RESEARCH ARTICLE `OPEN ACCESS`

A Polarizable CASSCF/MM Approach Using the Interface Between OpenMMPol Library and Cfour

Tommaso Nottoli | Mattia Bondanza | Filippo Lipparini | Benedetta Mennucci

Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Pisa, Italy

Correspondence: Tommaso Nottoli (tommaso.nottoli@dcci.unipi.it) | Benedetta Mennucci (benedetta.mennucci@unipi.it)

ABSTRACT

We present a polarizable embedding quantum mechanics/molecular mechanics (QM/MM) framework for ground- and excited-state Complete Active Space Self-Consistent Field (CASSCF) calculations on molecules within complex environments, such as biological systems. These environments are modeled using the AMOEBA polarizable force field. This approach is implemented by integrating the OpenMMPol library with the CFour quantum chemistry software suite. The implementation supports both single-point energy evaluations and geometry optimizations, facilitated by the availability of analytical gradients. We demonstrate the methodology by applying it to two distinct photoreceptors, exploring the impact of the protein environment on the structural and photophysical properties of their embedded chromophores.

1 | Introduction

Multiscale approaches that couple quantum mechanical (QM) descriptions with classical models have become a widely used strategy for investigating properties and processes of molecular systems in increasingly complex environments. Among these hybrid QM-classical methods, a particularly effective approach is the use of atomistic models based on molecular mechanics (MM) force fields for the classical component [1–6]. The success of the resulting QM/MM approach is largely attributed to its versatility. Any embedded system can, in principle, be treated at the QM/MM level, given the availability of an appropriate MM force field for the classical region and a model that accurately describes QM–MM interactions. The nature of these QM–MM interactions distinguishes the various formulations of the QM/MM approach. Currently, two primary approaches are widely used: the standard electrostatic embedding and the more sophisticated polarizable embedding. The key difference between these methods is that polarizable embedding accounts for mutual polarization between the QM and MM regions, in addition to the electrostatic interactions. This is expected to play a significant role, especially in processes where changes in the QM charge density are important such as in reactivity and light-induced phenomena [7, 8].

Polarizable embedding QM/MM models are available in various formulations, where the polarization of the MM component is represented using methods such as fluctuating charges [9–12], Drude oscillators [13–16], or induced point dipoles [7, 17–26]. The growing adoption of these models has been facilitated by the integration of these formulations into electronic structure codes or through interfaces between these codes and external libraries [27–30]. Notably, the latter strategy has proven highly effective, as it enables a straightforward coupling of the polarizable embedding model with different QM codes and levels of theory.

Recently, our group has presented the OpenMMPol, an open-source library for coupling QM codes with induced point dipole (IPD)-based polarizable MM models [30]. OpenMMPol offers a

modular implementation of all the functionalities to perform polarizable embedding calculations using either the standard IPD force-field (often referred to as MMPol) [31] or the more involved AMOEBA [32] force-field. Thanks to a simple but flexible multi-language interface, OpenMMPol can easily work with most existing QM software; the library also implements all the purely MM terms of the forcefield, including a simple and straightforward way to handle link-atoms and all the energy (and geometrical gradients) for the purely MM terms required by AMOEBA and AMBER forcefields. Thanks to those features, OpenMMPol is an appealing plug-and-play solution that allows, after a minimal interfacing effort, to perform complex QM/MM calculations on a broad variety of real-life systems without any external software required.

Here, the OpenMMPol library is interfaced with CFOUR [33], a suite of QM programs that specialize in highly accurate post-Hartree Fock calculations, including a large manifold of methods from the Coupled Cluster hierarchy, but also the Complete Active Space – Self-consistent field (CASSCF) method [34–36]. The latter can be used for both ground- and excited-state calculations [37, 38] and, thanks to a recent implementation [39, 40] based on the Cholesky Decomposition of the two-electron integrals [41–44], it can be applied to large molecular systems.

CASSCF methods have been previously integrated with polarizable force fields [45–48, 49]. In this work, the CFOUR-OpenMMPol interface enables an efficient implementation of the CASSCF/AMOEBA approach, allowing for the calculation of ground and excited state energies in both State-Average and State-Specific formulations. Additionally, this implementation supports geometry optimizations using analytical gradients.

The developed software was used to investigate two photoresponsive proteins (see Figure 1). The first system is the Dronpa variant of green fluorescent protein (GFP), in which the chromophore spontaneously forms from the condensation of a tyrosine residue with adjacent glycine and cysteine. The second system is the orange carotenoid protein (OCP), where the chromophore is canthaxanthin (CAN). In both systems, the chromophores are involved in significant hydrogen-bonding interactions with nearby residues. However, in Dronpa, the chromophore is covalently bonded to the protein matrix and carries a charge, whereas the CAN chromophore in OCP is noncovalently bound and is neutral. These characteristics make the two systems valuable test cases for the CASSCF/AMOEBA approach. They allow for the investigation of both specific hydrogen-bonding interactions and nonspecific electrostatic and polarization effects exerted by the protein on the chromophore geometry and response to light. To reach a deeper understanding of these effects, we also compare the AMOEBA force-field with an electrostatic embedding model.

2 | Method and Implementation

A convenient way of deriving a generic polarizable embedding model coupled to a QM method consists in building a Lagrangian [50]. Adopting the AMOEBA force field, where a fixed multipolar distribution (up to quadrupoles) is used in combination with induced dipole moments, the Lagrangian reads [26]:

$$\mathcal{L}(\kappa, c, \mu^d, \mu^p) = \mathcal{E}^{\text{CAS}}(\kappa, c) + \mathcal{E}^{\text{self}}(M) + \mathcal{E}^{\text{ele}}(\kappa, c, M)$$
$$+ \mathcal{E}^{\text{pol}}(\kappa, c, M) + \frac{1}{2}\langle \mu^p, \mathbf{T}\mu^d - \mathbf{E}(\kappa, c) - \mathbf{E}^d(M)\rangle$$

$$(1)$$

where the first term, $\mathcal{E}^{\text{CAS}}(\kappa, c)$ is the CASSCF energy and the other four terms represent the QM-AMOEBA specific terms. We recall that AMOEBA, due to different screening rules used in the polarization step and in energy calculations, introduces two independent sets of induced dipoles that are called 'polarization' (μ^p) and 'direct' (μ^d) [32].
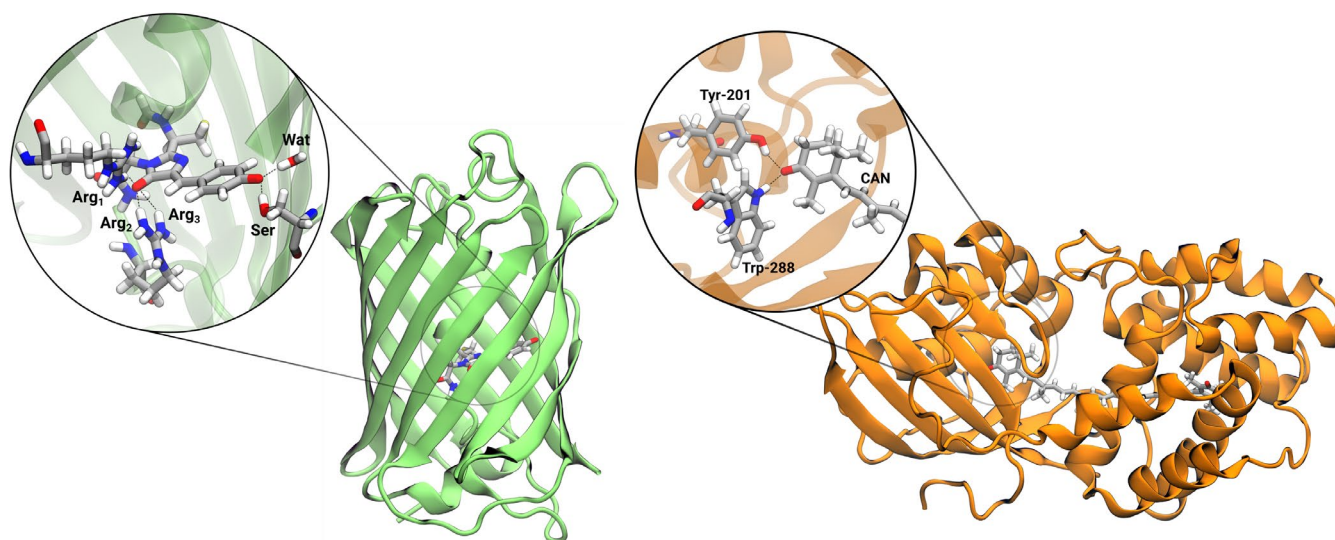

FIGURE 1 | Representation of Dronpa (left) and OCP (right) structures. The insets show the hydrogen bond networks of each chromophore in its binding pocket: Serin-142 (Ser), Arginine-91 (Arg$_1$), two hydrogen atoms of Arginine-66 (Arg$_2$ and Arg$_3$), and a water molecule (Wat) for Dronpa and Tyrosine-201 (Tyr-201) and Tryptophan-288 (Trp-288) for OCP.

$\mathcal{E}^{\mathrm{CAS}}(\kappa, c)$ only depends on the QM parameters, *i. e.* orbital rotation (κ) and configurational coefficients (c) and it can be written in terms of one- and two-body reduced density matrices and integrals

$$\mathcal{E}^{\mathrm{CAS}} = \sum_i \left(h_{ii} + F_{ii}^I \right) + \sum_{uv} \gamma_{uv} F_{uv}^I + \frac{1}{2} \sum_{uvxy} \Gamma_{uvxy} (uv|xy) + \mathcal{E}^{\mathrm{nuc}} \tag{2}$$

where h_{ii} are conventional one-electron integrals, $(uv|xy)$ are two-electron integrals written using Mulliken convention, γ_{uv} and Γ_{uvxy} are the one- and two-body reduced density matrices, respectively, $\mathcal{E}^{\mathrm{nuc}}$ is the nuclear repulsion energy, and $\mathbf{F}^I$ is the inactive Fock matrix.

In Equation (1)

$$\mathcal{E}^{\mathrm{self}}(M) = \frac{1}{2} \sum_{k \neq l=1}^{N_{\mathrm{MM}}} s_{kl}^m \sum_{L,L'=0}^{2} M_k^L T_{kl}^{LL'} M_l^{L'} \tag{3}$$

is the self-interaction energy of the AMOEBA multipolar distribution (M^L), where $L = 0, 1, 2$ for charges, dipoles, and quadrupoles, respectively. s_{kl}^m is a screening factor used to remove unwanted interactions and

$$T_{kl}^{LL'} = \frac{\partial^L}{\partial r_k^L} \frac{\partial^{L'}}{\partial r_l^{L'}} \frac{1}{|r_k - r_l|} \tag{4}$$

is the generalized Coulomb kernel [51].

The second and third terms of Equation (1) are the electrostatic and polarization interactions between the QM and the MM components. The electrostatic interaction energy, $\mathcal{E}^{\mathrm{ele}}$, between the MM multipoles and the QM charge density, can be written as

$$\mathcal{E}^{\mathrm{ele}}(\kappa, c, M) = \sum_k \sum_L M_k^L \left(\sum_\alpha T_{k\alpha}^{L0} Z_\alpha - \sum_{pq} \gamma_{pq} \langle \phi_p | \hat{t}_k^L | \phi_q \rangle \right) \tag{5}$$

where ϕ_p is a molecular orbital, and

$$\hat{t}_k^L = \frac{\partial^L}{\partial r_k^L} \frac{1}{|r_k - r|} \tag{6}$$

The polarization energy can be written as the scalar product between the induced dipoles (μ^d) and the sum of the QM electric field ($\mathbf{E}(\kappa, c)$) and the *polarization* field due to the multipolar distribution

$$\mathcal{E}^{\mathrm{pol}}(\kappa, c, M) = -\frac{1}{2} \langle \mu^d, \mathbf{E}(\kappa, c) + \mathbf{E}^p(M) \rangle \tag{7}$$

$$E_k^p(M) = -\left(\sum_l s_{kl}^p \sum_L \mathcal{T}_{kl}^{1L} M_l \right) \tag{8}$$

with

$$\mathcal{T}_{kl}^{LL'} = \frac{\partial^L}{\partial r_k^L} \frac{\partial^{L'}}{\partial r_l^{L'}} \frac{\lambda(|r_k - r_l|)}{|r_k - r_l|} \tag{9}$$

as the damped Coulomb kernel, which is used to avoid the so-called polarization catastrophe [31, 51, 52].

The coupled CASSCF/AMOEBA equations can be obtained by imposing the stationarity conditions on the Lagrangian in Equation (1).

By differentiating with respect to κ_{pq}, we get the CASSCF orbital gradient

$$g_{pq} = 2\left(F_{pq} - F_{qp} \right) \tag{10}$$

where F is the generalized Fock matrix which contains a modified inactive Fock matrix

$$\overline{F}_{pq}^I = F_{pq}^I - \sum_k^{N_{\mathrm{MM}}} \sum_L M_k^L \langle \phi_p | \hat{t}_k^L | \phi_q \rangle - \sum_k^{N_{\mathrm{MM}}} \frac{\mu_k^d + \mu_k^p}{2} \langle \phi_p | \hat{t}_k^1 | \phi_q \rangle \tag{11}$$

which includes the effects of the environment as additional one-electron contributions. Here we exploited a nested, first-order, algorithm to get the optimal CASSCF parameters. Therefore, the full CI problem is solved at each orbital optimization step up to a numerical threshold, which depends on the norm of the orbital gradient. In particular, we used the Super-CI algorithm as the orbital optimization solver [34, 53–55].

Differentiation of Equation (1) with respect to μ^d and μ^p leads, respectively, to the following polarization equations that need to be solved at each CASSCF step

$$\mathbf{T}\mu^d = \mathbf{E}(\kappa, c) + \mathbf{E}^d(M) \tag{12}$$

$$\mathbf{T}\mu^p = \mathbf{E}(\kappa, c) + \mathbf{E}^p(M) \tag{13}$$

where the kl–th block of matrix $\mathbf{T}$ is defined as follows

$$T_{kl} = \alpha_k^{-1} \delta_{kl} + (1 - \delta_{kl}) \mathcal{T}_{kl}^{11} \tag{14}$$

and α_k is the polarizability of the k-th MM atom. Here $\mathbf{E}^d(M)$ is the *direct* field

$$E_k^d(M) = -\left(\sum_l s_{kl}^d \sum_L \mathcal{T}_{kl}^{1L} M_l \right) \tag{15}$$

which differs from the polarization field reported in Equation (8) because the screening factors are different.

The OpenMMPol-CFOUR interface works as follows. The electrostatic contribution to the Fock matrix (second term of Equation (11)) is computed at the beginning by contracting the appropriate integrals in the AO basis with the multipoles, which can be accessed by the OpenMMPol object. At each CASSCF optimization step, the one-body reduced density matrix is used to compute the electronic component of the electric field and passed as input to the OpenMMPol subroutine `set_external_field`, which in turn solves the polarization equations.

Successively, the induced dipoles are used to compute the third term of Equation (11).

Finally, the geometrical gradients can be obtained by differentiating Equation (1) with respect to the atomic coordinates. Moreover, only explicit derivatives need to be computed for all the contributions. We consider here only the gradient with respect to QM coordinates (r_α) and for a single CASSCF state. Differentiation of the first term of Equation (1) leads to the standard CASSCF molecular gradient. The other contributions stem from the electrostatic interaction

$$\frac{\partial \mathcal{E}^{\mathrm{ele}}}{\partial r_\alpha} = \sum_k \sum_L M_k^L \left(T_{k\alpha}^{L1} Z_\alpha - \sum_{pq} \gamma_{pq} \frac{\partial \langle \phi_p | \hat{t}_k^L | \phi_q \rangle}{\partial r_\alpha} \right) \quad (16)$$

and from the polarization energy

$$\frac{\partial \mathcal{E}^{\mathrm{pol}}}{\partial r_\alpha} = -\frac{1}{2} \sum_k \left(\mu_k^d + \mu_k^p \right) \frac{\partial E_k(\kappa, c)}{\partial r_\alpha} \quad (17)$$

The OpenMMPol-CFour interface can also be used for computing excitation energies. Our method is based on state-averaged CASSCF and allows for the description of several excited states. During the CASSCF procedure, the environment polarization is converged to the state-averaged density of the QM system [47]. In order to recover a state-specific polarization of the environment for each excitation, a correction is introduced following the approach proposed by Li et al. [45]. At the convergence of the state-average calculation, for the two states involved in the excitation, we compute the induced dipoles using the corresponding state density (γ^X) and we recompute the electrostatic and polarization terms of the energy

$$\mathcal{E}_{\mathrm{corr}}^X = \mathcal{E}^{\mathrm{ele-pol}}(\gamma^X, \mu(\gamma^X)) - \mathcal{E}^{\mathrm{ele-pol}}(\gamma^X, \mu(\overline{\gamma})) \quad (18)$$

Finally, we correct the state-averaged excitation energy ($\Delta \mathcal{E}_{\mathrm{SA}}^{IF}$) to get the state-specific corrected one as it follows:

$$\Delta \mathcal{E}_{\mathrm{SS}}^{IF} = \Delta \mathcal{E}_{\mathrm{SA}}^{IF} + \mathcal{E}_{\mathrm{corr}}^{IF} \quad (19)$$

$$\Delta \mathcal{E}_{\mathrm{corr}}^{IF} = \mathcal{E}_{\mathrm{corr}}^F - \mathcal{E}_{\mathrm{corr}}^I \quad (20)$$

where I and F indicate the initial and the final state involved in the excitation, respectively.

3 | Two Applications

In this section, we report the results of the CASSCF/AMOEBA description obtained via the CFOUR-OpenMMPol interface to study the impact of the protein environment on the geometry and excitation energies of the two chromophores within Dronpa and OCP. To reach a deeper understanding of these effects, we compare the AMOEBA force-field with the commonly utilized electrostatic embedding model (specifically, AMBER99SB force-field), where the environment is described using the Amber force field. Furthermore, we assess the performance of CASSCF/AMOEBA relative to the (TD)DFT level of theory.

The starting structures for the optimization procedure, are taken from previous works on OCP [56] and Dronpa [57] by our group and have been prepared similarly: the crystal structure (2Z1O for Dronpa [58] and 4XB5 for OCP [59]) was inserted in a solvent box containing water molecules and enough Na$^+$ and Cl$^-$ ions to obtain a neutral system and a salt concentration of 0.1 M. The details of the equilibration procedures are explained in the reference papers. In our optimizations, only solvent molecules within 5 Å distance from the protein are included.

All CASSCF calculations were performed using unrestricted natural orbitals (UNO) as the initial guess. In addition to being well suited for the CASSCF optimization algorithm, UNOs allow for automatic determination of the active space based on natural occupation numbers [60, 61]. The optimal strategy is to include all orbitals with occupation numbers (ON) between 0.01 and 1.99. However, if this results in an excessively large active space, it is reduced to ensure that the calculation remains feasible. In both systems, the QM region includes only the chromophore. All CASSCF geometry optimizations were performed with the MM atoms kept frozen and the 6-31G(d) basis set [62], whereas the 6-31+G(d) was employed for excited-state calculations. Various active spaces were tested (see SI), but the results presented here are based on a (12,12) space for Dronpa and a (8,12) space for OCP. Specifically, for Dronpa we included all orbitals as provided by the UNO criterion plus additional π ones, while for OCP we considered as active the orbitals with ON between 0.04 and 1.93. For the SA–CASSCF calculations, we considered two states for Dronpa, with the exception of the Amber calculation for which three states were averaged to improve the convergence. On the other hand, three states were always considered for OCP. For the (TD)DFT calculations, the B3LYP exchange correlation functional [63] in conjunction with the 6-31G(d) basis set was used for geometry optimization, while the CAM–B3LYP functional [64] was applied, using the 6-31+G(d) basis, to compute vertical excitations.

3.1 | Dronpa

The chromophore of Dronpa can be described in terms of two resonance structures, depending on the formal position of the negative charge. In the benzenoid form, depicted in Figure 2, the charge is primarily localized on the oxygen atom attached to the benzene ring, while in the quinoid form, the charge resides on the oxygen atom of the imidazole ring. To investigate the effect of the protein in tuning charge distribution thus affecting the weight of the two limiting resonance structures, we use the bond length alternation (BLA) parameter. BLA is defined as the difference between the average lengths of the single and double bonds highlighted in green in Figure 2. A positive BLA value indicates a predominance of the benzenoid form, while a negative BLA signifies the quinoid form.

In Table 1, we report the values of the BLA computed for the isolated chromophore (vacuum) and in the protein with the two selected force fields. For the isolated chromophore, the optimization was performed with the dihedral angle that connects the nitrogen of the imidazole ring to the nitrogen of the
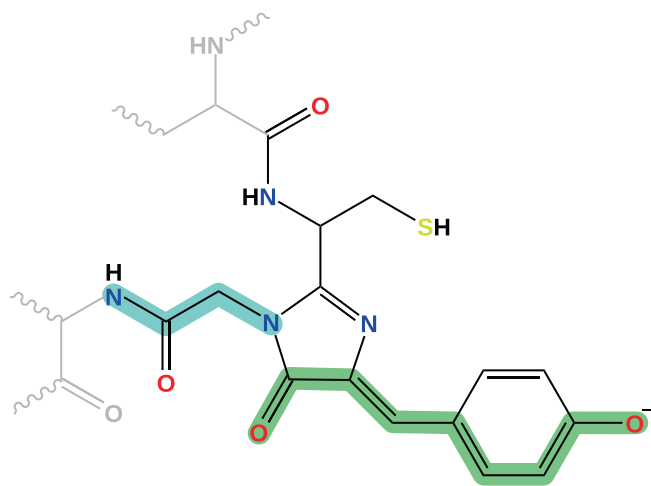
FIGURE 2 | Chemical representation of the Dronpa chromophore. In gray we show the residues to which it is covalently bonded, which are treated at the MM level either using the Amber or AMOEBA force field. The atoms considered to compute the BLA are highlighted in green, whereas those that were kept frozen during the optimizations performed in vacuum are highlighted in cyan.

TABLE 1 | DFT and CASSCF BLA values (in Å) and dihedral angles (in degrees) for Dronpa's chromophore computed in vacuum and in protein with the two force fields.

Method	Vacuum	Amber	AMOEBA
BLA			
B3LYP	0.014	0.004	0.016
CAS(12,12)	0.041	0.010	−0.020
Dihedral angle			
B3LYP	179	174	175
CAS(12,12)	180	174	174

condensed glycine (see atoms highlighted in cyan in Figure 2) frozen to prevent the formation of a hydrogen bond between the oxygen of the imidazole ring and the amide hydrogen. The dihedral angle was fixed to the value found in the protein cavity.

CASSCF and DFT give a qualitatively similar description for both the isolated chromophore and the chromophore embedded in an electrostatic embedding (Amber). In contrast, the two QM methods show qualitative differences when using the polarizable AMOEBA force field. Specifically, DFT/AMOEBA yields a value similar to the vacuum case, while CASSCF/AMOEBA gives a negative BLA indicative a quinoid-like character.

Additional insights can be gained by examining the dihedral angle around the single bond adjacent to the benzene ring, as shown in Table 1. The dihedral angle is smaller in protein than in vacuum, where the molecule adopts a planar conformation, but it shows little sensitivity to the environmental model used. For both force fields, the dihedral angle remains

nearly identical (around 174°). This tiny distortion of the chromophore appears to be primarily driven by the steric effects of the protein pocket. The discrepancy between the embedding models in describing the BLA, cannot therefore be attributed to the distortion of the dihedral angle, but is driven by local interactions, such as hydrogen bonds, present within the protein. The two oxygen atoms at the extremities of the conjugated chain considered in the BLA calculation in fact engage in multiple hydrogen bonds with nearby residues. The hydrogen bond network within the protein cavity is illustrated in the inset of Figure 1.

Table 2 presents the lengths of the key hydrogen-bonds between the oxygen atom of the benzene ring and Serine-142 (Ser), as well as a water molecule (Wat), and between the oxygen atom of the imidazole ring and Arginine-91 (Arg$_1$), alongside two hydrogen atoms of Arginine-66 (Arg$_2$ and Arg$_3$). Moving from the electrostatic to the polarizable embedding results a noticeable increase in the bond lengths is observed for both DFT and CASSCF methods. Specifically, CASSCF exacerbates these differences, with the most significant change being a 0.165 Å increase in the hydrogen bond length with Arg$_3$ when switching from Amber to AMOEBA. This different description of H-bonds may explain the reversed BLA signs between Amber and AMOEBA computed at the CASSCF level.

The effect of the embedding model has also been assessed by computing excitation energies. In order to disentangle the impact of the geometry, for AMOEBA calculations of excitation energies, we performed two calculations using Amber (AMOEBA@Amber) and AMOEBA (AMOEBA@AMOEBA) optimized geometries.

Table 3 presents the first excitation energy calculated at the SA–CAS(12,12)/6-31+G(d) and TDCAM–B3LYP/6-31+G(d) level on CASSCF geometries for the two embedding models and in vacuum. For CAS/AMOEBA calculations we report also the state-specific corrected energies ($\Delta\mathcal{E}_{SS}$). To have a more correct comparison with $\Delta\mathcal{E}_{SS}$ for TDDFT/AMOEBA calculations we report the values obtained within a State-Specific approximation generally indicated as "corrected linear response" cLR [23].

Starting from the CAS values, we observe that by keeping fixed the model used to compute the excitation energy, the effect of a different geometry is not negligible: $\Delta\mathcal{E}_{SA}$ of AMOEBA@AMOEBA is red-shifted by 0.13 eV to the AMOEBA@Amber case. However, this difference is reduced to 0.07 eV when the state-specific correction is added to the AMOEBA calculations. Instead, by keeping constant the Amber geometry and changing the model for the energy calculations, we note that AMOEBA and Amber excitation energies differ by 0.17 eV. Comparing Vacuum, Amber, and AMOEBA@AMOEBA, we observe a systematic increase in the excitation energy moving from the isolated molecule to the protein with an overall blue shift of 0.17 eV for AMOEBA of which 0.1 eV is its difference to Amber. If we do not include the state-specific correction, the latter difference almost disappears making the two embedding models very similar. All these data clearly show that including an SS correction for the polarizable model is important. In this case, it can also be justified by a somewhat large

TABLE 2 | DFT and CASSCF values of the H-bond lengths between Dronpa's chromophore and the residues/water computed with the two force fields. All values are in Å.

Method	Ser		Wat		Arg$_1$		Arg$_2$		Arg$_3$	
	Amber	AMOEBA	Amber	AMOEBA	Amber	AMOEBA	Amber	AMOEBA	Amber	AMOEBA
B3LYP	1.821	1.857	1.814	1.819	1.932	1.934	2.433	2.524	1.826	1.907
CAS(12,12)	1.751	1.787	1.742	1.828	1.898	1.970	2.455	2.508	1.787	1.952

TABLE 3 | Excitation energies (in eV) of the first excited state of Dronpa's chromophore computed using SA–CAS(12,12) and TDCAM–B3LYP level on top of CAS geometries. For CAS/AMOEBA calculations we report also the state-specific corrected energies ($\Delta\mathcal{E}_{SS}$). For TDDFT calculations we also report the excitation energies calculated using DFT optimized geometries (TDDFT/DFT). The notation AMOEBA@Amber and AMOEBA@AMOEBA means that the excitation energy has been computed with QM/AMOEBA at the geometry optimized with QM/Amber and QM/AMOEBA model, respectively.

Method	SA–CAS/CAS		TDDFT/CAS	TDDFT/DFT
	$\Delta\mathcal{E}_{SA}$	$\Delta\mathcal{E}_{SS}$	$\Delta\mathcal{E}$	$\Delta\mathcal{E}$
Vacuum	3.41		3.15	3.04
Amber	3.48		3.13	3.22
AMOEBA@Amber	3.63	3.65	3.22	3.30
AMOEBA@AMOEBA	3.50	3.58	3.24	3.16

difference between the ground- and excited-state dipole, 17 and 11 Debye, respectively.

Moving to the energies computed at the TDDFT level on the same CAS geometries and on the DFT geometries, both qualitative and quantitative variations are observed. The excitation energy is red-shifted relative to the CAS value for the isolated molecule using the same geometry, with an additional red-shift observed when using DFT geometry. Considering environmental effects, further differences emerge. The behavior we obtain at TDDFT when moving from vacuum to AMOEBA is qualitatively similar to what found at CAS level if we consider consistent calculations (e.g., AMOEBA@AMOEBA): for both CAS and DFT geometries, we observe a blue-shift of about 0.1 eV. In contrast, the Amber-AMOEBA shift strongly depends on the selected geometry and shows a blue-shift for CAS geometries and a red-shift for DFT geometries. All these data show a larger sensitivity of the TDDFT description to geometries.

3.2 | OCP

The analysis presented for Dronpa is here repeated for OCP.

As before we start the analysis by presenting the calculated values of the BLA parameter for CAN. In this case, we included all atoms involved in the conjugated chain of CAN, as depicted in Figure 3.

The results obtained at the CAS and DFT level of theories and with the different embedding models are reported in Table 4.

At the DFT level, both force fields exhibit nearly identical BLA. However, CASSCF results show a modest increase in BLA when using AMOEBA. Overall, CASSCF tends to overestimate the BLA compared to DFT – which have been shown to be consistent with higher levels of theory [65]; this effect can be attributed to the absence of dynamic correlation in CASSCF – a critical
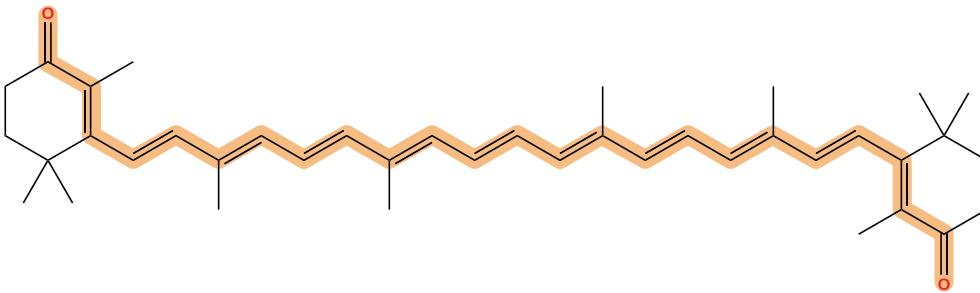
FIGURE 3 | Chemical representation of the CAN chromophore. The atoms considered to compute the BLA are highlighted.

TABLE 4 | CAS and DFT BLA values (in Å) for CAN computed in vacuum and in protein with the two force fields.

Method	Vacuum	Amber	AMOEBA
B3LYP	0.110	0.101	0.102
CAS(8,12)	0.151	0.147	0.149

TABLE 5 | DFT and CASSCF values of the H-bond lengths between OCP's chromophore and the residues computed with the two force fields. All values are in Å.

Method	Trp-288		Tyr-201	
	Amber	AMOEBA	Amber	AMOEBA
B3LYP	1.853	1.916	1.815	1.888
CAS(8,12)	1.879	1.925	1.869	1.924

factor in the geometry optimization of conjugated molecules [65, 66].

As reported in the Introduction, CAN is engaged in two hydrogen bonds with two adjacent residues, namely a tryptophan (Trp-288) and a tyrosine (Tyr-201). In Table 5 we report the corresponding bond lengths as obtained through CAS and DFT optimizations using either Amber or AMOEBA. As already found for Dronpa, we observe a general increase in the distances when moving from a non-polarizable to a polarizable force field, both for DFT and CASSCF data. The hydrogen bonds computed at the DFT and CASSCF level formed with Trp-288 reasonably agree. On the other hand, we note a somewhat higher difference of up to 0.05 Å for Tyr-201.

Moreover, the directional nature of the hydrogen bond network can be assessed by looking at the angle formed between the oxygen atom of CAN and the hydrogen atoms of Trp-288 and Tyr-201. Here, the two QM levels of theory provide almost the same result, 90° with Amber and 88° for AMOEBA.

In Table 6 we compare the excitation energy to the lowest bright state of CAN in OCP calculated using CASSCF and TDDFT, the latter performed either on top CASSCF optimized geometries or DFT optimized ones.

The CASSCF-computed excitation energies are consistently blue-shifted relative to the TDDFT results, which is in line

TABLE 6 | Excitation energies (in eV) of the first excited state of OCP's chromophore computed using SA–CAS(8,12)/6-31+G(d) and TDCAM-B3LYP/6-31+G(d) level on top of CASSCF(8,12)/6-31G(d) geometries. For CAS/AMOEBA calculations we report also the state-specific corrected energies ($\Delta\mathcal{E}_{SS}$). For TDDFT calculations we also report the excitation energies calculated using B3LYP/6-31G(d) optimized energies (TDDFT/DFT). The notation AMOEBA@Amber and AMOEBA@AMOEBA means that the excitation energy has been computed with the QM/AMOEBA at the geometry optimized with the QM/Amber and QM/AMOEBA model, respectively.

Method	SA–CAS/CAS		TDDFT/ CAS	TDDFT/ DFT
	$\Delta\mathcal{E}_{SA}$	$\Delta\mathcal{E}_{SS}$	$\Delta\mathcal{E}$	$\Delta\mathcal{E}$
Vacuum	4.52		3.03	2.26
Amber	4.24		2.73	2.30
AMOEBA@ Amber	4.30	4.35	2.89	2.31
AMOEBA@ AMOEBA	4.50	4.59	2.93	2.36

with expectations [67]. As observed for Dronpa, the impact of using different geometries for CASSCF excitation energies and the Amber force field for the environment is negligible. In contrast, AMOEBA@Amber and AMOEBA@AMOEBA differ by 0.24 eV, indicating greater sensitivity to molecular geometry with the AMOEBA force field. Comparing consistent calculations (Vacuum, Amber and AMOEBA@AMOEBA), we find that introducing an electrostatic embedding causes a significant red shift (around 0.3 eV) from vacuum, but this shift is fully counteracted by AMOEBA. This behavior is confirmed by the TDDFT/CAS results: a red shift of 0.3 eV is observed with Amber, which is largely reduced by AMOEBA. In contrast, the TDDFT/DFT results show a blue shift for both Amber and AMOEBA. At the TDDFT level, the effect of the geometry on the AMOEBA excitation energies is much smaller than what is found at the CAS level.

4 | Conclusions

We have presented an implementation of a polarizable embedding QM/MM that can be used to perform ground- and excited-state CASSCF calculations of molecules embedded in a polarizable environment described with the AMOEBA force field. The implementation has been achieved by coupling the

recently developed OpenMMPol library [30] and the CFour suite of quantum-mechanical programs. OpenMMPol offers a versatile and user-friendly realization of all the machinery required to run embedded calculations, including an implementation of all the purely MM contributions to the energy (e.g., bonded, van der Waals) and the forces, plus electrostatic drivers that can be used to compute the induced point dipoles and the various QM/MM interaction terms. The coupling with the CASSCF program in CFour, which is well suited for calculations on medium and large systems thanks to an efficient implementation based on the Cholesky Decomposition of the two-electron integrals, opens the way to the description of complex molecular systems embedded in biological matrices, and is particularly useful to obtain a qualitative description of systems that exhibit a marked multireference character in their ground and excited states.

This work opens the door to several future directions that can leverage the many capabilities of CFour to perform highly accurate calculations of molecular energies, structures, and properties. In particular, we plan to extend the present implementation to embedded Coupled-Cluster (CC) theory, including Equation of Motion CC (EOM-CC), and to complete the CASSCF-OpenMMPol interface to allow for linear response calculations of excitation energies, transition moments, and frequency dependent properties. We plan to include the interface described in this paper in the next public release of CFour.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. H. M. Senn and W. Thiel, "QM/MM Methods for Biomolecular Systems," *Angewandte Chemie International Edition* 48, no. 7 (2009): 1198–1229.

2. M. W. van der Kamp and A. J. Mulholland, "Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology," *Biochemistry* 52, no. 16 (2013): 2708–2728.

3. A. Warshel, "Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture)," *Angewandte Chemie International Edition* 53, no. 38 (2014): 10020–10031.

4. E. Brunk and U. Rothlisberger, "Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States," *Chemical Reviews* 115, no. 12 (2015): 6217–6263.

5. S. F. Sousa, A. J. M. Ribeiro, R. P. P. Neves, et al., "Application of Quantum Mechanics/Molecular Mechanics Methods in the Study of Enzymatic Reaction Mechanisms," *Wiley Interdisciplinary Reviews: Computational Molecular Science* 7, no. 2 (2016): e1281–e1289.

6. U. N. Morzan, J. Diego, A. de Armiño, et al., "Spectroscopy in Complex Environments From QM–MM Simulations," *Chemical Reviews* 118, no. 7 (2018): 4071–4113.

7. M. Bondanza, M. Nottoli, L. Cupellini, F. Lipparini, and B. Mennucci, "Polarizable Embedding QM/MM: The Future Gold Standard for Complex (Bio)systems?," *Physical Chemistry Chemical Physics* 22, no. 26 (2020a): 14433–14448.

8. D. Loco, L. Lagardère, O. Adjoua, and J.-P. Piquemal, "Atomistic Polarizable Embeddings: Energy, Dynamics, Spectroscopy, and Reactivity," *Accounts of Chemical Research* 54, no. 13 (2021): 2812–2822.

9. A. K. Rappe and W. A. Goddard, "Charge Equilibration for Molecular Dynamics Simulations," *Journal of Physical Chemistry* 95, no. 8 (1991): 3358–3363.

10. S. W. Rick, S. J. Stuart, and B. J. Berne, "Dynamical Fluctuating Charge Force Fields: Application to Liquid Water," *Journal of Physical Chemistry* 101, no. 7 (1994): 6141–6156.

11. F. Lipparini, C. Cappelli, and V. Barone, "Linear Response Theory and Electronic Transition Energies for a Fully Polarizable QM/Classical Hamiltonian," *Journal of Chemical Theory and Computation* 8, no. 11 (2012): 4153–4165.

12. P. P. Poier and F. Jensen, "Describing Molecular Polarizability by a Bond Capacity Model," *Journal of Chemical Theory and Computation* 15, no. 5 (2019): 3093–3107.

13. E. Boulanger and W. Thiel, "Solvent Boundary Potentials for Hybrid QM/MM Computations Using Classical Drude Oscillators: A Fully Polarizable Model," *Journal of Chemical Theory and Computation* 8, no. 11 (2012): 4527–4538.

14. S. Riahi and C. N. Rowley, "The Charmm–Turbomole Interface for Efficient and Accurate Qm/Mm Molecular Dynamics, Free Energies, and Excited State Properties," *Journal of Computational Chemistry* 35, no. 28 (2014): 2076–2086.

15. L. Zhenyu and Y. Zhang, "Interfacing Ab Initio Quantum Mechanical Method With Classical Drude Osillator Polarizable Model for Molecular Dynamics Simulation of Chemical Reactions," *Journal of Chemical Theory and Computation* 4, no. 8 (2008): 1237–1248.

16. J. A. Lemkul, J. Huang, B. Roux, and A. D. MacKerell, "An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications," *Chemical Reviews* 116, no. 9 (2016): 4983–5013.

17. A. Warshel and M. Levitt, "Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme," *Journal of Molecular Biology* 103 (1976): 227–249.

18. J. Applequist, J. R. Carl, and K.-K. Fung, "Atom Dipole Interaction Model for Molecular Polarizability. Application to Polyatomic Molecules and Determination of Atom Polarizabilities," *Journal of the American Chemical Society* 94, no. 9 (1972): 2952–2960.

19. J. Gao, "Energy Components of Aqueous Solution: Insight From Hybrid QM/MM Simulations Using a Polarizable Solvent Model," *Journal of Computational Chemistry* 18, no. 8 (1997): 1061–1071.

20. P. T. Van Duijnen, and M. Swart, "Molecular and Atomic Polarizabilities: Tholes Model Revisited," *Journal of Physical Chemistry. A* 102, no. 14 (1998): 2399–2407.

21. C. Curutchet, A. Muñoz-Losa, S. Monti, J. Kongsted, G. D. Scholes, and B. Mennucci, "Electronic Energy Transfer in Condensed Phase Studied by a Polarizable QM/MM Model," *Journal of Chemical Theory and Computation* 5, no. 7 (2009): 1838–1848.

22. J. M. H. Olsen and J. Kongsted, "Chapter 3 - Molecular Properties Through Polarizable Embedding," in *In Advances in Quantum*

Chemistry, vol. 61 (Amsterdam, Netherlands: Academic Press, 2011), 107–143, https://doi.org/10.1016/b978-0-12-386013-2.00003-6.

23. D. Loco, É. Polack, S. Caprasecca, et al., "A QM/MM Approach Using the AMOEBA Polarizable Embedding: From Ground State Energies to Electronic Excitations," *Journal of Chemical Theory and Computation* 12, no. 8 (2016): 3654–3661.

24. J. Dziedzic, Y. Mao, Y. Shao, et al., "TINKTEP: A Fully Self-Consistent, Mutually Polarizable QM/MM Approach Based on the AMOEBA Force Field," *Journal of Chemical Physics* 145, no. 12 (2016): 124106.

25. W. Xiaojing, J.-M. Teuler, F. Cailliez, C. Clavaguéra, D. R. Salahub, and A. de la Lande, "Simulating Electron Dynamics in Polarizable Environments," *Journal of Chemical Theory and Computation* 13, no. 9 (2017): 3985–4002.

26. M. Nottoli, M. Bondanza, P. Mazzeo, et al., "QM/AMOEBA Description of Properties and Dynamics of Embedded Molecules," *Wiley Interdisciplinary Reviews: Computational Molecular Science* 13, no. 6 (2023): e1674.

27. J. M. H. Olsen, V. Bolnykh, S. Meloni, et al., "MiMiC: A Novel Framework for Multiscale Modeling in Computational Chemistry," *Journal of Chemical Theory and Computation* 15, no. 6 (2019): 3810–3823.

28. I. A. Kaliman and L. V. Slipchenko, "LIBEFP: A New Parallel Implementation of the Effective Fragment Potential Method as a Portable Software Library," *Journal of Computational Chemistry* 34, no. 26 (2013): 2284–2292.

29. M. Scheurer, P. Reinholdt, E. R. Kjellgren, J. M. H. Olsen, A. Dreuw, and J. Kongsted, "CPPE: An Open-Source C++ and Python Library for Polarizable Embedding," *Journal of Chemical Theory and Computation* 15, no. 11 (2019): 6154–6163.

30. M. Bondanza, T. Nottoli, M. Nottoli, L. Cupellini, F. Lipparini, and B. Mennucci, "The OpenMMPol Library for Polarizable QM/MM Calculations of Properties and Dynamics," *Journal of Chemical Physics* 160, no. 13 (2024, Release availble on): 134106, https://github.com/Molecolab-Pisa/OpenMMPol.

31. J. Wang, P. Cieplak, J. Li, et al., "Development of Polarizable Models for Molecular Mechanical Calculations Ii: Induced Dipole Models Significantly Improve Accuracy of Intermolecular Interaction Energies," *Journal of Physical Chemistry. B* 115 (2011a): 3100.

32. J. W. Ponder, W. Chuanjie, P. Ren, et al., "Current Status of the AMOEBA Polarizable Force Field," *Journal of Physical Chemistry. B* 114, no. 8 (2010): 2549–2564.

33. J. F. Stanton, J. Gauss, L. Cheng, M. E. Harding, D. A. Matthews, and P. G. Szalay, "Coupled-Cluster Techniques for Computational Chemistry: The CFOUR Program Package," *Journal of Chemical Physics* 152, no. 21 (2020): 214108.

34. B. O. Roos, "The Complete Active Space SCF Method in a Fock-Matrix-Based Super-CI Formulation," *International Journal of Quantum Chemistry* 18, no. S14 (1980): 175–189.

35. H.-J. Werner, "Matrix-Formulated Direct Multiconfiguration Self-Consistent Field and Multiconfiguration Reference Configuration-Interaction Methods," *Advances in Chemical Physics* 69 (1987): 1–62, https://doi.org/10.1002/9780470142943.ch1.

36. R. Shepard, "The Multiconfiguration Self-Consistent Field Method," *Advances in Chemical Physics* 69 (1987): 63–200.

37. F. Lipparini and J. Gauss, "Cost-Effective Treatment of Scalar Relativistic Effects for Multireference Systems: A CASSCF Implementation Based on the Spin-Free Dirac-Coulomb Hamiltonian," *Journal of Chemical Theory and Computation* 12 (2016): 4284–4295.

38. F. Lipparini, T. Kirsch, A. Köhn, and J. Gauss, "Internally Contracted Multireference Coupled Cluster Calculations With a Spin-Free Dirac–Coulomb Hamiltonian: Application to the Monoxides of Titanium, Zirconium, and Hafnium," *Journal of Chemical Theory and Computation* 13, no. 7 (2017): 3171–3184.

39. T. Nottoli, J. Gauss, and F. Lipparini, "Second-Order CASSCF Algorithm With the Cholesky Decomposition of the Two-Electron Integrals," *Journal of Chemical Theory and Computation* 17, no. 11 (2021): 6819–6831, https://doi.org/10.1021/acs.jctc.1c00327.

40. T. Nottoli, S. Burger, S. Stopkowicz, J. Gauss, and F. Lipparini, "Computation of NMR Shieldings at the CASSCF Level Using Gauge-Including Atomic Orbitals and Cholesky Decomposition," *Journal of Chemical Physics* 157, no. 8 (2022): 084122.

41. N. H. F. Beebe and J. Linderberg, "Simplifications in the Generation and Transformation of Two-Electron Integrals in Molecular Calculations," *International Journal of Quantum Chemistry* 12, no. 4 (1977): 683–705.

42. I. Røeggen and T. Johansen, "Cholesky Decomposition of the Two-Electron Integral Matrix in Electronic Structure Calculations," *Journal of Chemical Physics* 128, no. 19 (2008): 194107.

43. F. Aquilante, L. Boman, J. Boström, et al., *Cholesky Decomposition Techniques in Electronic Structure Theory* (Dordrecht, Netherlands: Springer, 2011), 301–343.

44. H. Koch, A. Sánchez de Merás, and T. B. Pedersen, "Reduced Scaling in Electronic Structure Calculations Using Cholesky Decompositions," *Journal of Chemical Physics* 118, no. 21 (2003): 9481–9484.

45. Q. Li, B. Mennucci, M. A. Robb, L. Blancafort, and C. Curutchet, "Polarizable QM/MM Multiconfiguration Self-Consistent Field Approach With State-Specific Corrections: Environment Effects on Cytosine Absorption Spectrum," *Journal of Chemical Theory and Computation* 11, no. 4 (2015): 1674–1682.

46. H. Liu, A. J. Jenkins, A. Wildman, et al., "Time-Dependent Complete Active Space Embedded in a Polarizable Force Field," *Journal of Chemical Theory and Computation* 15, no. 3 (2019): 1633–1641.

47. C. Song, "State Averaged CASSCF in AMOEBA Polarizable Water Model for Simulating Nonadiabatic Molecular Dynamics With Nonequilibrium Solvation Effects," *Journal of Chemical Physics* 158, no. 1 (2023a): 014101.

48. C. Song and L.-P. Wang, "A Polarizable QM/MM Model That Combines the State-Averaged CASSCF and AMOEBA Force Field for Photoreactions in Proteins," *Journal of Chemical Theory and Computation* 20, no. 15 (2024): 6632–6651.

49. C. Sepali, L. Goletto, P. Lafiosca, M. Rinaldi, T. Giovannini, and C. Cappelli, "Fully Polarizable Multiconfigurational Self-Consistent Field/Fluctuating Charges Approach," *Journal of Chemical Theory and Computation* 20, no. 22 (2024): 9954–9967.

50. M. Nottoli and F. Lipparini, "General Formulation of Polarizable Embedding Models and of Their Coupling," *Journal of Chemical Physics* 153, no. 22 (2020): 224108.

51. B. T. Thole, "Molecular Polarizabilities Calculated With a Modified Dipole Interaction," *Chemical Physics* 59, no. 3 (1981): 341–350.

52. J. Sala, E. Guardia, and M. Masia, "The Polarizable Point Dipoles Method With Electrostatic Damping: Implementation on a Model System," *Journal of Chemical Physics* 133, no. 23 (2010): 234101.

53. P. E. M. Siegbahn, J. Almlöf, A. Heiberg, and B. O. Roos, "The Complete Active Space SCF (CASSCF) Method in a Newton–Raphson Formulation With Application to the HNO Molecule," *Journal of Chemical Physics* 74, no. 4 (1981): 2384–2396.

54. C. Kollmar, K. Sivalingam, B. Helmich-Paris, C. Angeli, and F. Neese, "A Perturbation-Based Super-CI Approach for the Orbital Optimization of a CASSCF Wave Function," *Journal of Computational Chemistry* 40, no. 14 (2019): 1463–1470.

55. C. Angeli, S. Evangelisti, R. Cimiraglia, and D. Maynau, "A Novel Perturbation-Based Complete Active Space–Self-Consistent-Field Algorithm: Application to the Direct Calculation of Localized Orbitals," *Journal of Chemical Physics* 117, no. 23 (2002): 10525–10533.

56. M. Bondanza, L. Cupellini, F. Lipparini, and B. Mennucci, "The Multiple Roles of the Protein in the Photoactivation of Orange Carotenoid Protein," *Chem* 6, no. 1 (2020b): 187–203.

57. R. Nifosì, B. Mennucci, and C. Filippi, "The Key to the Yellow-To-Cyan Tuning in the Green Fluorescent Protein Family Is Polarisation," *Physical Chemistry Chemical Physics* 21, no. 35 (2019): 18988–18998.

58. H. Mizuno, T. K. Mal, M. Wälchli, et al., "Light-Dependent Regulation of Structural Flexibility in a Photochromic Fluorescent Protein," *Proceedings of the National Academy of Sciences USA* 105, no. 27 (2008): 9227–9232.

59. R. L. Leverenz, M. Sutter, A. Wilson, et al., "A 12 Å Carotenoid Translocation in a Photoswitch Associated With Cyanobacterial Photoprotection," *Science* 348, no. 6242 (2015): 1463–1466.

60. P. Pulay and T. P. Hamilton, "UHF Natural Orbitals for Defining and Starting MC-SCF Calculations," *Journal of Chemical Physics* 88, no. 8 (1988): 4926–4933.

61. Z. Tóth and P. Pulay, "Comparison of Methods for Active Orbital Selection in Multiconfigurational Calculations," *Journal of Chemical Theory and Computation* 16, no. 12 (2020): 7328–7341.

62. W. J. Hehre, R. Ditchfield, and J. A. Pople, "Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules," *Journal of Chemical Physics* 56, no. 5 (1972): 2257–2261.

63. A. D. Becke, "Density-Functional Thermochemistry. 3. The Role of Exact Exchange," *Journal of Chemical Physics* 98, no. 7 (1993): 5648–5652.

64. T. Yanai, D. P. Tew, and N. C. Handy, "A New Hybrid Exchange–Correlation Functional Using the Coulomb-Attenuating Method (CAM-B3LYP)," *Chemical Physics Letters* 393, no. 1-3 (2004): 51–57.

65. M. Bondanza, D. Jacquemin, and B. Mennucci, "Excited States of Xanthophylls Revisited: Toward the Simulation of Biologically Relevant Systems," *Journal of Physical Chemistry Letters* 12, no. 28 (2021): 6604–6612.

66. M. J. G. Peach, E. I. Tellgren, P. Sałek, T. Helgaker, and D. J. Tozer, "Structural and Electronic Properties of Polyacetylene and Polyyne From Hybrid and Coulomb-Attenuated Density Functionals," *Journal of Physical Chemistry. A* 111, no. 46 (2007): 11930–11935.

67. B. Helmich-Paris, "Benchmarks for Electronically Excited States With CASSCF Methods," *Journal of Chemical Theory and Computation* 15, no. 7 (2019): 4170–4179.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.