



Universiteit
Leiden
The Netherlands

Ab initio study of the optical properties of green fluorescent protein

Zaccheddu, M.

Citation

Zaccheddu, M. (2008, April 24). *Ab initio study of the optical properties of green fluorescent protein*. Retrieved from <https://hdl.handle.net/1887/12836>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/12836>

Note: To cite this publication please use the final published version (if applicable).

Chapter 3

Chromophore in vacuum

3.1 Chromophore models of GFP

Before studying the neutral (protonated) and anionic (deprotonated) forms of GFP within the protein environment, it is important to understand the performance of our theoretical techniques on simpler models. We therefore begin our investigation of GFP with the computation of the electronic excitations of model chromophores in the gas phase since the lower level of complexity of these systems allow us to push the limits of our theoretical tools and better understand their limitations. Moreover, absorption experiments and calculations using highly-correlated techniques are available for several chromophore models in the gas phase [58–60].

The chromophore models studied in this Chapter are depicted in Fig. 3.1, and can be divided in three groups: The anionic (deprotonated) chromophores (A, B, C), the neutral (protonated) chromophores (D, E, F), and a positively charged chromophore (G). For both the anionic and the neutral case, three models of increasing size are constructed. The geometries of all models are optimized within all-electron DFT with a cc-pVTZ basis and two different functionals, BLYP and B3LYP, using the Gaussian03 code [53]. We always refer to Fig. 3.1 and the its labels when describing the models below.

Anionic models

The anionic *minimal* model (A) is the smallest possible representation of the GFP chromophore with the additional nice feature to posses C_s symmetry, which significantly accelerates most quantum chemical computations. Given its favorable size and symmetry, it is a simple starting point to investigate the performance of quantum Monte Carlo in describing GFP. Even though no experimental characterization has been done on this system, correlated

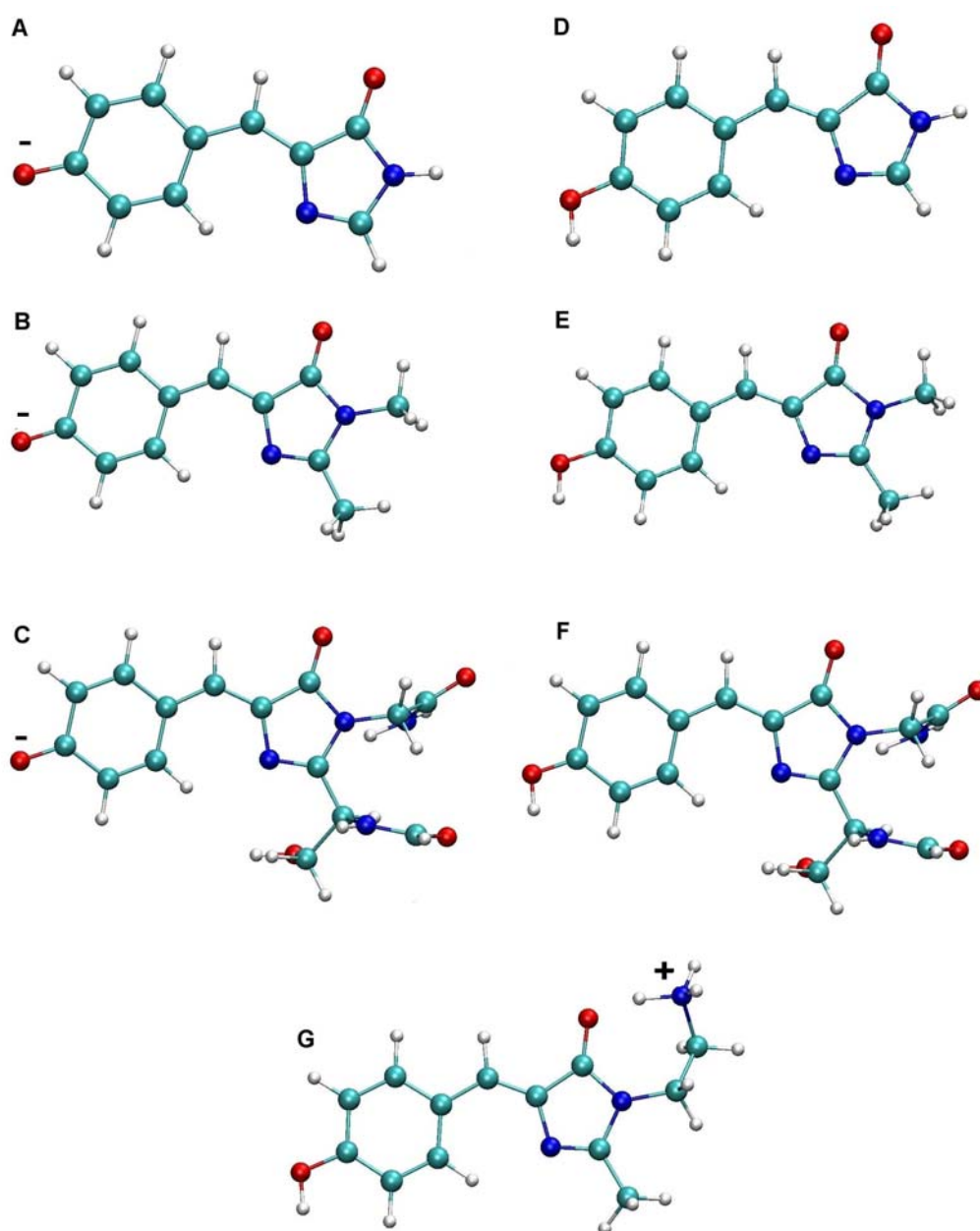


Figure 3.1: Gas-phase chromophore models: The three anionic chromophores in the minimal (A), methyl-terminated (B), and protein-cut (C) models; the three neutral chromophore in the minimal (D), methyl-terminated (E), and protein-cut (F) models; the positively charged (neutral⁺) model (G).

CASPT2 calculations [60] are available to which we can compare.

The anionic methyl-terminated model (B) is only slightly larger than the minimal model (A). Since it only differs in the termination with the two hydrogens substituted with methyl groups, we expect that the electronic properties of models (A) and (B) are rather similar. Even though model (B) has the disadvantage that C_s symmetry is lost and the system has now no symmetry, we construct this chromophore since it was synthesized and investigated in the gas-phase spectroscopic experiments described above. Therefore, for model (B), we can directly compare the calculated excitation energies with experiments, which place its absorption maximum at 2.59 eV [58].

The largest anionic chromophore (C) is the *protein-cut* model with a total of 39 (24 C, N and O) atoms. It corresponds to the chromophore we employ as the QM part of the QM/MM simulations when the protein environment is included for a realistic study of wild-type GFP. The geometry of model (C) is optimized in vacuum and a comparison of the structural and electronic properties of this model with the one in the protein will allow us to understand the geometrical and electrostatic changes induced by the protein environment.

Neutral and neutral⁺ models

For the neutral chromophore, we also construct three models of increasing size, which are analogous to the anionic case, that is the minimal (D), methyl-terminated (E) and protein-cut (F) models. To date, no highly-correlated quantum chemical calculations nor experiments are available for these neutral chromophores in the gas phase. Nevertheless, the study of the neutral chromophores allow us to analyze the changes in electronic properties with respect to the anionic gas-phase chromophores as well as as a function of system size. Moreover, we can investigate the geometrical and electronic impact of the protein environment on the large model (F).

The experimental technique employed for the study of the anionic gas-phase chromophore (B) makes use of an electrostatic ion storage ring and can therefore be applied only to negatively or positively charged molecules. Therefore, not being able to study neutral chromophores, the same experimental group synthesized a positively charged chromophore (G) by attaching to the methyl-terminated model (E) a positively charged NH_3^+ group [59]. We refer to model (G) as the neutral⁺ chromophore. The relevance of this chromophore lies in the claim by the same experimental group that the positively charged group is just a spectator, and that the absorption spectrum of chromophore (G) should therefore be very close to the one of the neutral chromophore. In the photodestruction spectroscopy experiment, the absorp-

tion maximum of model (G) is identified at 2.99 eV and an excitation of 3.11 eV is attributed to the neutral chromophore by correcting the experimental value by the TDDFT/B3LYP difference between the vertical excitation of the (G) and the (E) models. We note that, in the neutral⁺ chromophore, the positively charged group is hydrogen bonded to the oxygen of the imidazole ring.

3.2 Structural analysis of the models

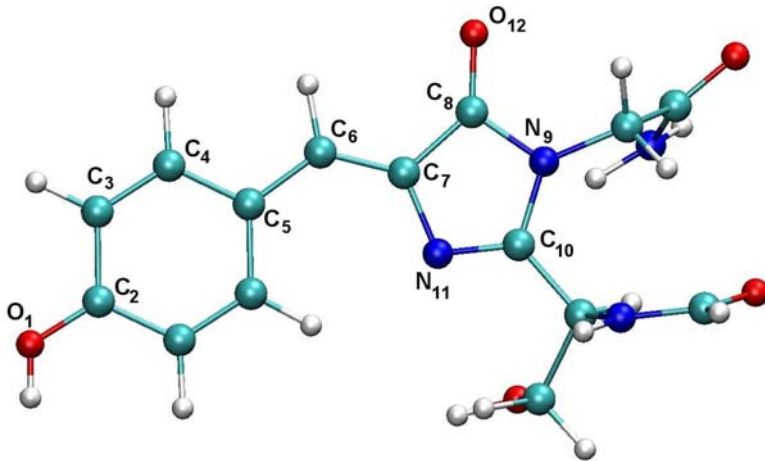


Figure 3.2: Atom numbering used for the chromophore of GFP. The benzene ring is essentially symmetric respect to the C₂-C₅ axis.

The structural features of the chromophore play a fundamental role in determining its excited state properties, and it is therefore important to accurately describe the geometry and how this is affected by the charge state of the chromophore and further modified when the chromophore is embedded in the protein environment. In particular, the degree of bond-length alternation in the conjugate chain running through the chromophore is correlated to the size of the bright $\pi \rightarrow \pi^*$ transition, with a stronger bond-length alternation yielding a larger excitation energy. For π -conjugate linear chains, this correspondence between gap and bond-length alternation is well known as, in going from shorter to longer chains, the energy gap decreases while the double bonds lengthen and the single bonds shorten. The GFP chromophore

is also a π -conjugated system even though the presence of the rings makes it a more complicated system than a linear oligomer.

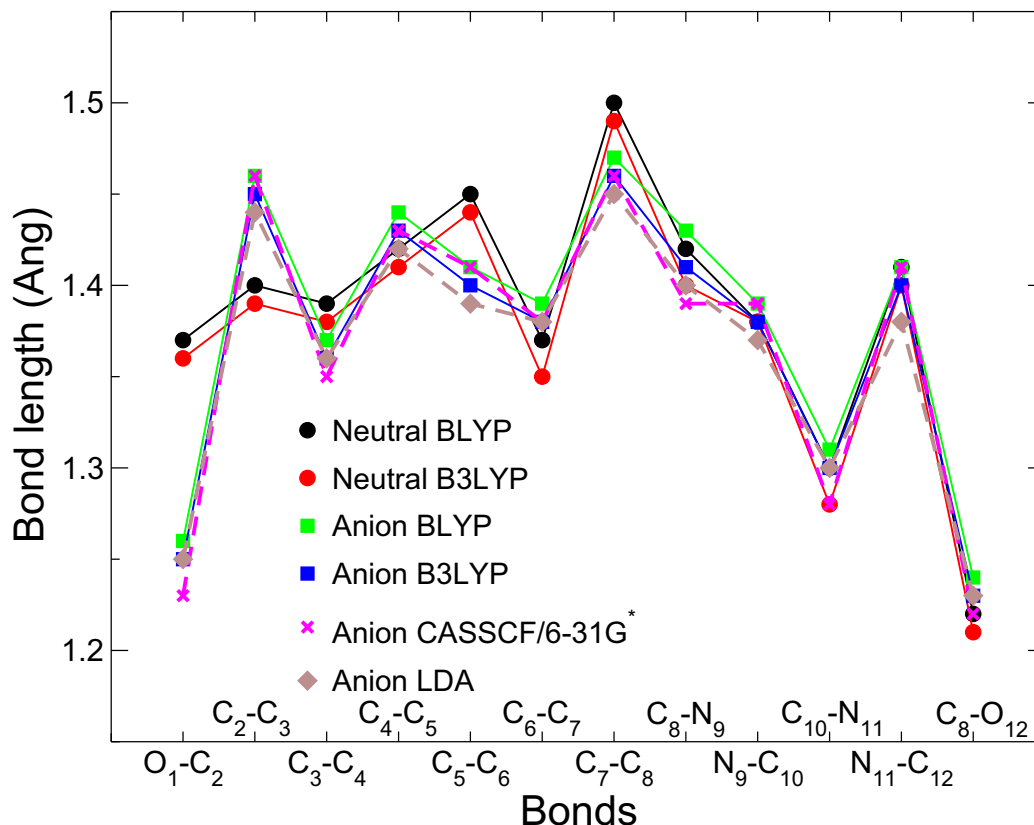


Figure 3.3: Bond lengths of the minimal anionic (A) and neutral (D) models of the GFP chromophore as obtained in a DFT/cc-pVTZ geometry optimization with the BLYP, B3LYP and LDA functionals. The results from a CASSCF/6-31G* calculation [60] are also shown. The bonds of the central carbon bridge are C₅—C₆ and C₆—C₇.

The structures of the chromophores are constructed by relaxing them in the ground state using DFT and various exchange-correlation functionals. Even though all models have been generated with at least the BLYP and B3LYP functionals, we only discuss in detail how the functional influences the geometry of the chromophore for the minimal anionic (A) and neutral (D) models as the conclusions apply equally well to all other models. The bond lengths along the chromophore of the minimal models computed with various functionals are shown in Fig. 3.3. For the atom labelling, we refer the reader to Fig. 3.2, where the heavy atoms of the chromophore are numbered starting from the phenolic oxygen along the top ridge of the phenol, through

the bridge and around the imidazole ring.

For both the anionic (A) and the neutral (D) minimal model, the geometries obtained with the BLYP and the B3LYP functional are very similar, with the use of B3LYP shortening all bonds by a very small amount (about 0.01 Å). The structural parameters obtained using the PBE functional which we employ to generate the QM/MM protein models of GFP are not shown as they are practically indistinguishable from the BLYP bond lengths with an average agreement of 0.001 Å. For the anionic model (A), we also compute the bond lengths with the LDA functional as the local density instead of a generalized gradient approximation functional is used in the QM/MM calculations by Marques *et al.* [12] which we extensively discuss in Chapter 4. As expected, we find that LDA overbinds with respect to B3LYP but the difference is not very large with an average deviation of 0.01 Å.

In summary, all exchange-correlation DFT functionals yield equivalent geometries for both the minimal anionic (A) and neutral (D) chromophores, and a similar finding holds for all other models depicted in Fig. 3.1. No reference structure exists for the neutral chromophores in the gas phase but, for the minimal anionic (A) model, we can compare our DFT geometries to the correlated CASSCF/6-31G* calculations by Martin *et al.* [60]. As shown in Fig. 3.3, we find good agreement between the CASSCF and the DFT structures and, in particular, the B3LYP bond lengths are quite close to the CASSCF values with a maximum deviation of only 0.02 Å. Since there is no evidence that anionic molecules are better described by DFT than neutral ones, we can safely assume that the accuracy of DFT for all models of GFP is well within the spread of the different functionals we tested. Therefore, when we compare excitation spectra for gas-phase models, the eventual differences one observes should be attributed to the theoretical approaches employed to compute the excitation energy rather than to the approach followed to optimize the geometry.

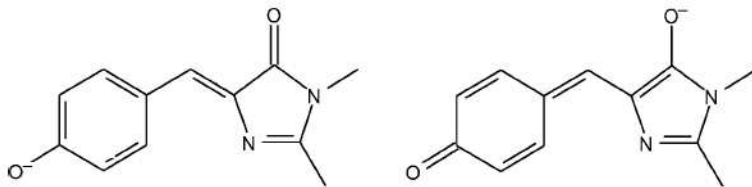


Figure 3.4: Scheme of the two resonant forms of the anionic chromophore: Benzenoid (left) and quinonoid (right).

While the main structural features are largely independent on the functional, the most evident difference is between the bond lengths of the neutral and the anionic model. To better understand the geometrical changes with the charge state of the chromophore, we show the two resonant forms of the anionic chromophore in Fig. 3.4. In the benzenoid form, the negative charge is localized on the phenolic oxygen and this bond structure is therefore also characteristic of the neutral chromophore. Upon deprotonation, the quinonoid form is also accessible where the negative charge has migrated to the imidazole oxygen. The change in bond length alternation and its reduction in the anionic chromophore is a measure of the mixing between the two forms. As we will see in Chapter 4, the protein environment can further tune the mixing by driving the resonance towards the benzenoid or the quinonoid form and therefore changing the bond structure of the chromophore.

The structural changes of Fig. 3.3 between the minimal neutral (D) to the anionic (A) model can now be more easily understood in terms of the two resonant forms. The neutral model is characterized by the aromaticity of the phenolic ring (with rather similar bond lengths between all carbon atoms of the ring) and by a double-single bond-length alternation at the carbon bridge given by the bonds C_5-C_6 and $C_6=C_7$. In the anionic model, the hydroxyl group is deprotonated and the oxygen-carbon bond, O_1-C_2 , shortens by about 0.1 Å as compared to the neutral model, loosing its single-bond character. As a consequence, the aromaticity of the phenolic ring is reduced, yielding a quinoid structure of the ring (double bonds between C_2-C_3 and the opposite carbon bond). The degree of bond alternation in the central carbon bridge is also decreased in the anionic chromophore, yielding a significantly stronger bond C_5-C_6 than in the neutral model: The two central bonds, C_5-C_6 and $C_6=C_7$ differ by about 0.08 Å in the neutral model and only 0.02 Å in the anionic chromophore. Beyond the central bridge, the deprotonation of the phenolic oxygen does not have as large an effect, and the two most significant changes are a reduction by 0.03 Å of the single bond C_7-C_8 and a lengthening by 0.02 Å of the C_8-O_{12} bond.

Finally, we note that the main structural changes between the anionic and neutral chromophores we describe for the minimal (A) and (D) models are also present when the larger models of the chromophore are considered. In Table 3.1, we list the main bond lengths optimized within DFT/BLYP for the anionic and neutral chromophores of the minimal (A), methyl-terminated (B) and protein-cut (C) models and the corresponding (D, E, F) neutral models. Both in the anionic and the neutral case, the bond lengths of the three models are practically identical, so the addition of the longer tails in the protein-cut chromophore has no effect on the relevant geometrical parameters of the chromophore.

	Anionic model			Neutral model		
	A	B	C	D	E	F
Bond length (Å)						
O ₁ —C ₂	1.26	1.26	1.26	1.37	1.38	1.37
C ₂ —C ₃	1.46	1.46	1.47	1.40	1.40	1.41
C ₃ —C ₄	1.37	1.37	1.37	1.39	1.39	1.38
C ₄ —C ₅	1.44	1.44	1.44	1.42	1.42	1.43
C ₅ —C ₆	1.41	1.41	1.41	1.45	1.45	1.45
C ₆ —C ₇	1.39	1.39	1.40	1.37	1.37	1.38
C ₇ —C ₈	1.47	1.46	1.46	1.50	1.49	1.50
C ₈ —N ₉	1.43	1.43	1.44	1.42	1.42	1.43
N ₉ —C ₁₀	1.39	1.40	1.40	1.38	1.40	1.39
C ₁₀ —N ₁₁	1.31	1.31	1.32	1.30	1.31	1.33
N ₁₁ —C ₇	1.41	1.41	1.41	1.41	1.41	1.41
C ₈ —O ₁₂	1.24	1.24	1.24	1.22	1.23	1.23
Dihedral angle (°)						
D(C ₄ C ₅ C ₆ C ₇)	180.0	180.0	178.3	180.0	180.0	179.5

Table 3.1: DFT/BLYP structural parameters of the minimal (A), methyl-terminated (B), and protein-cut (C) anionic models and the corresponding (D, E, F) neutral models. We list the most representative bond lengths and one dihedral angle. See Fig. 3.2 for the labeling of the atoms.

3.3 TDDFT excited states

We compute the low-lying TDDFT vertical excitations of the various model chromophores in the gas phase and locate the excitation with the largest oscillator strength which corresponds to the maximum light absorption. We compare these bright excitations to the available reference data, that is, to photodistraction spectroscopy experiments for the anionic methyl-terminated (B) [58] and the neutral⁺ (G) model [59], and to theoretical CASPT2 calculations for the anionic minimal (A) model [60]. These gas-phase calculations will already give us a feeling of which performance we may expect from TDDFT when the chromophores are embedded in the protein environment.

The all-electron linear-response TDDFT calculations are performed with the BLYP and B3LYP functionals at the corresponding BLYP and B3LYP ground state structures. We use the Gaussian03 code [53] and a cc-pVTZ basis as for the ground state calculations. For all model chromophores, we report the lowest two singlet excitations with their oscillator strength and

character which allow us to distinguish bright from dark states, and identify the electronic transition involved in the excitation. We also give two quantities which we use as indicators of the reliability of the TDDFT excitations, that is, minus the Kohn-Sham energy of the highest occupied molecular orbitals (HOMO) which indicates the start of the DFT ionization continuum, and the Kohn-Sham gap between the lowest unoccupied (LUMO) and the highest occupied molecular orbitals.

Before presenting the TDDFT results, we briefly discuss what we should expect for the excitations of the model chromophores in the gas phase. First, the excitation energy should decrease when progressing to larger chromophore sizes in the series (A, B, C) and (D, E, F) for the anionic and neutral models, respectively. This can be easily understood based on the simple argument of how the level spacing for a particle in a box behaves as the box is made larger. Moreover, we would expect the excitations of the anionic chromophores to be lower than the ones of the corresponding neutral models due to a combination of geometrical and electronic considerations. As discussed in the previous Section, the degree of bond alternation in proximity of the central carbon bridge is reduced in the anionic as compared to the neutral models, so the excitation of the bright $\pi \rightarrow \pi^*$ will be reduced. Moreover, as the anionic state can be described as a mixing of the benzenoid or the quinonoid forms, we expect also the excitation to be more delocalized and therefore lower than the neutral case. We will now see whether our picture of the behavior of the excitations is met in practice in the TDDFT calculations.

Anionic models

The TDDFT results for anionic minimal (A), methyl-terminated (B), and protein-cut (C) models are summarized in Table 3.2. For all three models and both exchange-correlation functionals, the excitation energy with the largest oscillator strength has a dominant HOMO $\rightarrow$ LUMO character. As the highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbitals are rather similar for the three models and the two functionals, we only plot them for the BLYP functional and the protein-cut model in Fig. 3.5. The highest occupied orbital has π -bonding character on the two carbon bonds of the central bridge while the lowest unoccupied orbital is a π -antibonding orbital on both bonds, and some degree of charge transfer can be observed across the bridge from the phenolic to the imidazole ring.

As expected, the electronic properties of the minimal (A) and the methyl-terminated (B) models are rather similar. The addition of the methyl groups only lowers the BLYP and B3LYP excitation energies by 0.08 and 0.07 eV, respectively, while preserving the same character of the excitation and the

oscillator strength. In the protein-cut (C) model, the BLYP and B3LYP functionals behave instead rather differently as the BLYP excitation with the largest oscillator strength is no longer the lowest state. However, if we consider the bright excitation for both the BLYP and B3LYP functionals, we find that the absorption maximum is further lowered by 0.03 and 0.05 eV with respect to the smaller methyl-terminated (B) model. Therefore, as expected, the excitation energy decreases with increasing size of the model.

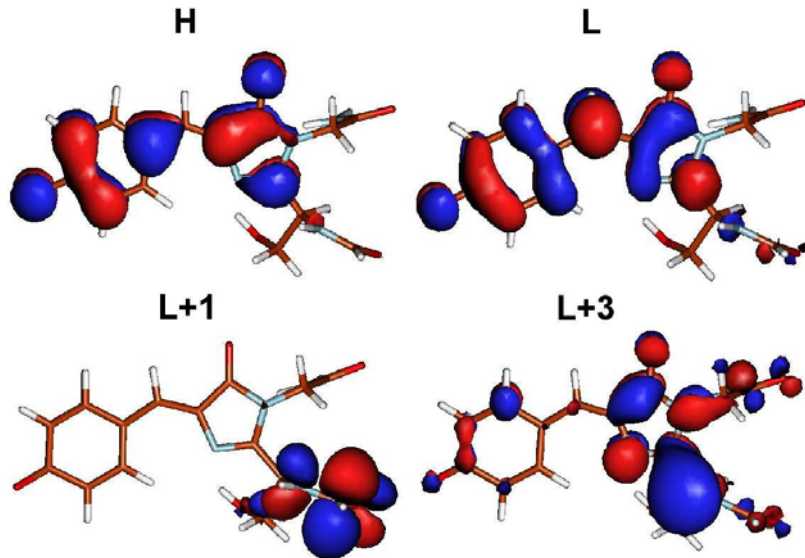


Figure 3.5: The DFT/BLYP orbitals for the anionic protein-cut (C) model chromophore in the gas phase. An isosurface of 0.025 is shown in red and an isosurface of -0.025 in blue. We only show the orbitals which are involved in the TDDFT/BLYP excitations.

We note that, as the model becomes larger, more orbitals are involved in the description of the BLYP excited states while the character of the lowest B3LYP excitation remains unchanged and predominantly $\text{HOMO} \rightarrow \text{LUMO}$. In particular, for the protein-cut (C) model, the lowest BLYP excitation has $\text{HOMO} \rightarrow (\text{LUMO}+1)$ character while the excitation with the largest oscillator strength has non negligible contributions from transitions to the $(\text{LUMO}+1)$ and $(\text{LUMO}+3)$ orbitals, which are depicted in Fig. 3.5. While the $(\text{LUMO}+3)$ orbital is largely localized on the imidazole ring, the $(\text{LUMO}+1)$ orbital is confined in the far tail of the model. As we expect equivalent electronic properties for a protein-cut chromophore obtained by differently setting the QM/MM boundary to shorten this tail, it appears

rather unphysical that the (LUMO+1) orbital would play such a relevant role in the excited states of the chromophore.

We now compare the linear-response TDDFT/BLYP and B3LYP excitation energies with the available reference data. For the methyl-terminated (B) anionic model, experiments [58] place the absorption maximum at 2.59 eV, significantly lower than the excitations of 2.89 and 3.09 eV obtained with the BLYP and B3LYP functionals, which therefore overestimate the experiment by as much as 0.30 and 0.50 eV, respectively. Correlated CASPT2 calculations by Martin *et al.* [60] for the anionic minimal (A) model give a vertical excitation energy of 2.67 eV which is in very good agreement with experiment as we consider that the electronic properties of models (A) and (B) must be rather similar. For the minimal (A) model, BLYP and B3LYP yield instead a significantly higher excitation energy of 2.97 and 3.16 eV, respectively. We discuss below the possible reasons of the apparent failure of TDDFT in describing the excitation energies of the anionic chromophore in the gas phase.

Neutral and neutral⁺ models

The TDDFT results for neutral minimal (D), methyl-terminated (E), and protein-cut (F) models are summarized in Table 3.3. Similarly to the anionic case, the B3LYP functional gives the lowest excitation as the one with the largest oscillator strength and HOMO $\rightarrow$ LUMO character for all three models. When the BLYP functional is employed, this is still true for the smaller models but not for the protein-cut (F) chromophore where the excitation with the largest oscillator strength is the second one, and has a large HOMO $\rightarrow$ LUMO component but a dominant (HOMO-2) $\rightarrow$ LUMO transition.

We plot the orbitals which are involved in the BLYP excitations for the protein-cut (F) model in Fig. 3.6. The highest occupied (HOMO) and lowest unoccupied (LUMO) molecular orbitals are rather similar for the three models and the two functionals. The HOMO orbital has π -anti-bonding/bonding character on the first/second bond of the central carbon bridge while the situation is reversed in the LUMO where the sequence is instead π -bonding/anti-bonding. The (HOMO-2) orbital which has dominant weight in the two lowest BLYP excitations is localized in the upper tail of the chromophore. Similarly to case of the BLYP excitations of the anionic protein-cut model, it is rather unphysical that the (HOMO-2) $\rightarrow$ LUMO excitation is so relevant as we could generate a different chromophore with equivalent electronic properties but a shorter upper tail by placing a different QM/MM boundary. This is likely an indication that the adiabatic TDDFT/BLYP description of the largest model is encountering some problems.

If we focus on the excitations with the largest oscillator strength for both functionals, we see that the B3LYP excitation energies are always larger than the BLYP ones by 0.3 eV, and that the excitation correctly decreases as the size of the model increases as it was the case for the anionic chromophores. We also note that the excitations of the neutral chromophores are always larger than the ones of the corresponding anionic models by roughly 0.2-0.3 and 0.4 eV for the BLYP and the B3LYP functional, respectively.

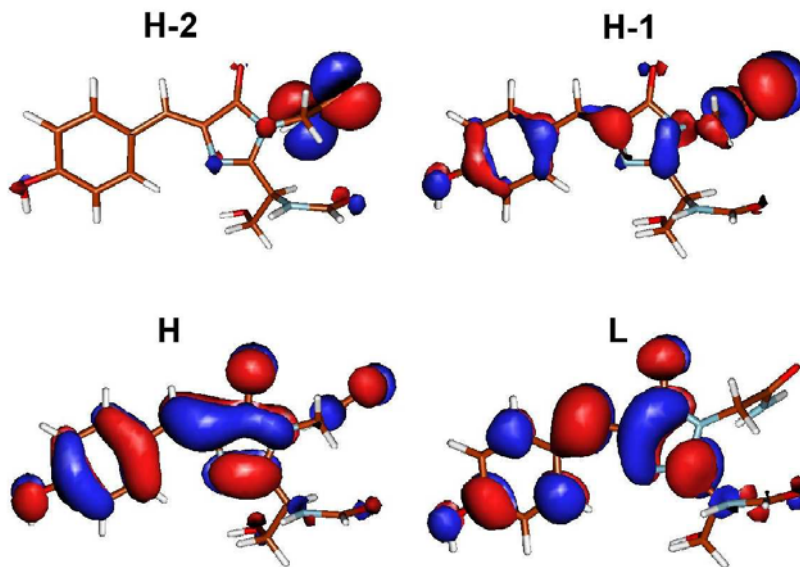


Figure 3.6: The DFT/BLYP orbitals for the neutral protein-cut (C) model chromophore in the gas phase. An isosurface of 0.025 is shown in red and an isosurface of -0.025 in blue. We only show the orbitals which are involved in the TDDFT/BLYP excitations.

In Table 3.4, we report the TDDFT excitations for the neutral⁺ (G) model, which must be compared to the absorption maximum of 2.99 eV measured in photodestruction spectroscopy experiments [59]. To estimate the excitation of the methyl-terminated neutral (E) model, the same experimental group corrects the absorption maximum of the neutral⁺ (G) chromophore by the difference between the theoretical TDDFT/B3LYP excitations of the (G) and (E) models optimized at the MP2 level, and obtains an excitation of 3.11 eV for the (E) model. We prefer to avoid this procedure as we obtain for instance different TDDFT corrections of 0.08 and 0.25 eV, when using the BLYP and the B3LYP functional with the corresponding DFT geometries. We instead focus on the direct comparison with the photodestruction

spectroscopy experiments for the neutral⁺ (G) model and the available experimental data obtained with the same spectroscopy technique for the anionic methyl-terminated (B) model [58]. For the neutral⁺ (G) model, we find that TDDFT/BLYP gives an excitation of 3.02 eV in very close agreement with experiments while the B3LYP excitation of 3.21 eV overestimates experiments by 0.22 eV. The experimental shift of 0.4 eV between the absorption maximum of the anionic methyl-terminated (B) model (2.59 eV) and the neutral⁺ (G) model (2.99 eV) is not reproduced by the TDDFT calculations. The BLYP and B3LYP shifts are equal to 0.13 and 0.12 eV, respectively, and therefore significantly smaller than the experimental value.

3.3.1 Assessing the performance of TDDFT

Before discussing the mixed performance of TDDFT when compared to the reference data for the anionic (A and B) and the neutral⁺ (G) models, we analyze some general features observed in the excitations of the gas-phase chromophores which will also be present when the chromophore is embedded in the protein environment. For the excited states of the smaller models, pure (BLYP) and hybrid (B3LYP) functionals yield excited states characterized by the same type of electronic transitions. For the larger models, BLYP mixes several transitions in contrast to the hybrid B3LYP functional which preserves a dominant single-excitation character. For all excited states, B3LYP yields higher excitation energies than the corresponding pure functional.

That B3LYP yields higher TDDFT excitation energies than BLYP can be in part understood from the fact that TDDFT applies a correction to the single-particle Kohn-Sham excitations given by eigenvalue differences. The Kohn-Sham eigenvalues behave rather differently if the functional contains a fraction of exchange as in B3LYP. The occupied orbitals will be lower as the hybrid functional is partially self-interaction corrected and the exchange-correlation potential no longer decays exponentially. On the other hand, the unoccupied orbitals will be higher as the virtual orbitals see a different number of electrons than the occupied ones due to the presence of Hartree-Fock exchange in the functional. Consequently, the Kohn-Sham excitations will be higher for the B3LYP than for the BLYP functional, as can be seen by inspecting for instance the B3LYP and BLYP HOMO-LUMO gaps in Tables 3.2, 3.3 and 3.4.

As TDDFT corrects the Kohn-Sham eigenvalue differences, it is important to try to establish, if possible, when these corrections are meaningful. Due to the spatial locality of the approximate adiabatic exchange-correlation kernel f_{xc} , the TDDFT correction to the Kohn-Sham excitations would vanish for a pure functional if the dominant transition is between non-overlapping

occupied and virtual orbitals. Therefore, if the TDDFT/BLYP excitation reduces to the Kohn-Sham eigenvalue difference we are in the presence of an excitation characterized by charge transfer which is poorly described by adiabatic TDDFT. When a hybrid functional is employed, the kernel is non-local in space and the TDDFT correction may not necessarily vanish. Therefore, as most excited states have a dominant HOMO $\rightarrow$ LUMO transition, we always report in the Tables the HOMO-LUMO gap since a comparison of the TDDFT/BLYP excitation with the corresponding gap reveals if the TDDFT excitation has charge transfer character and is consequently unreliable.

Another important indicator of the reliability of the TDDFT excitations is minus the eigenvalue of the HOMO orbital ($-\epsilon_{\text{HOMO}}$) as it locates the start of the TDDFT continuum and represents the ionization threshold. This threshold is usually underestimated in DFT as the Kohn-Sham potential for pure functionals decays exponentially and therefore too quickly. The use of hybrid functionals corrects to some degree this problem as they are partially self-interaction corrected. If the TDDFT excitation lies above the DFT ionization threshold, its value cannot in general be trusted.

We can now discuss the TDDFT performance for the anionic models. In Table 3.2, we observe that, for all models, the TDDFT/BLYP excitation with the largest oscillator strengths has HOMO $\rightarrow$ LUMO character and its energy lies well above the HOMO-LUMO gap. Therefore, this indicates that, for all models, the TDDFT/BLYP corrections to the Kohn-Sham excitations are not negligible and the brightest excited states are not charge-transfer excitations. For the protein-cut anion (C) model, the lowest BLYP excitation is predominantly a HOMO $\rightarrow$ (LUMO+1) transition and its energy of 2.65 eV closely agrees with the HOMO-(LUMO+1) gap of 2.68 eV. We had already noted the unphysical relevance of such a charge-transfer excitation as the (LUMO+1) orbital is localized in the far tails of the chromophore.

While charge transfer does not appear to pose a severe problem for the anionic chromophores, we note that the excitation energies of all models and for both functionals are significantly above the DFT ionization threshold. The use of B3LYP raises the ionization threshold as compared to BLYP but not sufficiently to bring it above the lowest excitation energy. As the chromophores are anionic, one may wonder whether the lowest excited state may actually be a quasi-stable excited state in the continuum which autoionizes after absorption, and therefore really lies above the ionization threshold. To investigate this point, we perform TDDFT calculations for the minimal (A) and the protein-cut (C) anionic models with the LB94 potential and the statistical average of orbital potential (SAOP) approach, which both yield the correct Coulombic tail ($-1/r$) in the exchange-correlation potential.

The TDDFT/LB94 and SAOP results obtained with the ADF code and

a ETpVQZ basis are reported in Table 3.5 for the ground state DFT/BLYP geometries. The use of an asymptotically correct exchange-correlation potential raises the ionization threshold well above the relevant excitation energy. Therefore, the BLYP/B3LYP excited states are not quasi-bound states in the continuum but the BLYP/B3LYP ionization threshold is simply underestimated. The TDDFT/LB94 and SAOP excitation energies for the minimal (A) model differ only by -0.08 and 0.02 eV from the BLYP excitation, respectively. For the protein-cut (C) anionic chromophore, the TDDFT/SAOP excitations are very similar to the BLYP ones with the lowest excitation having HOMO $\rightarrow$ (LUMO+1) character and the (LUMO+1) orbital being localized in the tails of the chromophore. The second SAOP excitation with the largest oscillator strength is only 0.07 eV higher than the brightest BLYP excitation. Therefore, even though the use of an asymptotically corrected potential has removed the issue of the underestimation of the DFT ionization threshold, the picture remains practically unchanged with respect to the use of the BLYP functional.

For the excitations of the neutral chromophores of Table 3.3, the underestimation of the DFT ionization threshold does not pose a problem as, for all models and both functionals, the DFT ionization threshold is above the excitation with the largest oscillator strength. As far as the charge-transfer character of the excitations, the behavior of the neutral minimal (D) and methyl-terminated (E) models is very similar to their anionic counterparts: The TDDFT/BLYP excitations have predominantly HOMO $\rightarrow$ LUMO character and the energy is significantly higher than the HOMO-LUMO gap. On the other hand, for the protein-cut (F) model, the lowest excitation has (HOMO-2) $\rightarrow$ LUMO character and the BLYP excitation energy of 2.97 eV is very close to the (HOMO-2)-LUMO gap of 2.99 eV indicating the charge-transfer character of the excitation. For this model, the second excitation has the largest oscillator strength but its energy is rather close to the lowest state and the predominant transition is still (HOMO-2) $\rightarrow$ LUMO with the (HOMO-2) orbital being localized on one the tails of the chromophores. This may point at some difficulties of adiabatic TDDFT in describing the excitations of the larger neutral chromophore.

In summary, it is not clear why TDDFT/BLYP and B3LYP should perform poorly in the description of the smaller anionic minimal (A) and methyl-terminated (B) models as compared to experimental and CASPT2 reference data. The excitations of these smaller models do not appear to be characterized by charge transfer and curing the underestimation of the ionization threshold with the use of asymptotically corrected functionals leaves the excitation energy practically unvaried. Therefore, it is not evident why TDDFT should be superior in describing the excitations of the corresponding neutral

(D) and (F) models or the neutral⁺ (F) chromophore. We note that, for the larger protein-cut models, the presence of significant contributions from charge-transfer transitions also in the excitation with the largest oscillator strength raises some doubts about the reliability of the TDDFT excitations, in particular for the neutral protein-cut (F) model.

Table 3.2: TDDFT/BLYP and B3LYP excitation energies (eV) and oscillator strengths (in parenthesis) for the minimal (A), methyl-terminated (B), and protein-cut (C) anionic models of the GFP chromophore, computed using a cc-pVTZ basis. DFT/BLYP and B3LYP geometries are consistently used. The dominant electronic transitions and their contributions in parenthesis (if > 0.1) are also listed. The lowest two singlet excitations are given together with minus the Kohn-Sham energy of the highest occupied molecular orbitals ($-\epsilon_{\text{HOMO}}$) and the Kohn-Sham gap between the lowest unoccupied and the highest occupied molecular orbitals ($\Delta\epsilon_{\text{HL}}$). The experimental absorption maximum for the methyl-terminated (B) chromophore is 2.59 eV [58] while CASPT2 calculations for the minimal (A) model give a vertical excitation of 2.67 eV [60].

Model	minimal	methyl-terminated	protein-cut
BLYP functional			
$S_0 \rightarrow S_1$	2.97(0.75)	2.89(0.77)	2.65(0.17)
	H $\rightarrow$ L(0.54)	H $\rightarrow$ L(0.54) H $\rightarrow$ L+2(0.13) H $\rightarrow$ L+3(0.11)	H $\rightarrow$ L+1(0.62) H $\rightarrow$ L(0.29)
$S_0 \rightarrow S_2$	3.25(0.00)	3.60(<0.01)	2.86(0.62)
	H-2 $\rightarrow$ L(0.70)	H-2 $\rightarrow$ L(0.59) H $\rightarrow$ L+5(0.18) H $\rightarrow$ L+3(0.17)	H $\rightarrow$ L(0.45) H $\rightarrow$ L+1(0.32) H $\rightarrow$ L+3(0.18)
$\Delta\epsilon_{\text{HL}}$	1.82	1.79	1.77
$-\epsilon_{\text{HOMO}}$	0.59	0.53	1.29
B3LYP functional			
$S_0 \rightarrow S_1$	3.16(0.88)	3.09(0.92)	3.04(0.96)
	H $\rightarrow$ L(0.57)	H $\rightarrow$ L(0.58)	H $\rightarrow$ L(0.58)
$S_0 \rightarrow S_2$	4.01(0.00)	4.24(<0.01)	4.02(0.02)
	H-3 $\rightarrow$ L(0.70)	H-2 $\rightarrow$ L(0.61) H $\rightarrow$ L+6(0.14) H $\rightarrow$ L+3(0.14)	H $\rightarrow$ L+2(0.70)
$\Delta\epsilon_{\text{HL}}$	2.96	2.93	2.91
$-\epsilon_{\text{HOMO}}$	1.29	1.22	1.98

Table 3.3: TDDFT/BLYP and B3LYP excitation energies (eV) and oscillator strengths (in parenthesis) for the minimal (D), methyl-terminated (E), and protein-cut (F) neutral models of the GFP chromophore, computed using a cc-pVTZ basis. DFT/BLYP and B3LYP geometries are consistently used. See caption in Table 3.2 for further explanations.

Model	minimal	methyl-terminated	protein-cut
BLYP functional			
$S_0 \rightarrow S_1$	3.22(0.59)	3.10(0.52)	2.97(0.20)
	H $\rightarrow$ L(0.57)	H $\rightarrow$ L(0.57)	H-2 $\rightarrow$ L(0.51)
	H-5 $\rightarrow$ L(0.13)	H-5 $\rightarrow$ L(0.14)	H $\rightarrow$ L(0.36)
		H-2 $\rightarrow$ L(0.13)	H-1 $\rightarrow$ L(0.22)
$S_0 \rightarrow S_2$	3.66(0.01)	3.56(0.13)	3.10(0.34)
	H-2 $\rightarrow$ L(0.57)	H-2 $\rightarrow$ L(0.63)	H-2 $\rightarrow$ L(0.44)
	H-3 $\rightarrow$ L(0.27)	H $\rightarrow$ L+2(0.19)	H $\rightarrow$ L(0.39)
	H $\rightarrow$ L+1(0.19)		H-4 $\rightarrow$ L(0.22)
	H $\rightarrow$ L+2(0.18)		
$\Delta\epsilon_{\text{HL}}$	2.23	2.19	2.13
$-\epsilon_{\text{HOMO}}$	4.98	4.76	4.87
B3LYP functional			
$S_0 \rightarrow S_1$	3.54(0.68)	3.46(0.66)	3.42(0.71)
	H $\rightarrow$ L(0.61)	H $\rightarrow$ L(0.60)	H $\rightarrow$ L(0.60)
$S_0 \rightarrow S_2$	3.56(<0.01)	3.66(<0.01)	3.58(<0.01)
	H-1 $\rightarrow$ L(0.69)	H-1 $\rightarrow$ L(0.69)	H-1 $\rightarrow$ L(0.66)
			H-3 $\rightarrow$ L(0.21)
$\Delta\epsilon_{\text{HL}}$	3.51	3.55	3.49
$-\epsilon_{\text{HOMO}}$	5.85	5.62	5.78

Table 3.4: TDDFT/BLYP and B3LYP excitation energies (eV) and oscillator strengths (in parenthesis) for the neutral⁺ (G) model of the GFP chromophore, computed using a cc-pVTZ basis. DFT/B3LYP ground state geometries are always used. See caption in Table 3.2 for further explanations.

	BLYP	B3LYP
$S_0 \rightarrow S_1$	3.02(0.73)	3.21(0.86)
	H $\rightarrow$ L(0.56)	H $\rightarrow$ L(0.60)
	H-1 $\rightarrow$ L(0.10)	
$S_0 \rightarrow S_2$	3.26(<0.01)	3.73(0.02)
	H-3 $\rightarrow$ L(0.50)	H-1 $\rightarrow$ L(0.67)
	H-1 $\rightarrow$ L(0.46)	H $\rightarrow$ L+2(0.13)
	H-2 $\rightarrow$ L(0.11)	
$\Delta\epsilon_{\text{HL}}$	2.01	3.19
$-\epsilon_{\text{HOMO}}$	7.80	8.63

Table 3.5: TDDFT/LB94 and SAOP excitation energies (eV) and oscillator strengths (in parenthesis) for the anionic minimal (A) and protein-cut (C) models computed using a ET-pVQZ basis and the BLYP ground state geometries. See caption in Table 3.2 for further explanations.

	minimal		protein-cut
	LB94	SAOP	SAOP
$S_0 \rightarrow S_1$	2.89(0.75)	2.99(0.79)	2.75(0.29)
	H $\rightarrow$ L(0.95)	H $\rightarrow$ L(0.95)	H $\rightarrow$ L+1(0.67)
			H $\rightarrow$ L(0.32)
$S_0 \rightarrow S_2$			2.93(0.62)
			H $\rightarrow$ L(0.60)
			H $\rightarrow$ L+1(0.32)
$\Delta\epsilon_{\text{HL}}$	1.73	1.86	1.83
$-\epsilon_{\text{HOMO}}$	6.31	5.10	5.85

3.4 QMC excitation energies

The use of QMC methods in combination with a careful construction of the many-body trial wave function has proven successful in describing the excitations of small prototypical photoactive molecules in the gas phase [61–64]. Here, we begin our exploration of the applicability of quantum Monte Carlo to the study of excitations in realistic biosystems, by computing the QMC vertical excitations of the model chromophores of GFP in the gas phase. For the smaller anionic minimal (A) model, we will also perform a thorough investigation of the possible limitations of QMC calculations for excited states. In particular, the dependence of the excitation energy on the trial wave function will be analyzed by optimizing several thousands parameters in excited-state trial wave functions of increasing complexity, using a robust and efficient optimization method we have recently developed [65].

For the QMC calculations in the gas phase, we focus on the chromophores which are relevant when comparing to reference experimental or CASPT2 data, that is, the anionic minimal (A) and methyl-terminated (B), the neutral minimal (D), and the neutral⁺ (G) models. We employ the ground state geometries optimized within DFT/B3LYP for all models except for the anionic minimal (A) model where, for historical reasons, we have used the DFT/BLYP ground state geometry. In quantum Monte Carlo, we compute the vertical $\pi \rightarrow \pi^*$ excitations as the difference between the total energies of the ground and excited states, which we obtain in a three step procedure. First, a conventional state-average (SA) complete active space (CAS) self-consistent field (SCF) calculation is performed. The resultant SA-CASSCF wave functions are then multiplied by a Jastrow factor to include dynamical correlation, and partially or fully reoptimized for the ground and excited states. Finally, the Jastrow-Slater wave functions are used in a variational Monte Carlo (VMC) calculation, and the VMC results are further improved via diffusion Monte Carlo (DMC).

The trial Jastrow-Slater wave functions

We briefly remind the reader about the many-body wave functions we use since they are the key ingredient which determines the quality of our QMC calculations, and are here chosen of the Jastrow-Slater type, with the particular form:

$$\Psi_I^{\text{VMC}} = \Psi_I^{\text{CAS}} \prod_{A,i,j} \mathcal{J}(r_{ij}, r_{iA}, r_{jA}), \quad (3.1)$$

where I labels the state of interest, r_{ij} denotes the distance between electrons i and j , and r_{iA} the distance of electron i from nucleus A . For most calcula-

tions, we use a Jastrow factor $\mathcal{J}$ which correlates pairs of electrons and each electron separately with a nucleus, and employ different Jastrow factors to describe the correlation with different atom types. We also investigate the effect of including electron-electron-nucleus correlation terms in the Jastrow factor. We refer the reader to Section 2.2.3 for further information on the specific form of the Jastrow factor.

The determinantal component consists of a complete active space (CAS) expansion which includes all possible space- and spin-adapted configuration state functions (CSF) obtained by placing n electrons in the active space of m orbitals, which is known as a CASSCF(n,m) wave function. Since the relevant excited state of all our models has the same symmetry as the ground state, the ground and excited-state wave functions are obtained in a state average (SA) CASSCF calculation, so the orbitals minimize the weighted average of the energies of the states of interest and the expansion coefficients are determined to preserve their orthogonality as explained in Section 2.2.1. Therefore, the wave functions of the different states share the same Jastrow factor and the same orbitals, but have different linear expansion coefficients on the CSFs. We note that, for the largest CASSCF wave functions, we may only keep the CSFs whose coefficient is above a chosen threshold. In this case, the threshold is separately applied to the ground- and the excited-state determinantal expansion, and the union of the surviving CSFs is then kept in both wave functions.

For the minimal anionic (A) model, we optimize all parameters in the Jastrow and the determinantal component of the wave function by energy minimization. Since the optimal orbitals and expansion coefficients in Ψ_I^{CAS} may differ from the CASSCF values obtained in the absence of the Jastrow factor $\mathcal{J}$, it is important to investigate the impact on the excitation energies when we reoptimize the determinantal parameters in the presence of the Jastrow component. If the wave function were the lowest state of a given symmetry, we could simply follow the energy-minimization approach of Ref. [66]. However, since the excited state in our models is not the lowest in its symmetry, we obtain both the Jastrow and orbitals parameters which minimize the average energy over the state of interest and the lower states, while the linear coefficients in the CSF expansion ensure that orthogonality is preserved among the states [65] as described in Section 2.2.3. For the other chromophore models studied in QMC, we either optimize a subset of parameters as the Jastrow factor and the linear coefficients, or simply optimize the Jastrow parameters in energy minimization for the ground state, and use the same Jastrow factor for the excited state calculations with an unoptimized SA-CASSCF determinantal component.

The trial wave functions of both states are then used in two separate

diffusion Monte Carlo (DMC) calculations, which produce the best energy within the fixed-node approximation, that is, the lowest-energy state with the same zeros (nodes) as the trial wave function.

All QMC calculations are performed with scalar-relativistic energy-consistent Hartree-Fock pseudopotentials specifically constructed for use in QMC, and with the corresponding Gaussian basis sets [67]. For most calculations, we employ a cc-pVDZ basis but we explore the effect of augmenting the basis with diffuse functions [68]. For this purpose, we generate an augmented cc-pVDZ basis (aug-cc-pVDZ) by adding one s and one p function with exponents 0.0469 and 0.04041 for carbon, 0.07896 and 0.06856 for oxygen, and 0.06124 and 0.05611 for nitrogen, respectively, and one s function with exponent 0.02974 for hydrogen.

3.4.1 The anionic minimal model: A case study

The anionic minimal (A) model represents a perfect playground to understand what the correct ingredients for our QMC calculations are. This model has the smallest number of atoms, so the calculations are faster, and has C_s symmetry which reduces the number of parameters in the determinantal component by roughly a factor of two. This reduction is convenient if we want to reoptimize the parameters of the SA-CASSCF wave function after including the Jastrow factor, and also because it is easier to converge very large CASSCF calculation with the quantum chemistry code GAMESS. In addition to these computational advantages, it is important to understand the anionic minimal (A) model which appears to be incorrectly described by adiabatic linear-response TDDFT (see Tables 3.2 and 3.5).

The determinantal CASSCF component

The first step in the generation of the many-body trial wave function is the construction of the SA-CASSCF determinantal component. Therefore, we want to explore the dependence of the excitation energy on the dimensions of the active space of the CAS wave function, and construct CASSCF(n,n) expansions of n electrons distributed over n orbitals with increasing values of n . The active space is build over the π/π^* orbitals (A'' symmetry) as the excitation of interest has $\pi \rightarrow \pi^*$ character and these orbitals are expected to be most relevant in describing the excitation. As 8 π orbitals are doubly occupied at the Hartree-Fock level, the maximum number of electrons in the active space can be $n = 16$.

In Table 3.6, we show the SA-CASSCF(n,n) excitation energies as a function of the dimension n of the active space for the anionic minimal (A) model

Table 3.6: SA-CASSCF(n,n) lowest excitation energy in eV of the anionic minimal (A) model as a function of the dimension n of the active space. We employ three different Gaussian basis sets, that is, cc-pVDZ, cc-pVTZ and aug-cc-pVDZ.

n	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ
2	4.10	4.05	3.92
4	3.57	3.53	3.38
6	3.47	3.44	3.21
8	3.17	3.12	3.02
10	3.25	3.21	3.20
12	3.21	—	—
14	3.11	—	—

in combination with different Gaussian basis sets. With the code GAMESS which we use to perform all CASSCF calculations, we are able to converge the SA-CASSCF(n,n) wave functions only up to $n = 14$ when the cc-pVDZ basis is employed, while we only reach $n = 10$ for the larger basis sets. From the calculations with the cc-pVDZ basis, we note that enlarging the active space significantly reduces the CASSCF excitation, which reaches however a roughly constant value beyond $n = 8$. As far as the dependence from the basis set, we observe that increasing the valence character of the basis from double (cc-pVDZ) to triple (cc-pVTZ) reduces the excitation energy by less than 0.05 eV. The augmentation of the cc-pVDZ basis set with diffuse functions has a slightly larger effect on the excitation energy which is lowered by almost 0.2 eV. However, this gain appears to be lost if the active space is enlarged to $n = 10$ where the difference between the excitation energies with and without augmentation is only 0.05 eV. This finding is in agreement with the all-electron CASSCF/6-31G* calculations by Martin *et al.* who observe a reduction of 0.06 eV in the CASSCF(12,11) excitation energy by augmenting the basis with diffuse functions. We note however that their CASSCF/6-31G* excitation of 3.68 eV is higher than our CASSCF(10,10) value, a difference which is possibly due to their calculation being all-electron.

In summary, from the CASSCF study, we can draw several conclusions which are useful in setting up the quantum Monte Carlo wave functions. First, we do not need to increase the valence nature of the basis and using a cc-pVDZ basis should be sufficient. It is however important to check the effect of augmentation on the cc-pVDZ basis, which may be visible if one makes use of small CAS expansions. As far as active space, we should investigate the effect of increasing the dimensions of the CAS at least up to $n = 8$.

The impact of the trial wave function on the QMC excitation

Having established the necessary ingredients in the determinantal component of the wave function, we now want to determine where DMC places the excitation of the anionic minimal (A) model, and perform several tests which are summarized in Table 3.7. For these calculations, we employ different determinantal components and an electron-nucleus and electron-electron Jastrow factor, and optimize the Jastrow and determinantal parameters within energy minimization in a state average approach. For some of the wave function forms, we also report the energy obtained without reoptimizing the CASSCF component, and using for both states the optimal Jastrow factor determined for the ground state by energy minimization.

Table 3.7: Variational (VMC) and diffusion Monte Carlo (DMC) excitation energies in eV of the minimal anionic (A) model of the GFP chromophore. The dimension of the CAS(n,n) determinantal component, the threshold on the CSFs and the number of CSFs included in the wave function are listed. Energies are given for the fully optimized (Optimized) wave function and for the wave function (Unoptimized) where only the Jastrow factor is optimized in correspondence of the ground state. The CASSCF excitation energy is given in the last column. The basis set used is specified.

	CSF		Unoptimized		Optimized		CASSCF
	Thr	Number	VMC	DMC	VMC	DMC	
cc-pVDZ basis							
CAS(2,2)	0.00	3	3.38(4)	3.18(4)	3.42(4)	3.15(4)	4.10
CAS(4,4)	0.00	20	3.33(4)	3.29(5)	3.31(4)	3.25(4)	3.57
CAS(6,6)	0.10	7	–	–	3.18(4)	3.11(7)	3.47
CAS(8,8)	0.10	9	–	–	3.06(4)	3.03(5)	3.17
CAS(8,8)	0.07	15	–	–	3.24(4)	3.10(5)	3.17
CAS(8,8)	0.05	25	–	–	3.11(4)	3.04(5)	3.17
aug-cc-pVDZ basis							
CAS(2,2)	0.00	3	3.24(4)	3.11(7)	–	–	3.92

We begin our analysis of the DMC results with the simplest CAS(2,2) wave function which is constructed from the HOMO and the LUMO orbitals as this ansatz will also be employed in the preliminary QMC study

of the chromophores embedded in the protein environment. Optimizing the orbitals and linear coefficients of the CAS(2,2) determinantal component in the presence of the Jastrow factor significantly lowers the absolute VMC and DMC energies of the two states (not shown). However, the VMC and DMC excitation energies for the optimized and unoptimized wave functions are equivalent within statistical error. When moving to a larger CAS(4,4) expansion, the situation is rather similar in the sense that optimization does not significantly affect the VMC and DMC excitation energies. In comparison to the CAS(2,2) results, the DMC excitations energies of the CAS(4,4) wave functions appear higher even though the difference with the CAS(2,2) values is still within two standard deviations.

When going to larger CAS expansions, we only keep the CSFs with coefficient above a chosen threshold as the wave function would otherwise not be tractable within quantum Monte Carlo. The number of CSFs grows indeed very rapidly with the size of the active space: For instance, the CAS(8,8) wave function contains 1764 CSFs which become 19404 in a CAS(10,10) expansion. Since each CSF consists of several determinants, the full determinantal wave function can then be formed by several thousand determinants, which would render the QMC calculation exceedingly slow. We note that the wave function obtained by applying a threshold to a given CAS expansion does not reduce to one of the smaller CAS wave functions as different excitations may now be important and survive the threshold. Using a CAS(6,6) wave function with a rather large threshold of 0.1, we obtain a VMC excitation energy which is 0.22(6) eV smaller than the CAS(2,2) value, and a DMC excitation of 3.11(7) eV which is still compatible with both the CAS(2,2) and CAS(4,4) results within statistical error. Finally, we optimize the Jastrow, orbital, and linear parameters of a CAS(8,8) with a threshold of 0.1, and subsequently, reduce the threshold to 0.07 and 0.05 reoptimizing only the linear expansion coefficients. The VMC excitation does not behave monotonically as, by lowering the threshold to the intermediate 0.05 value, CSFs more relevant for the ground than the excited state may have entered the wave function. However, within statistical error, the DMC excitation energies are always rather comparable with each other falling in the range 3.03(5)-3.10(5) eV, and are certainly lower by at least 0.1 eV than the DMC values obtained with the optimized CAS(2,2) and CAS(4,4) wave functions.

In summary, it appears that the effect of increasing the CAS expansion and therefore improving the many-body wave function is to reduce the difference between the VMC and the DMC excitation energies, with the VMC gap approaching the DMC value from above. The DMC energy is very robust and depends not too strongly on the size of the active space with the DMC gap being lowered by only roughly 0.1-0.2 eV when going from a CAS(2,2) to a

CAS(8,8) wave function. The effect on the DMC excitation of optimizing the determinantal component in the presence of the Jastrow factor appears to be rather small. Finally, we note that the use of the cc-pVDZ basis is sufficient as augmenting the basis does not affect the DMC excitation energy of an unoptimized CAS(2,2) wave function and we know from the CASSCF study that the impact of diffuse functions would anyhow be washed out in going to larger CAS expansions. Further tests on the effect of using electron-electron-nucleus terms in the Jastrow factor, of performing a time-step extrapolation of the DMC results, and of treating the non-local pseudopotentials beyond the locality approximation do not change the present picture.

While this analysis reassures us about the robustness of our DMC approach in computing the brightest excitation energy of the minimal anionic (A) model, we must note that the DMC excitation energy is in the range 3.0-3.2 eV and therefore in rather good agreement with the TDDFT values of 2.97 and 3.16 eV obtained with the BLYP and B3LYP functionals, respectively. Consequently, the DMC excitation energy is significantly higher than the CASPT2/6-31G* value of 2.67 eV for the minimal anionic model [60], and not in good agreement with the absorption maximum of 2.59 eV obtained in photodestruction spectroscopy experiments for the closely related anionic methyl-terminated (B) model [58].

3.4.2 The neutral and anionic models at comparison

To better understand the apparent overestimation by DMC of the excitation of the minimal anionic (A) model, we want to compare the DMC results obtained for different model chromophores, that is, the minimal neutral (D), the anionic methyl-terminated (B) and the neutral⁺ (G) model. This will allow us to observe how the excitation correlates to the charge state of the chromophore, and to compare the QMC excitations to other available reference data. Since the excitation energy of the anionic minimal (A) model is not particularly affected by either the optimization of the determinantal component or the use of large CAS expansions, we perform all QMC calculations consistently using an unoptimized CAS(2,2) determinantal component in the wave function. We will take into account in our analysis that the excitations may be overestimated by roughly 0.1-0.2 eV due to the use of the small CAS expansion. The results are summarized in Table 3.8.

We first compare the results for the minimal anionic (A) and neutral (D) models and note that DMC yields an excitation for the neutral species 0.9(1) eV higher than the one for the anionic counterpart. This finding is in agreement with the difference of 0.8 eV found in semi-empirical configuration interactions (CI) calculations for larger anionic and neutral chromophore

Table 3.8: Variational (VMC) and diffusion Monte Carlo (DMC) energies in eV for the minimal anionic (A) and neutral (D), the anionic methyl-terminated (B), and the neutral⁺ (G) model, computed using an unoptimized CAS(2,2) wave function. The TDDFT/BLYP and B3LYP energies are also listed together with the photodistraction spectroscopy experimental absorption maxima from Refs. [58]^a and [59]^b.

	VMC	DMC	BLYP	B3LYP	Expt.
Minimal anionic	3.38(4)	3.18(4)	2.97	3.16	–
Minimal neutral	4.27(7)	4.09(7)	3.22	3.54	–
Methyl-term. anionic	3.43(5)	3.20(5)	2.89	3.09	2.59 ^a
Neutral ⁺	3.67(8)	3.37(7)	3.02	3.21	2.99 ^b

models in the gas phase [69]. The semi-empirical CI excitations energies for the anionic and neutral models are equal to 3.52 and 2.70 eV, respectively, but the parameters of the semiempirical approach were tuned to yield an excitation for the anionic model reasonably close to the experimental gas-phase value [70]. On the other hand, TDDFT yields a significantly smaller difference between the excitations of the neutral and the anionic minimal models, that is, 0.25 and 0.38 eV with the BLYP and the B3LYP functional, respectively. Therefore, while the TDDFT excitation for the anionic chromophore is in reasonable agreement with the DMC value, the TDDFT excitation of the neutral is lower than the DMC energy by 0.87 and 0.54 eV for the BLYP and the B3LYP functional, respectively.

If we compare the DMC excitation energies of the anionic methyl-terminated (B) model and the neutral⁺ (G) models with TDDFT and available experimental numbers, we should keep in mind that the DMC excitations are overestimated by 0.1-0.2 eV since, for the moment, we only employed simple CAS(2,2) wave functions. While a reduction of 0.1-0.2 eV brings the DMC energies of both models in close agreement with the TDDFT/B3LYP excitations, the resulting excitation for the neutral⁺ (G) model remains higher by 0.1-0.2 eV than the absorption maximum located in photodistraction spectroscopy experiments. Moreover, as already discussed, DMC appears to be overestimating the excitation of the anionic form by roughly 0.4-0.5 eV.

3.5 Conclusions

Before studying the neutral and anionic forms of GFP within the protein environment, we constructed a series of model chromophores of GFP in the gas phase to begin exploring the performance of adiabatic TDDFT and quantum Monte Carlo approaches. We built three models of increasing complexity for both the anionic (deprotonated) and the neutral (protonated) chromophore of GFP, and also investigated a particular cationic model which was recently characterized in photodestruction spectroscopy experiments. Reference experimental data obtained with the same spectroscopy technique as well as CASPT2 theoretical results are available for two of the smaller anionic models we study.

The results are rather puzzling. Adiabatic TDDFT reproduces the experimental absorption maximum of the cationic model reasonably well if a pure functional is used, but appears to be overestimating the excitation energies of two small anionic chromophores by 0.3-0.5 eV depending on the functional employed. Moreover, the TDDFT excitation energies for the neutral models are not too dissimilar from the excitations of the corresponding anionic chromophores while one would expect a significant shift in the excitation upon deprotonation. We analysed various possible shortcomings of TDDFT in describing the anionic form of GFP but we were not able to identify evident problems of TDDFT, especially in the calculations for the smaller models of the GFP chromophore. The excitations of these models do not appear to be characterized by charge transfer and curing the underestimation of the DFT ionization threshold with the use of asymptotically corrected functionals does not change the excitation energy. Therefore, it is not evident why TDDFT should be superior in describing the excitations of the corresponding neutral models or of the cationic chromophore. On the other hand, for the larger model chromophores that we will employ within the protein, we found significant contributions from charge-transfer transitions in the TDDFT excitations, which are a signature of potential problems with the TDDFT description of these larger models, especially in the neutral form.

Using quantum Monte Carlo approaches, we performed a thorough study of the excitation energy of the smallest anionic model using sophisticated trial many-body wave functions and state-of-the-art optimization techniques. We find that the DMC excitation energy is very close to the TDDFT result, and therefore higher than the reference photodestruction spectroscopy experimental result by roughly 0.4-0.5 eV. We could attempt to use even more complex wave functions but, based on the tests done so far, the DMC result appears to be rather robust. Differently from TDDFT, when comparing the anionic

and neutral chromophores, we observed a sizable shift in the excitation as the DMC excited state of the neutral form is 0.9(1) eV higher than the energy of the corresponding anionic model. Finally, the shift between the anionic and cationic model chromophores observed in photodestruction spectroscopy experiments is not well reproduced within DMC but, for both models, the DMC excitations appear to be rather close to the TDDFT values obtained using hybrid exchange-correlation functionals.

3.5. Conclusions

3. Chromophore in vacuum