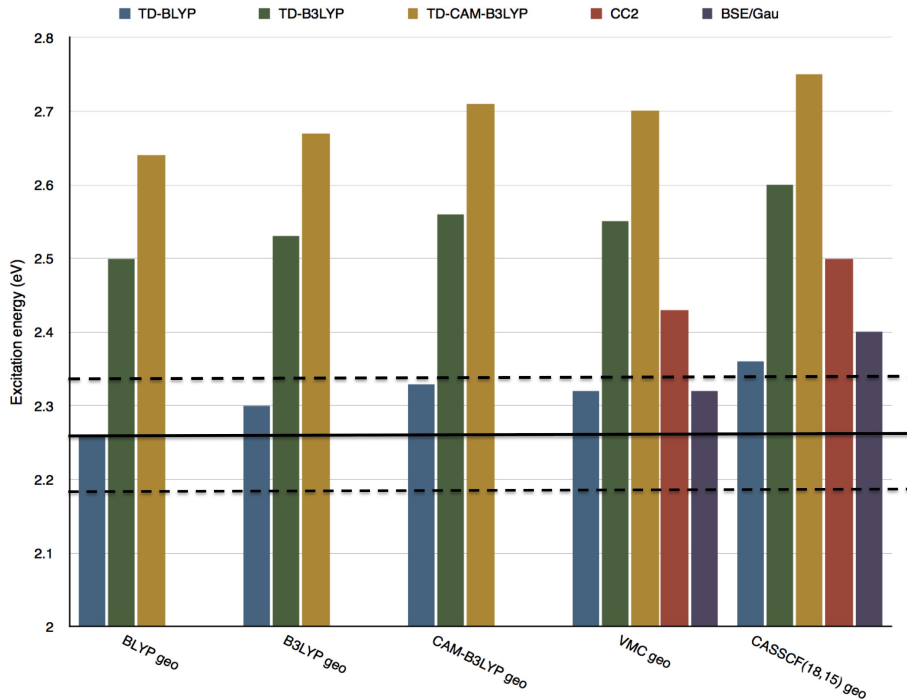


Figure 2: Excitation energies (in eV) for the keto-1 form using TDDFT (BLYP, B3LYP¹⁴ and CAM-B3LYP with Def2-SVPD basis set), CC2 (cc-pVTZ basis set) and evGW/BS (cc-pVQZ basis set). Optimized ground-state geometries from DFT (BLYP, B3LYP and CAM-B3LYP), CASSCF(18,15)¹² and VMC. Horizontal lines correspond to the experimental value of 2.26 eV (solid) and the related error (± 0.08 eV, dashed). The corresponding TDDFT values are collected in Table S2 of SI.

results show a poor agreement with the experimental value, with the exception of the semi-local BLYP functional, probably because of a cancellation of errors, as already suggested in Ref.¹⁴ TDDFT calculations have been performed using the Def2-SVPD basis set, as done in Ref.¹⁴ and, as pointed out in the same reference, the results are clearly affected by the choice of the exchange correlation kernel. On the other hand, it is important to stress that the fine convergence with respect to the basis set is of fundamental importance when determining the relative shift in energy between the keto-1 and enol-1 excitations. The complete list of results is reported in SI in Tab. S2 and S3.

From Fig. 2 it is clear that B3LYP and CAM-B3LYP functionals overestimate the excitation energy of the keto-1 form, regardless of the optimized structure: the B3LYP excitation energy spans from 2.50 to 2.60 eV, the CAM-B3LYP is always larger than 2.60 eV. TDDFT

values with the M062X functional show the same trend of those using CAM-B3LYP (Table S2 in SI). As expected, the net effect of the ground-state structure is to induce a blue shift, when the same excited-state approach is employed, in conjunction with a decrease of the correlation along the conjugated bonds: the CASSCF(18,15) structure produces the highest S_1 energy, as also found for the enol-1 form as shown in Figure 3. In summary, the level of theory employed for the structural optimization produces a moderate blue shift in the vertical excitation, smaller than 0.1 eV when the difference between single and double bonds increases. This is a much more reduced effect than those observed for linear chromophores like the peridinin carotenoid.⁴⁰ We note here that CAM-B3LYP and VMC geometries determine essentially indistinguishable excitations for the keto-1 form (Figure 2 and Table S2), provided the same functional for TDDFT is used.

TDDFT excitation energies computed using the 6-311+G* basis set, reported in Table S3 of SI, show no substantial difference with the Def2-SVPD set of results.

Excluding the poor TD-BLYP results, only the combination of VMC and BS is able to accurately match with the experimental value of 2.26(8) eV, as shown in Figure 2. The evGW/BS excitation energy for the keto-1 form (2.32 eV) was obtained using the cc-pVQZ Gaussian basis set starting from CAM-B3LYP eigenvalues and eigenstates. We have tested the dependence on the choice of the exchange and correlation functional in the ground state in the partially self-consistent evGW/BSE procedure using also LDA functional obtaining a very similar result, 2.34 eV (see Table S5). The convergence study (on the VMC structure of the keto-1 form) of the excitation energy as a function of the size of the basis set is reported for TDDFT (CAM-B3LYP) and evGW/BS in Tables S4 and S5 of SI, respectively: the energy decreases when increasing the complexity of the basis set, the cc-pVQZ being the largest basis set one can reasonably apply to OxyLuc systems. The convergence for the evGW/BS excitation energy has been also validated with respect a plane wave calculation starting from LDA ground state (Table S5).

Similarly, we report in Figure 3 the values of the excitation energy for the enol-1 form,

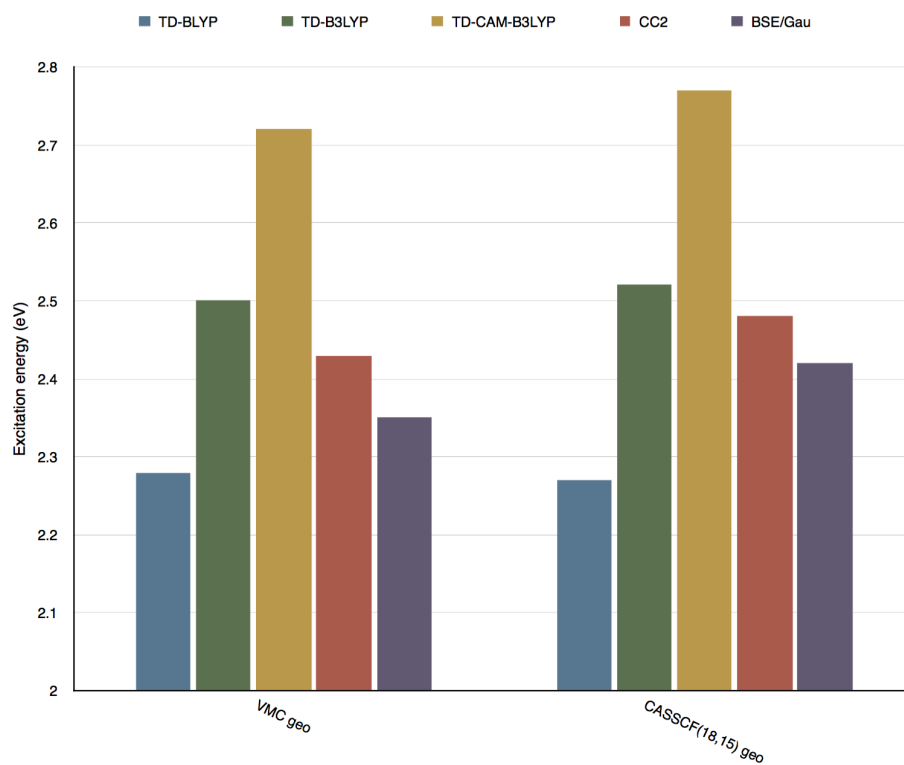


Figure 3: Excitation energies (in eV) for the enol-1 form using TDDFT (BLYP, B3LYP and CAM-B3LYP with Def2-SVPD basis set), CC2 (cc-pVTZ basis set) and evGW/BS (cc-pVQZ basis set). Optimized ground-state geometries from VMC and CASSCF(18,15).¹² The corresponding TDDFT values are collected in Table S2 in SI.

using the ground-state optimized VMC and CASSCF(18,15) structures, at TDDFT (BLYP, B3LYP and CAM-B3LYP), CC2 and evGW/BS level. TDDFT excitation energies are also reported in S2 of SI, while BS and CC2 energies are in Tables 3 and 4, respectively. The CASSCF(18,15) structure produces a systematic shift towards larger energies regardless of the excited-state approach. In particular, the net effect of using the CASSCF(18,15) geometry of the enol-1 form is to increase the excitation energy by 0.07 eV, when calculated in GW/BS framework, and by 0.05 eV for CC2. Also in this case, the TDDFT results span a large energy range (around 0.4 eV) according to the functional employed for the calculation, regardless of the chosen geometry.

Looking at the keto-1/enol-1 S_1 energy shift (defined as the difference between the keto-1 and the enol-1 excitation energies, see Table 5), one can compare the behavior of the different excited-state techniques applied on the VMC and CASSCF(18,15) geometries. A robust computational reference for the gas phase $S_1 \leftarrow S_0$ excitation energies is given by the MS-CASPT2 results,¹³ even though the transition energies slightly overestimate the experimental transition value (2.43 eV for the keto-1 form): the MS-CASPT2 keto-1/enol-1 shift (computed on the CASSCF(18,15) structures) is -0.05 eV. The relative keto-1/enol-1 shifts computed at different levels of theory are collected in Table.5. At TDDFT level, only the use of CAM-B3LYP and M062X functionals on the VMC structures can correctly reproduce the sign and the magnitude of the shifts predicted by the accurate MS-CASPT2 analysis. Indeed, a shift of -0.02 (CAM-B3LYP) and -0.03 eV (M062X) is found for the enol-1 form with respect to the keto-1 one, while BLYP and B3LYP yield a positive S_1 energy gap. The BS shift is characterized by the same sign of the MS-CASPT2 one, and an absolute value equal to that of TDDFT with CAM-B3LYP and M062X: -0.03 (VMC geometry) and -0.02 eV (CASSCF(18,15) geometry).

It is important to note that while the keto-1 evGW/BS excitation energy is only barely affected by the choice of the DFT initial guess, passing from CAM-B3LYP (2.32 eV) to LDA (2.34 eV), a large effect (i.e. blue shift) is instead observed for enol-1 passing from 2.35 eV

to 2.45 eV when LDA ground state is considered as a starting point. Such a difference needs to be ascribed to the different localization of the wave functions at CAM-B3LYP and LDA level for the enol-1 form. The LDA-based partially self consistent evGW/BS excitations for VMC and CASSCF(18,15) structures are reported in Tab. S7.

Using a cc-pVTZ basis set, the CC2 shift is zero when the VMC structure is used, and 0.02 eV for the CASSCF(18,15). Unfortunately, CC2 calculations with a cc-pVQZ basis set are not affordable since they are extremely memory demanding; however, the trend in the excitation energies (see Table S6) passing from cc-pVDZ to cc-pVTZ indicates that, for an even larger basis set, the sign of the keto-1/enol-1 shift could be the same of the one obtained in evGW/BS and MS-CASPT2 framework.

Table 3: evGW/BS vertical $S_1 \leftarrow S_0$ absorption energies using the cc-pVQZ basis set for the keto-1 and enol-1 forms with VMC and CASSCF structures. CAM-B3LYP ground state is used as starting point for the partially self-consistent evGW. The experimental excitation energy is 2.26 ± 0.08 eV (548 ± 10 nm). ^aStructure from Ref. 12.

S_0 Geo	Form	ΔE (nm)	ΔE (eV)
VMC	keto-1	534	2.32
	enol-1	528	2.35
CASSCF(18,15) ^a	keto-1	517	2.40
	enol-1	512	2.42

Table 4: CC2 vertical vertical $S_1 \leftarrow S_0$ absorption energies using the cc-pVTZ basis sets on keto-1 and enol-1 forms with VMC and CASSCF structures. ^aStructure from Ref. 12. The experimental excitation energy is 2.26 ± 0.08 eV (548 ± 10 nm).

S_0 Geo	Form	ΔE (nm)	ΔE (eV)
VMC	keto-1	510	2.43
	enol-1	510	2.43
CASSCF(18,15) ^a	keto-1	496	2.50
	enol-1	500	2.48

Table 5: Relative shifts between keto-1 and enol-1 forms for vertical absorption energies. The Def2-SVPD basis set has been used for the TDDFT calculations. ^aStructures from Ref. 12. ^bData from Ref. 13.

S ₀ Geo	S ₁ ← S ₀	Shift (nm)	Shift (eV)
VMC	BLYP	-9	0.04
	B3LYP	-10	0.05
	CAM-B3LYP	4	-0.02
	CAM-B3LYP ^a	3	-0.02
	GW/BS	6	-0.03
	CC2	0	0.00
CASSCF(18,15) ^a	MS-CASPT2 ^b	10	-0.05
	BLYP	-23	0.09
	B3LYP	-15	0.08
	CAM-B3LYP	3	-0.02
	GW/BS	5	-0.02
	CC2	-4	0.02

3.3 S₁ ← S₀ vertical absorption for enol-2

The study of the ground and excited-state properties of the gas phase enol-2 form represents a challenge for any theoretical approach, since a double negative charge characterizes the electronic structure of the system (Figure 1). A deeper insight into the properties of the HOMO and LUMO energies is therefore needed in order to understand whether the two electrons in excess can maintain the electronic density bound in gas phase conditions, without a stabilizing counterion.

The HOMO and LUMO energies of the gas-phase keto-1, enol-1 and enol-2 forms are reported in Table S8 of SI. These calculations were done considering the CAM-B3LYP structures, which, as previously mentioned, are close to the VMC ones. First we observe that for the singly-charged systems, the HOMO orbital is bound with a negative energy, regardless of the specific choice for the calculation, whereas the LUMO orbital lies in the continuous part of the eigenvalue spectrum. Energies are sensitive to the addition of diffuse functions to the cc-pVTZ basis set, but the gap remains substantially unchanged (the largest difference being of 0.1 eV for the enol-1 using M062X and aug-cc-pVTZ), confirming that the electronic density is well localized around the nuclei of the keto-1 and enol-1 species.

The behaviour of HOMO and LUMO energies, and of the corresponding gap dramatically changes when looking at the enol-2 form, showing a strong dependence on the chosen basis set. The HOMO orbital is characterized by a positive energy, which significantly decreases (about 0.3-0.4 eV) when augmentation is added. A similar enhanced effect is found for the LUMO energy, decreasing of about 1.0-1.8 eV when passing from cc-pVTZ to aug-cc-pVTZ, thus displaying an evident decrease of the gap. The same issue regarding the convergence of the gap is found when plane waves are used as basis. In Table S9 of SI, DFT and Hartree-Fock HOMO and LUMO energies for increasing supercell size are reported. DFT HOMO energy is always positive regardless the box size, while the LUMO energy drastically reduces when the simulation box increases. At Hartree-Fock level, the HOMO orbital becomes weakly bound when increasing the box size, while a not clear convergence is seen for LUMO; moreover, convergence issues in the DFT self consistent cycle arise when even larger simulation boxes are used.

The results of the tests performed both in Gaussian basis set and in plane waves clearly show that ground state density for the gas phase enol-2 is unbound, thus questioning the reliability of previous gas phase calculations present in the literature.

The ANO-RCC-VDZP basis set used to compute MS-CASPT2 absorption of the gas phase enol-2 in Ref. ¹³ contains diffuse functions, but this does not guarantee the robustness of the results for the reason above. Convergence studies with respect to the basis set size for the enol-2 form have then been carried out including solvation effects via a polarizable continuum model¹⁰³ considering the chromophore embedded in water and toluene. Consistently with the calculations shown above, geometry relaxation of the three forms in solvent has been performed using the CAM-B3LYP functional and the Def2-SVPD basis set. The values of the HOMO-LUMO gaps obtained with three functionals for three basis sets are reported in Table 6. The corresponding energies are instead given in Table S10 of SI. HOMO and LUMO energies and the corresponding gaps for the keto-1 and enol-1 forms in water and toluene are instead reported in Tables S11 and S12, respectively. The direct interaction with

Table 6: HOMO-LUMO gaps (eV) at DFT level for enol-2 in water and toluene, using different basis sets. VTZ and aVTZ are abbreviations for cc-pvTZ and aug-cc-pvTZ, respectively.

eV	BLYP			B3LYP			CAM-B3LYP		
	VTZ	aVTZ	Def2-SVPD	VTZ	aVTZ	Def2-SVPD	VTZ	aVTZ	Def2-SVPD
water	1.70	1.72	1.73	2.90	2.91	2.92	5.24	5.23	5.25
toluene	1.65	1.67	1.68	2.85	2.85	2.88	5.20	5.18	5.21

a polarizable medium stabilizes the electronic density of the dianion, making the HOMO-LUMO gap less sensitive to the basis set, while, as expected, an evident dependence on the DFT functional is still present. In water, both HOMO and LUMO energies become negative and the corresponding gap (for a given functional) is not strongly affected by the augmentation. The same stability is found for the enol-2 in toluene, a non-polar solvent, even if only the HOMO orbital has a bound energy.

The $S_1 \leftarrow S_0$ transition on the CAM-B3LYP structures of the three forms in water and toluene have been calculated using the CAM-B3LYP and the Def2-SVPD basis set. The results are shown in Table 7, together with the gas phase value for keto-1 (see Table S2) and enol-1.

It is interesting to note that the three forms are characterized by a different response with respect to the solvent: keto-1 excitation energy is red-shifted with respect to the gas phase, with only a small variation in water (0.04 eV); a large blue-shift is seen for the enol-1 form in water, while only small differences are found in toluene.

In water, enol-1 shows the maximum excitation energy; our calculations qualitatively confirm the relative shift in water among the three forms already computed at TDDFT/CAM-B3LYP level using the 6-31G** basis set,³³ though the ΔE for keto-1 and enol-1 are sensibly larger (2.89 and 3.23 eV, compared with our values 2.67 and 3.00 eV), indicating a not negligible basis and geometry effect. On the other hand, our calculations provide the same excitation energies in water of the one found in Ref.¹⁰⁴ at PCM/aug-cc-pVTZ level. More significantly, our results show a quantitative agreement with the experimental findings of Ref.³⁵ for the three forms in water.

Table 7: TDDFT/CAM-B3LYP absorption energies (in eV) for the three OxyLuc forms using the Def2-SVPD basis set and CAM-B3LYP structures optimized in gas phase, water and toluene. ^a Experimental absorption maxima from Ref. ³⁵

Form	Gas phase	Toluene	Water	Exp in Water ^a
keto-1	2.71	2.53	2.67	2.56
enol-1	2.73	2.71	3.00	3.05
enol-2	nc	2.85	2.94	2.92

4 Conclusions

The $S_1 \leftarrow S_0$ absorption energies of three different forms of the gas phase oxyluciferin have been explored by means of high-level methods such as the VMC for the ground state geometry and the evGW/BS formalism for the excitation energies. Unlike findings shown for other biological chromophores, like the protonated Schiff base of the retinal and the peridinin carotenoid,^{39,40} the accuracy in the determination of the ground-state geometry of keto-1 and enol-1 forms affects the value of the vertical excitation $S_1 \leftarrow S_0$ energy only slightly. In the case of OxyLuc forms, the optimized ground state structure plays a minor role in the determination of the optical properties, with only a moderate blue shift in the vertical absorption when the difference between single and double bonds increases, as for CASSCF(18,15) structures. The value of 2.32 eV obtained by applying the BS equation on the VMC structure of the keto-1 form shows an excellent agreement with the experimental 2.26(8) eV finding.

As shown in previous works, VMC is able to include a balanced description of the electronic correlation. Assuming the VMC as a reference structure, we have observed that the much less computationally demanding CAM-B3LYP or M062X DFT structural relaxation provides accurate results for complex, non-symmetric and conjugated molecular systems as the anionic forms of OxyLuc.

Regardless of the level of theory employed for the ground-state geometry optimization, the keto-1/enol-1 shift computed at BS level has the same sign of the very accurate MS-CASPT2 results while TDDFT calculations using semi-local functionals and even hybrids

as B3LYP fail to predict the correct sign of the shift. However, when passing from LDA to CAM-B3LYP wave functions in evGW/BS calculations a shift of 0.1 eV is observed for the enol-1 form. This indicates that particular care needs to be paid in the choice of the initial guess when determining the keto-1/enol-1 shift.

In the last years the BS approach has been shown to provide accurate excitation energies on well-known sets of small molecules and organic dyes. The present work contributes, together with previous ones by us and other authors,^{62,105,106} in showing that it is a promising tool to compute excited state energies in complex chromophores of biological interest.

Finally, studying the dianionic enol-2 form we have shown that particular attention needs to be paid to its unbound electronic character, and that meaningful results are obtained only when considering the effect of stabilizing solvents.

Furthermore, vibronic terms could have a not negligible weight in the applied theoretical model and a deeper and more quantitative insight on these effects on the vertical absorption of the OxyLuc forms requires further investigation.

Acknowledgement

EC thanks the Labex MiChem part of French state funds managed by the ANR within the Investissements d’Avenir programme (contract grant number: ANR-11-IDEX-0004-02), the computational centre of the Laboratoire de Chimie Théorique and the University of L’Aquila (SIR fellowship) for partial financial support. DV acknowledges partial support from EU Centre of Excellence MaX-Materials at the eXascale (H2020 grant no. 676598). Computational resources were provided by the Caliban computer Centre of the University of L’Aquila, the Hydra cluster of the computational centre of the University of Modena and Reggio Emilia, and by the PRACE research infrastructure (project BIOCHROMO). Authors thank S. Caprasecca for the careful reading of the manuscript. This work has been supported by the European Research Council Project MultiscaleChemBio (n. 240624) within the VII

Framework Program of the European Union.

Supporting Information Available

The following files are available free of charge. Table S1 contains the optimized angles and dihedrals (in degrees) of the DFT, VMC and CASSCF(18,15) structures of keto-1, enol-1 and enol-2 forms. TDDFT values on various structures of keto-1 and enol-1 are reported in Tables S2 and S3. Convergence analysis of the absorption energy with respect to the size of the basis set is shown in Table S4 (TDDFT), S5 (BS) and S6 (CC2). The evGW/BS excitations for keto-1 and enol-1 starting from an LDA ground state are reported in Table S7. HOMO, LUMO energies and gaps of gas-phase keto-1, enol-1 and enol-2, using various functionals and basis sets, are given in Table S8. In Table S9 the same quantities for enol-2 calculated using plane-wave basis sets are given. In Tables S10, S11 and S12 HOMO, LUMO energies and gaps in water and toluene are collected for enol-2, keto-1 and enol-1, respectively.

References

- (1) Navizet, I.; Liu, Y.-J.; Ferré, N.; Roca-Sanjúan, D.; Lindh, R. *ChemPhysChem* **2011**, *12*, 3064–3076.
- (2) Hosseinkhani, S. *Cell. Mol. Life Sci.* **2011**, *68*, 1167–1182.
- (3) da Silva, L. P.; da Silva, J. C. G. E. *ChemPhysChem* **2011**, *12*, 3002–3008.
- (4) da Silva, L. P.; da Silva, J. C. G. E. *J. Chem. Theory. Comput.* **2011**, *7*, 809–817.
- (5) Navizet, I.; Roca-Sanjúan, D.; Yue, L.; Liu, Y.-J.; Ferré, N.; Lindh, R. *Photochem. Photobiol.* **2013**, *89*, 319–325.
- (6) Ugarova, N. N.; Brovko, L. Y. *Luminescence* **2002**, *17*, 321–330.

- (7) Fraga, H. *Photochem. Photobiol. Sci.* **2008**, *7*, 146–158.
- (8) Orlova, G.; Goddard, J. D.; Brovko, L. Y. *J. Am. Chem. Soc.* **2003**, *125*, 6962–6971.
- (9) Ren, A.-M.; Goddard, J. D. *Photochem. Photobiol. B, Biol.* **2005**, *81*, 163–170.
- (10) Yang, T.; Goddard, J. D. *J. Phys. Chem. A* **2007**, *111*, 4489–4497.
- (11) Min, C.-G.; Ren, A.-M.; Guo, J.-F.; Li, Z.-W.; Zou, L.-Y.; Goddard, J. D.; Feng, J.-K. *ChemPhysChem* **2010**, *11*, 251–259.
- (12) Liu, Y.-J.; Vico, L. D.; Lindh, R. *Photochem. Photobiol. A, Chem.* **2008**, *194*, 261–267.
- (13) Chen, S.-F.; Liu, Y.-J.; Navizet, I.; Ferré, N.; Fang, W.-H.; Lindh, R. *J. Chem. Theory Comput.* **2011**, *7*, 798–803.
- (14) Støchkel, K.; Hansen, C. N.; Houmøller, J.; Nielsen, L. M.; Anggara, K.; Linares, M.; Norman, P.; Nogueira, F.; Maltsev, O. V.; Hintermann, L.; Nielsen, S. B.; Naumov, P.; Milne, B. F. *J. Am. Chem. Soc.* **2013**, *135*, 6485–6493.
- (15) Wood, K. V. *Photochem. Photobiol.* **1995**, *62*, 662–673.
- (16) Ugarova, N. N. *Photochem. Photobiol. Sci.* **2008**, *7*, 218–227.
- (17) McElroy, W. D.; Seliger, H. H.; White, E. H. *Photochem. Photobiol.* **1969**, *10*, 153–170.
- (18) Viviani, V. R. *Cell. Mol. Life Sci.* **2002**, *59*, 1833–1850.
- (19) Milne, B. F.; Marques, M. A. L.; Nogueira, F. *Phys. Chem. Chem. Phys.* **2010**, *12*, 14285–14293.
- (20) da Silva, L. P.; da Silva, J. C. G. E. *ChemPhysChem* **2011**, *12*, 951–960.
- (21) McCapra, F.; Gilfoyle, D. J.; Young, D. W.; N. J. Church, P. S. *Bioluminescence and Chemiluminescence: Fundamental and Applied Aspects*; Chichester: Wiley, New York, 1994; pp 387–391.

- (22) DeLuca, M. *Biochemistry* **1969**, *8*, 160–166.
- (23) Branchini, B. R.; Murtiashaw, M. H.; Magyar, R. A.; Portier, N. C.; Ruggiero, M. C.; Stroh, J. G. *J. Am. Chem. Soc.* **2002**, *124*, 2112–2113.
- (24) Branchini, B. R.; Southworth, T. L.; Murtiashaw, M. H.; Magyar, R. A.; Gonzalez, S. A.; Ruggiero, M. C.; Stroh, J. G. *Biochemistry* **2004**, *43*, 7255.
- (25) Nakatsu, T.; Ichiyama, S.; Hiratake, J.; Saldanha, A.; Kobashi, N.; Sakata, K.; Kato, H. *Nature* **2006**, *440*, 372–376.
- (26) Nakatani, N.; Hasegawa, J.; Nakatsuji, H. *J. Am. Chem. Soc.* **2007**, *129*, 8756–8765.
- (27) Nakatani, N.; Hasegawa, J.; Nakatsuji, H. *Chem. Phys. Lett.* **2009**, *469*, 191–194.
- (28) Navizet, I.; Liu, Y.-J.; Ferré, N.; Xiao, H.-Y.; Fang, W.-H.; Lindh, R. *J. Am. Chem. Soc.* **2010**, *132*, 706–712.
- (29) Min, C.-G.; Ren, A.-M.; Guo, J.-F.; Zou, L.-Y.; Goddard, J. D.; Sun, C.-C. *ChemPhysChem* **2010**, *11*, 2199–2204.
- (30) Tagami, A.; Ishibashi, N.; Kato, D.; Taguchi, N.; Mochizuki, Y.; Watanabe, H.; Ito, M.; Tanaka, S. *Chem. Phys. Lett.* **2009**, *472*, 118–123.
- (31) Cai, D.; Marques, M. A. L.; Nogueira, F. *J. Phys. Chem. B* **2011**, *115*, 329–332.
- (32) Song, C.-I.; Rhee, Y. M. *J. Am. Chem. Soc.* **2011**, *133*, 12040–12049.
- (33) Cheng, Y.-Y.; Liu, Y.-J. *J. Chem. Theory Comput.* **2015**, *11*, 5360–5370.
- (34) Naumov, P.; Ozawa, Y.; Ohkubo, K.; Fukuzumi, S. *J. Am. Chem. Soc.* **2009**, *131*, 11590–11605.
- (35) Ghose, A.; Rebarz, M.; Maltsev, O. V.; Hintermann, L.; Ruckebusch, C.; E. Fron, J. H.; Mély, Y.; Naumov, P.; Sliwa, M.; Didier, P. *J. Phys. Chem. B* **2015**, *119*, 2638–2649.

- (36) Pinto da Silva, L.; Esteves da Silva, J. C. G. *J. Phys. Chem. B* **2015**, *119*, 2140.
- (37) Foulkes, W. M. C.; Mitas, L.; Needs, R. J.; Rajagopal, G. *Rev. Mod. Phys.* **2001**, *73*, 33–83.
- (38) Barborini, M.; Guidoni, L. *J. Chem. Theory Comput.* **2015**, *11*, 4109–4118.
- (39) Coccia, E.; Varsano, D.; Guidoni, L. *J. Chem. Theory Comput.* **2013**, *9*, 8–12.
- (40) Coccia, E.; Varsano, D.; Guidoni, L. *J. Chem. Theory Comput.* **2014**, *10*, 501–506.
- (41) Varsano, D.; Coccia, E.; Mosca Conte, A.; Pulci, O.; Guidoni, L. *Comp. Theor. Chem.* **2014**, *1040-1041*, 338–346.
- (42) Casula, M.; Sorella, S. *J. Chem. Phys.* **2003**, *119*, 6500.
- (43) Casula, M.; Attaccalite, C.; Sorella, S. *J. Chem. Phys.* **2004**, *121*, 7110.
- (44) Barborini, M.; Sorella, S.; Guidoni, L. *J. Chem. Theory Comput.* **2012**, *8*, 1260–1269.
- (45) Coccia, E.; Chernomor, O.; Barborini, M.; Sorella, S.; Guidoni, L. *J. Chem. Theory Comput.* **2012**, *8*, 1952–1962.
- (46) Coccia, E.; Guidoni, L. *J. Comput. Chem.* **2012**, *33*, 2332–2339.
- (47) Barborini, M.; Guidoni, L. *J. Chem. Phys.* **2012**, *137*, 224309.
- (48) Barborini, M.; Guidoni, L. *J. Chem. Theory Comput.* **2015**, *11*, 508–517.
- (49) Varsano, D.; Caprasecca, S.; Coccia, E. *J. Phys.: Condens. Matter* **2017**, *29*, 013002.
- (50) Zen, A.; Coccia, E.; Luo, Y.; Sorella, S.; Guidoni, L. *J. Chem. Theory Comput.* **2014**, *10*, 1048–1061.
- (51) Zen, A.; Coccia, E.; Gozem, S.; Olivucci, M.; Guidoni, L. *J. Chem. Theory Comput.* **2015**, *11*, 992–1005.

- (52) Barborini, M.; Coccia, E. *J. Chem. Theory Comput.* **2015**, *11*, 5696–5704.
- (53) Chu, S.; Coccia, E.; Barborini, M.; Guidoni, L. *J. Chem. Theory Comput.* **2016**, *12*, 5696–5704.
- (54) Neuscamman, E. *J. Chem. Phys.* **2016**, *145*, 081103.
- (55) Zhao, L.; Neuscamman, E. *J. Chem. Theory Comput.* **2016**, *12*, 3719–3726.
- (56) Zhao, L.; Neuscamman, E. *J. Chem. Theory Comput.* **2016**, *12*, 3436–3440.
- (57) Fetter, A. L.; Walecka, J. D. *Quantum theory of many-body systems*; New York: McGraw-Hill, 1971.
- (58) Onida, G.; Reining, L.; Rubio, A. *Rev. Mod. Phys.* **2002**, *74*, 601.
- (59) Marini, A.; Hogan, C.; Grüning, M.; Varsano, D. *Comput. Phys. Commun.* **2009**, *180*, 1392–1403.
- (60) Hedin, L. *Phys. Rev.* **1965**, *139*, A796.
- (61) Strinati, G. *Phys. Rev. Lett.* **1982**, *49*, 1519.
- (62) Ma, Y.; Rohlfing, M.; Molteni, C. *J. Chem. Theory Comput.* **2009**, *6*, 257–265.
- (63) Blase, X.; Attaccalite, C. *Appl. Phys. Lett.* **2011**, *99*, 171909.
- (64) Baumeier, B.; Andrienko, D.; Rohlfing, M. *J. Chem. Theory Comput.* **2012**, *8*, 2790–2795.
- (65) Duchemin, I.; Blase, X. *Phys. Rev. B* **2013**, *87*, 245412.
- (66) Körbel, S.; Boulanger, P.; Duchemin, I.; Blase, X.; Marques, M. A.; Botti, S. *J. Chem. Theory Comput.* **2014**, *10*, 3934–3943.
- (67) Baumeier, B.; Rohlfing, M.; Andrienko, D. *J. Chem. Theory Comput.* **2014**, *10*, 3104–3110.

- (68) Jacquemin, D.; Duchemin, I.; Blase, X. *Mol. Phys.* **2015**, *114*, 957–967.
- (69) Jacquemin, D.; Duchemin, I.; Blase, X. *J. Chem. Theory Comput.* **2015**, *11*, 5340–5359.
- (70) Bruneval, F.; Hamed, S. M.; Neaton, J. B. *J. Chem. Phys.* **2015**, *142*, 244101.
- (71) Jacquemin, D.; Duchemin, I.; Blase, X. *J. Chem. Theory Comput.* **2015**, *11*, 3290–3304.
- (72) Stein, T.; Kronik, L.; Baer, R. *J. Chem. Phys.* **2009**, *131*, 244119.
- (73) Duchemin, I.; Deutsch, T.; Blase, X. *Phys. Rev. Lett.* **2012**, *109*, 167801.
- (74) Noguchi, Y.; Hiyama, M.; Akiyama, H.; Koga, N. *J. Chem. Phys.* **2014**, *141*, 044309.
- (75) Hiyama, M.; Noguchi, Y.; Akiyama, H.; Yamada, K.; Koga, N. *Photochem. Photobiol.* **2015**, *91*, 819–827.
- (76) Noguchi, Y.; Hiyama, M.; Shiga, M.; Sugino, O.; Akiyama, H. *J. Phys. Chem. B* **2016**, *120*, 8776–8783.
- (77) Frisch, M. J. et al. Gaussian-09 Revision E.01. Gaussian Inc. Wallingford CT 2009.
- (78) Rappoport, D.; Furche, F. *J. Chem. Phys.* **2010**, *133*, 134105.
- (79) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
- (80) Umrigar, C. J.; Toulouse, J.; Filippi, C.; Sorella, S.; Hennig, R. G. *Phys. Rev. Lett.* **2007**, *98*, 110201.
- (81) Sorella, S.; Casula, M.; Rocca, D. *J. Chem Phys.* **2007**, *127*, 14105.
- (82) Hammond, B. L.; Lester, W. A., Jr.; Reynolds, P. J. *Monte Carlo Methods in Ab-Initio Quantum Chemistry*; World Scientific, 1994.

- (83) Umrigar, C. J. *Int. J. Quan. Chem.* **1989**, *36*, 217–230.
- (84) Assaraf, R.; Caffarel, M. *J. Chem. Phys.* **2003**, *119*, 10536.
- (85) Sorella, S.; Capriotti, S. *J. Chem. Phys.* **2010**, *133*, 234111.
- (86) Drummond, N. D.; Towler, M. D.; Needs, R. J. *Phys. Rev. B* **2004**, *70*, 235119.
- (87) Marchi, M.; Azadi, S.; Casula, C.; Sorella, S. *J. Chem. Phys.* **2009**, *131*, 154116.
- (88) Zen, A.; Luo, Y.; Sorella, S.; Guidoni, L. *J. Chem. Theory Comput.* **2013**, *9*, 4332–4350.
- (89) Sorella, S. TurboRVB Quantum Monte Carlo package (accessed date February 21, 2017). <http://people.sissa.it/~sorella/web/index.html>.
- (90) Burkatzki, M.; Filippi, C.; Dolg, M. *J. Chem. Phys.* **2007**, *126*, 234105.
- (91) Burkatzki, M.; Filippi, C.; Dolg, M. *J. Chem. Phys.* **2008**, *129*, 164115.
- (92) Sharifzadeh, S.; Biller, A.; Kronik, L.; Neaton, J. B. *Phys. Rev. B* **2012**, *85*, 125307.
- (93) Baumeier, B.; Andrienko, D.; Ma, Y.; Rohlfing, M. *J. Chem. Theory Comput.* **2012**, *8*, 997–1002.
- (94) Blase, X.; Attaccalite, C.; Olevano, V. *Phys. Rev. B* **2011**, *83*, 115103.
- (95) Hogan, C.; Palumbo, M.; Gierschner, J.; Rubio, A. *J. Chem. Phys.* **2013**, *138*, 01B607.
- (96) Albrecht, S.; Reining, L.; Del Sole, R.; Onida, G. *Phys. Rev. Lett.* **1998**, *80*, 4510.
- (97) Bruneval, F.; Rangel, T.; Hamed, S. M.; Shao, M.; Yang, C.; Neaton, J. B. *Comput. Phys. Commun.* **2016**, *208*, 149–161.
- (98) Weigend, F. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285–4291.
- (99) Giannozzi, P. et al. *J. Phys.: Condens. Matter* **2009**, *21*, 395502.

- (100) Rozzi, C. A.; Varsano, D.; Marini, A.; Gross, E. K.; Rubio, A. *Phys. Rev. B* **2006**, *73*, 205119.
- (101) Godby, R.; Needs, R. *Phys. Rev. Lett.* **1989**, *62*, 1169.
- (102) Werner, H.-J. et al. MOLPRO, version 2015.1, a package of ab initio programs. 2015.
- (103) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.
- (104) Hiyama, M.; Akiyama, H.; Wang, Y.; Koga, N. *Chem. Phys. Lett.* **2013**, *577*, 121–126.
- (105) Yin, H.; Ma, Y.; Mu, J.; Liu, C.; Rohlfing, M. *Phys. Rev. Lett.* **2014**, *112*, 228301.
- (106) Faber, C.; Boulanger, P.; Duchemin, I.; Attaccalite, C.; Blase, X. *J. Chem. Phys.* **2013**, *139*, 194308.