



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 4

Treating the protein environment in QM/MM

4.1 Construction of the protein model

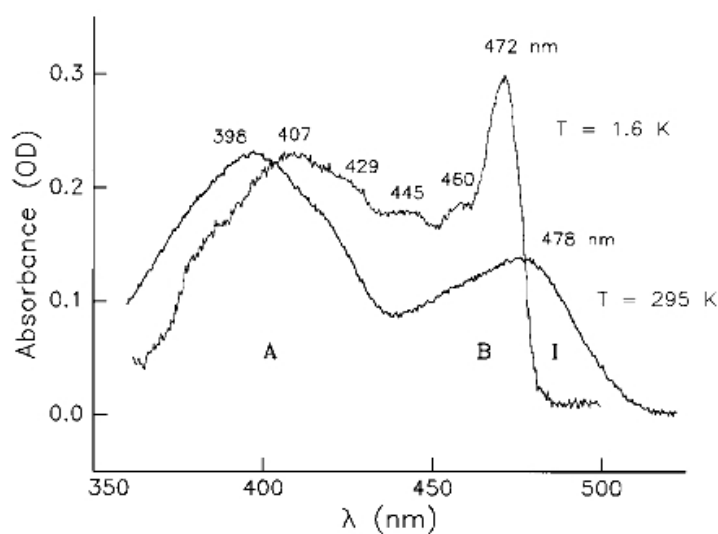


Figure 4.1: Absorption spectra of wild-type GFP at room temperature ($T=295$ K) and low temperature ($T=1.6$ K). Adapted from Ref. [7]

As the focus of this Chapter is the construction of a theoretical model to describe the photophysics of GFP, we briefly remind the reader some of the key spectral properties of this autofluorescent protein. As shown

4.1. The protein model 4. Treating the protein environment in QM/MM

in Fig. 4.1, the room temperature spectrum of wild-type GFP is characterized by two maxima which are attributed to two interconvertible states (protonated/deprotonated) of the protein. The absorption band at 398 nm (3.12 eV) is from the neutral A form while the band at 478 nm (2.59 eV) from the anionic B form of the protein. Upon photoexcitation of A, the excited chromophore transfers a proton through a hydrogen-bond network to the surrounding environment forming a transient intermediate anionic state (I^*) which emits in the region of 506 nm (2.45 eV). After decay to the ground state (I), the system usually returns to state A through a ground state inverse proton transfer process. The green fluorescence at 482 nm (2.57 eV) following direct excitation of the B state stems from direct decay of the excited B^* state. In Fig. 4.1, the spectrum at 1.6 K is also shown where the ratio of the absorbances of the A and B forms inverts and the two maxima shift at 407 nm (3.05 eV) and 472 nm (2.63 eV). The broad wing at the red side of the 472 maximum disappears and is attributed to the I form which is not populated at this low temperature. The 0-0 transitions of all three forms have been located in hole-burning spectroscopy experiments [7].

To compute the spectrum of wild-type GFP, we construct a representative configuration of the chromophore-protein structure. As we compare to low-temperature ($T=1.6$ K) spectroscopy experiments, it is a reasonable choice to evaluate the theoretical spectrum of a single conformation obtained in a simulated annealing procedure instead of following the more costly route to compute the average spectrum over several snapshots of a molecular dynamics (MD) simulation at the experimental temperature. Moreover, we will see below that the chromophore is kept rather rigidly within its binding site by a complex network of hydrogen bonds. In this Section, we describe how this representative conformation for the neutral A and anionic I and B forms of wild-type GFP is constructed starting from the available X-ray structures of neutral wild-type GFP and of the GFP mutant S65T where the mutations stabilize the anionic B form against the A form of GFP.

4.1.1 The neutral form

The starting structure for the construction of the neutral form of the protein is the X-ray structure [71] at 1.90 Å resolution (entry 1GFL in the Protein Data Bank [72]). During the crystallization process precedent to the X-ray scanning, the proteins pack together to form dimers so that the X-ray structure with code 1GFL contains not one but two GFP proteins. In our simulations of the neutral and anionic I forms, we keep the dimer for simplicity. The setup and a fast preliminary MM equilibration is performed using the Amber suite of programs [43] and the Amber force field is used in all

4. Treating the protein environment in QM/MM 4.1. The protein model

MM simulations. The structure is then refined in a simulated annealing run within QM/MM.

Equilibration of GFP with MM

Since X-ray diffraction experiments do not distinguish hydrogen atoms, the positions of the hydrogens are not given in the X-ray structure of GFP and must be added when setting up the simulation. This is in part automatically done using the module *Xleap* within the Amber package as most of the standard aminoacids exist in a given ionization state at neutral pH. Exceptions are the glutamine/glutamic acid, aspartic acid, lysine and histidine which can have different protonation states. Their charge state is assigned based on their hydrogen bonding configuration, or chosen as they most frequently appear in nature. Of these aminoacids, only the histidine and the glutamic acid are present in the hydrogen-bond network surrounding the chromophore. Since we do not perform long MD simulation but only refine the structure by simulated annealing, we focus our attention on the protonation state of the histidines and the glutamic acids in the binding site, and leave the Amber package to assign the protonation states of the other residues. Based on the most likely hydrogen-bonding configuration of the chromophore, the histidine residues numbered 25, 148, 181, 199 and 217 are protonated at their δ nitrogen while the remaining histidine residues are protonated at their ϵ nitrogen. For the protonation of the glutamic acids, particular attention is given to Glu-222 since this residue is involved in the proton shuffle between the neutral A and the anionic forms. As explained later, experiments and theoretical work indicate that Glu-222 is anionic in the A state while is the proton receptor in the anionic I and B form of GFP. We follow this criterion in our setup.

To simulate the protein *in vivo*, the protein is then set in a physiological solution, with a saline concentration of 0.15 M. The protein is placed at the center of a cubic MM simulation box, surrounded by 12 Å of TIP3P [73] water molecules in each direction. As periodic boundary conditions are applied, the box is sufficiently large to avoid interactions between the images of the protein. Finally, counter-ions are added to the water solution to achieve physiologic concentration and to ensure that the cell is completely neutral. As the total sum of the partial charges of the protein is not zero but equals -12 , we added more positive ions (63 Na^+) than negative ones (51 Cl^-) to have a simulation box with zero total charge. The total simulation box contains around 70000 atoms and its size is roughly $83 \times 97 \times 82 \text{ Å}^3$. In Fig. 4.2, we show part of the simulation cell comprising the protein and a few Å of the surrounding water.

4.1. The protein model 4. Treating the protein environment in QM/MM

We first perform a classical MM equilibration of the hydrogens and the waters using the module *Sander* of the Amber program in order to start the more costly QM/MM simulations with these many degrees of freedom in a stable configuration. During the relaxation, the coordinates of the heavy atoms of the protein are kept fixed to preserve a protein structure close to the original crystallographic one, and because the force field parameters for the chromophore are not available in the Amber library as the chromophore is not a standard aminoacid. The chromophore force field parameters are available at the Charmm level [74] but their use would then be incompatible with the subsequent QM/MM simulations.

Even though the heavy atoms of the chromophore are kept fixed, we need to assign a force field to the chromophore as its atoms interact with the hydrogens of the protein, the solution waters, and the ions which are being equilibrated. Fortunately, as the chromophore is kept fixed, we do not need a very accurate force field and we can proceed to generate one for our particular purpose as follows. We determine the partial charges of the atoms of the chromophore by computing a quantum mechanical electrostatic potential within Hartree-Fock using a 6-31G* basis and the Gaussian03 [53]

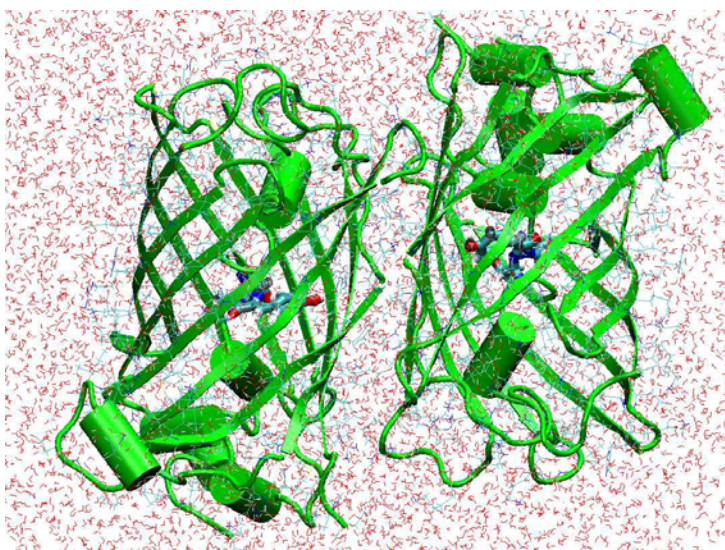


Figure 4.2: Model of the neutral A form of wild-type GFP in water solution. The dimerized structure is shown together with the chromophores in the protein cavities. The diameter of the barrel of a single GFP protein is approximately 24 Å and the height 42 Å. The distance between the two proteins is approximately 5 Å. The two chromophores do not interact as they are shielded by the protein barrel and are at a distance of about 17 Å.

4. Treating the protein environment in QM/MM 4.1. The protein model

code. The partial charges are then fitted to reproduce this potential on a large number of grid points covering the region around the chromophore. We assign the bond, angle, and dihedral parameters to each atom of the chromophore based on its position and chemical role following Ref. [75]. We know that this force field is not sufficiently accurate to describe the intra-molecular interactions of the heavy atoms of the chromophore since an energy relaxation test performed for this chromophore model does not yield a stable structure. However, we consider it sufficient to give an approximate description of the interaction of the chromophore atoms with the rest of the system.

To restrain the protein coordinates, we find that simply locking the protein coordinates leads to an unstable simulation. Therefore, we introduce a restoring potential, $k \Delta x^2$, where Δx is the displacement of the atomic coordinates respect to the reference X-ray structure. Finally, the MM equilibration is performed in the following three steps:

- 1) 1000 steps of energy minimization using the steepest descent algorithm in the first 10 steps, and the conjugate gradients algorithm afterwards.
- 2) 33 ps of classical isothermal and isobaric molecular dynamics at 300 K and 1 atm. We use an isotropic pressure scaling and the Andersen thermostat to couple the temperature to an external bath. We consider the system equilibrated when the temperature fluctuations are less than 5% and the density is constant within 2%. The time step is 0.0005 ps.
- 2) 3500 steps of energy minimization.

Finally, an equilibrated structure is obtained with a norm of the atomic forces of 0.1 kcal/mol/Å.

Simulated annealing within QM/MM

To obtain an accurate description of the structure of the chromophore binding site and of the internal geometry of the chromophore which was left so far unrelaxed, we start from the structure obtained in the preliminary MM equilibration and perform a simulated annealing procedure within QM/MM. In the QM/MM relaxation, the atoms of the chromophore are treated quantum mechanically and the rest of the protein and the solvent are treated classically. The QM/MM boundary is set through the single $\text{C}_{\text{OOH}}-\text{C}_\alpha$ bond of Phe-64 and the single $\text{N}-\text{C}_\alpha$ bond of Val-68 as shown in Fig. 4.3. We note that, as we keep the dimer structure of the original crystallographic structure, we only treat quantum mechanically the chromophore of one of the two GFP barrels and leave the other chromophore at the fixed crystallographic coordinates.

4.1. The protein model 4. Treating the protein environment in QM/MM

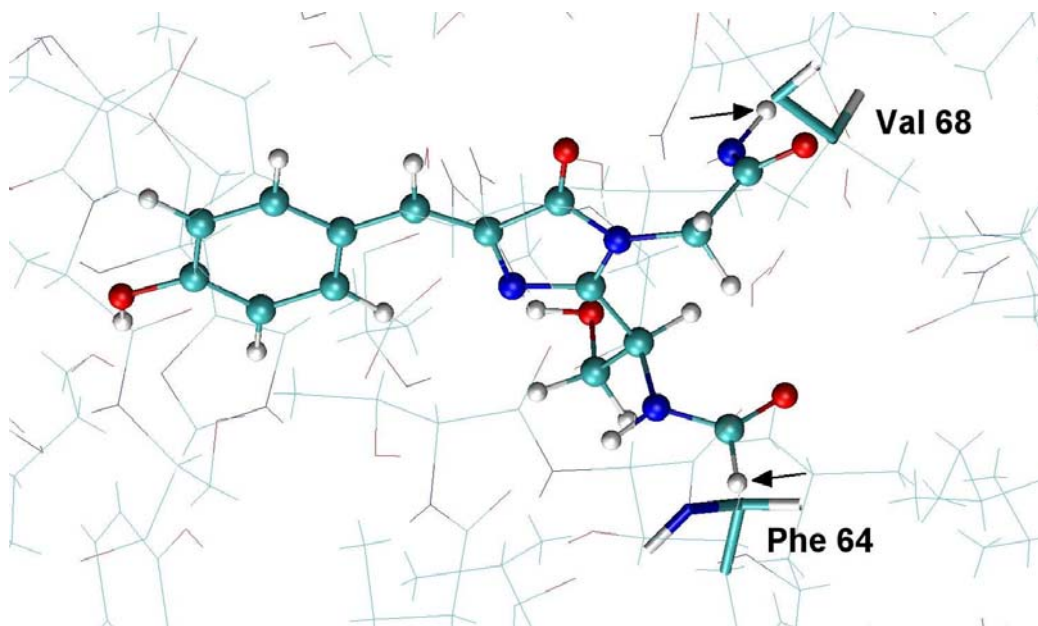


Figure 4.3: QM/MM partitioning for the A form of wild-type GFP. The QM/MM boundary is set through the single $\text{C}_{\text{OOH}}-\text{C}_{\alpha}$ bond of Phe-64 and the single $\text{N}-\text{C}_{\alpha}$ bond of Val-68. The two arrows indicate the two hydrogen-link atom between the QM and the MM. The MM part closest to the QM atoms is represented with cylinders.

The QM/MM simulations are performed with the density functional theory CPMD [46] code using the PBE generalized gradient approximation functional [52] and a plane wave cutoff of 70 Ry. The size of the supercell used for the QM calculations is roughly $21 \times 15 \times 13 \text{ \AA}^3$, and the chromophore is approximatively in the xy plane. A time step of about 0.7 fs is used.

The simulated annealing is performed through a molecular dynamics simulation where the velocities are rescaled at each step by a factor f . To quench the system and obtain a converged structure of the chromophore binding site, we perform:

- 1) 1.40 ps of molecular dynamics with $f = 0.999$.
- 2) 0.35 ps of molecular dynamics with $f = 0.99$.

At the end of the quenching, the amplitude of the oscillations for the bond lengths along the chromophore are of the order of $0.005 \text{ \AA}$, and thus negligible.

In Fig. 4.4, we show the final structure of the binding site of the neutral A form with the closest residues to the chromophore which are most relevant for either of the three forms of wild-type GFP. The A form is character-

ized by a hydrogen-bond network connecting the oxygen of the chromophore to the carboxylate of the negative Glu-222 through the water W1 and Ser-205. As we discuss below in Section 4.1.2, the accepted picture is that this hydrogen-bond chain provides the path for the proton leaving the photoexcited chromophore and being transferred to Glu-222. We also observe that the negatively charged Glu-222 is stabilized by a hydrogen bond donated by the oxygen of the tail of the chromophore (Ser-65). The stabilization of the anionic Glu-222 is crucial for the existence of the A form as perturbation of the hydrogen-bond network surrounding this residue leads to destabilization of the A state in favor of the B state.

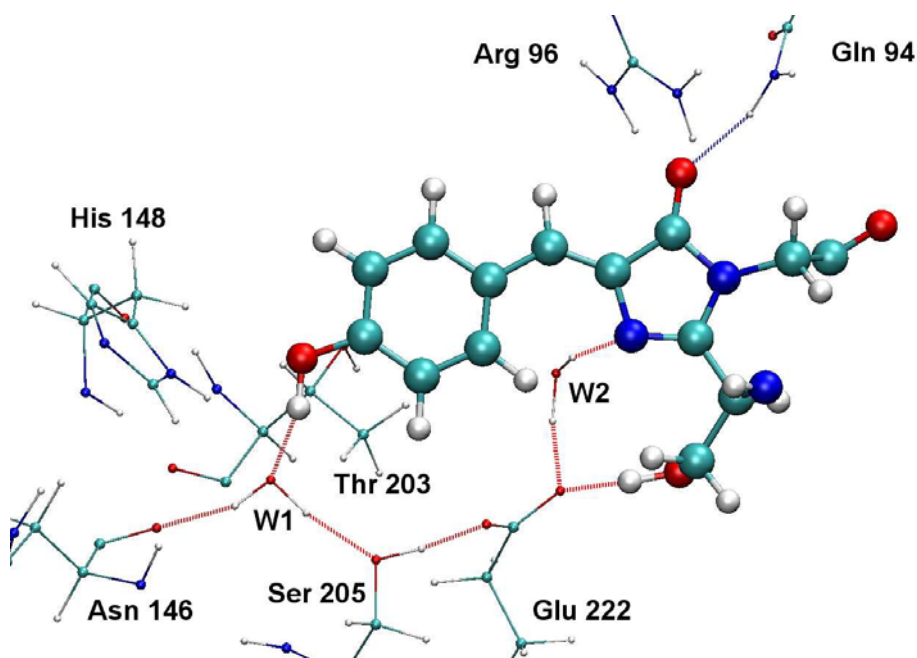


Figure 4.4: Binding site of the neutral A form of GFP. The residues closest to the chromophore which are relevant in either the A, B, or I form are shown. The hydrogen bonds are drawn if the bond length is less than 3 Å and the donor-hydrogen-acceptor angle is less than 30°. Glu-222 is negatively charged in the A form and is the acceptor of the proton leaving the hydroxyl group of the phenolic ring upon photoexcitation.

4.1.2 The intermediate form

Both the anionic intermediate and B forms differ from the neutral form as the proton of the hydroxyl group of the phenolic ring has left the chromophore

4.1. The protein model 4. Treating the protein environment in QM/MM

and has been transferred to the surrounding environment. For both forms, we assume that the proton leaving the phenolic ring is transferred to the negatively charged residue Glu-222 which becomes neutral upon protonation.

The identification of the residue accepting the proton leaving the chromophore has been a subject of debate for several years. Already in the early models proposed after structural and spectroscopic analysis of wild-type GFP and some of its mutants [76–78], it was suggested that Glu-222 is negatively charged in the A form and the acceptor of the proton upon photoexcitation. This picture was for instance put forward by Ormö *et al.* [76] in their X-ray characterization of the S65T mutant where the B form is stabilized by mutation of the side chain of the chromophore at position 65, which donates a hydrogen bond to Glu222 in the A form of wild-type GFP (see Fig. 4.4). This transfer mechanism was however doubted by successive FTIR (Fourier transform infrared spectroscopy) experiments [79], where Glu-222 appears to be protonated in both the neutral and the anionic form as the feature corresponding to the change in its protonation is absent in the FTIR difference spectrum. However, theoretical calculations of the vibrational frequencies of both forms [80] allowed to reinterpret the FTIR results by showing their compatibility with a protonation state of Glu-222 in the B form different from the one in the A form. Finally, the good agreement between classical MD calculations [69] and the X-Ray structure of the S65T mutant where the chromophore is stabilized in the anion form with a protonated Glu-222 also supports the picture of a proton transfer mechanisms from the chromophore to Glu-222 in wild-type GFP.

Finally, we note that the proton can be added to Glu-222 in the *syn*- or the *anti*-configuration. In the theoretical work by Tozzini and coworkers [69], it was observed that, during the classical molecular dynamics simulations at 298 K, the hydrogen jumps between the two configurations but that the *anti*-conformation is the one occupied for longer times. Therefore, we protonate the Glu-222 residue with the proton in the *anti*-configuration. In Fig. 4.5, we show the anionic intermediate form with the chromophore and the most important residues which are believed to be involved in the proton transfer mechanism.

For the calculation on the I form of the chromophore we followed the same procedure as described for the A form. A QM/MM simulated annealing procedure is performed on the structure following these steps:

- 1) 3 ps of molecular dynamics with $f = 0.999$.
- 2) 1 ps of classic molecular dynamics at constant volume at 300 K. Only the residues involved in the hydrogen-bond network of Fig. 4.5 are moved while the chromophore and the rest of the protein is kept fixed.

4. Treating the protein environment in QM/MM 4.1. The protein model

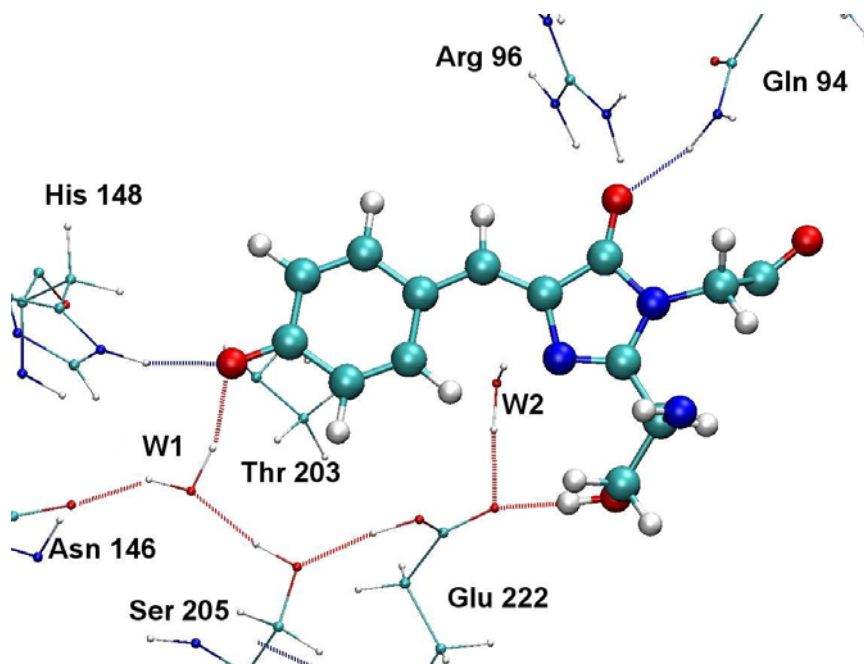


Figure 4.5: Binding site of the anionic I form of GFP. The residues closest to the chromophore which are relevant in either the A, B, or I form are shown. The hydrogen bonds are drawn if the bond length is less than 3 Å and the donor-hydrogen-acceptor angle is less than 20°. The chromophore is deprotonated and Glu-222 is commonly considered as the acceptor of the proton leaving the hydroxyl group of the phenolic ring.

This allows a faster rearrangement of the residues interested in the proton-shuffle.

3) 0.28 ps of molecular dynamics with $f = 0.999$.

In Fig. 4.5, the final structure of the binding site of the I form is shown with the closely residues which are the most relevant for either of the three forms of wild-type GFP. The hydrogen-bond network connecting the phenolic oxygen to Glu-222 is shown. In the anionic form, His-148 donates a hydrogen bond to the phenolic oxygen which is now deprotonated while Glu-222 is neutral. The electrostatic potential generated by the MM atoms of the protein and the solvent on the QM system is shown in Fig. 4.6. As expected, the potential is positive and therefore attractive for electrons around the positive Arg-94 and the hydrogen atom of Gln-94 which creates a hydrogen bond to the oxygen of the imidazole ring. It is also positive along the hydrogen-bond network running from the phenolic oxygen through the W1 water and Ser-205 to Glu-222.

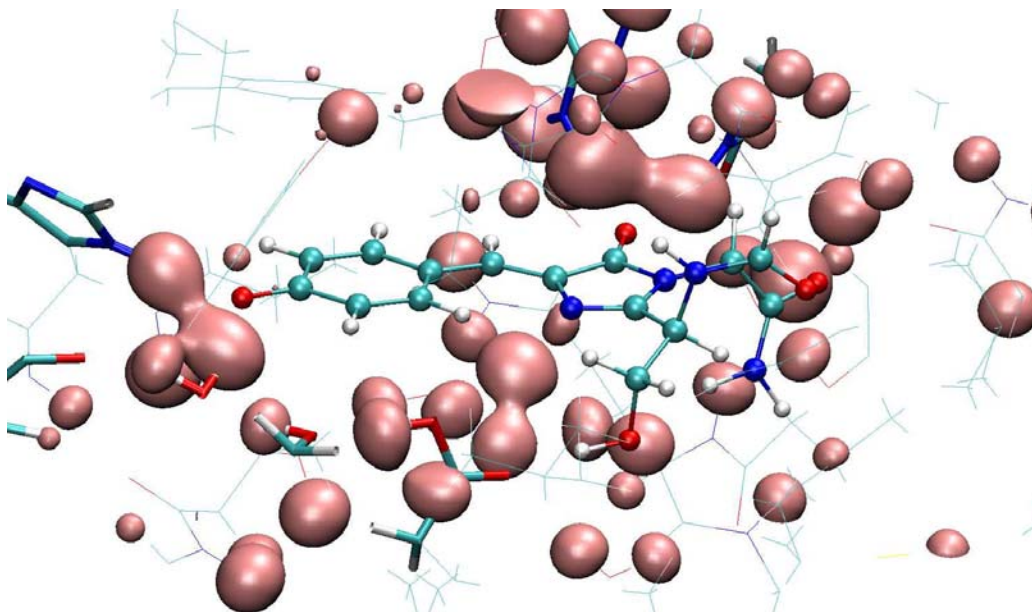


Figure 4.6: Electrostatic potential created by the MM atoms on the QM system. The residues which hydrogen bond to the chromophore are shown in a thicker representation (also compare with Fig. 4.5). An isosurface of 0.2 a.u. is shown in pink while no negative isosurface of -0.2 a.u. is close to the chromophore.

4.1.3 The B form

The construction of a model for the B form of wild-type GFP is more complex as the protein has been only crystallized in the neutral A form (entry 1GFL in the Protein Data Bank). The B form differs from the neutral one not only in the protonation of the chromophore and the Glu-222 residue but also because various residues in the binding pocket of the chromophore have a different, more stable conformations. The X-ray structure is however available for mutants of wild-type GFP where the protein is not in its original sequence but some mutations have been done to enhance the stability of the anionic B form. The environment of these mutants apart from the particular mutations is believed to be closer to the one of wild-type GFP in the B form.

The basic idea in the construction of the model for the B form is to start from the crystallographic structure of one of these mutants and “undo” the mutations to restore the protein sequence as in its wild-type expression. Following the theoretical work by Nifosi and Tozzini [69], we start from the crystallographic structure of mutant S65T (code 1EMG in the Protein Data Bank [72]). This X-ray structure shows two main differences with the wild-

4. Treating the protein environment in QM/MM 4.1. The protein model

type structure as in the 1GFL. First, the S65T substitution is present where the aminoacid serine at position 65 (Ser-65) is substituted with a threonine (Thr-65). Second, the biologic unit is now a monomer.

To restore the wild-type structure, we perform a homology modelling. To restore the aminoacid Ser-65, the methyl group of Thr-65 is substituted by a hydrogen atom, and the resulting Ser-65 oxygen tail is flipped in the other direction towards the close Glu-222 residue to form a hydrogen bond. This structure is then ready for the classic MD simulation for the preliminary equilibration and the subsequent QM/MM dynamics. In Fig. 4.7 we show the hydrogen bond network for the B form of the protein after our homology modeling construction as compared to the neutral and the intermediate anionic form.

Starting from the structure constructed by homology, we proceed as in the case of the neutral form. Using the Amber suite of programs, we add the missing hydrogens, the waters and the counterions. For the B form, we only add 10 Å of TIP3P water molecules in each direction to avoid the occurrence of uninfluential but unpleasant vacuum bubbles at the edges of the simulation cell which were present in the simulation cell of the neutral and intermediate forms. The total charge of the protein is -6 , the number of counterions added is 28 Na^+ and 22 Cl^- , and the box dimensions are approximately $65 \times 79 \times 78 \text{ Å}^3$ for a system of about 31000 atoms. In the B form, the number of atoms in the cell is approximately half the number for the A or I form because here the crystallized form is a monomer. As before, we equilibrate the hydrogens and the waters with a classic MD approach, and we then perform a simulated annealing run with the QM/MM approach. The procedure followed was:

- 1) 1.4 ps of molecular dynamics with $f = 0.999$.
- 2) 0.35 ps of molecular dynamics with $f = 0.997$.
- 3) 0.35 ps of molecular dynamics with $f = 0.995$.
- 4) 0.35 ps of molecular dynamics with $f = 0.991$.
- 5) 0.35 ps of molecular dynamics with $f = 0.987$.

The annealing factor is slowly decreased to allow a smoother quenching of the system. The main hydrogen-bond distances are smoothly dumped during this slow quenching until a temperature less than 0.02 K is reached.

In Fig. 4.7, we show the optimal binding site of the chromophore as obtained in our QM/MM simulation. The hydrogen-bond network surrounding the chromophore is very similar to the one of the I form (see Fig. 4.5) with a chain of hydrogen bonds running from the chromophore to Glu-222 and

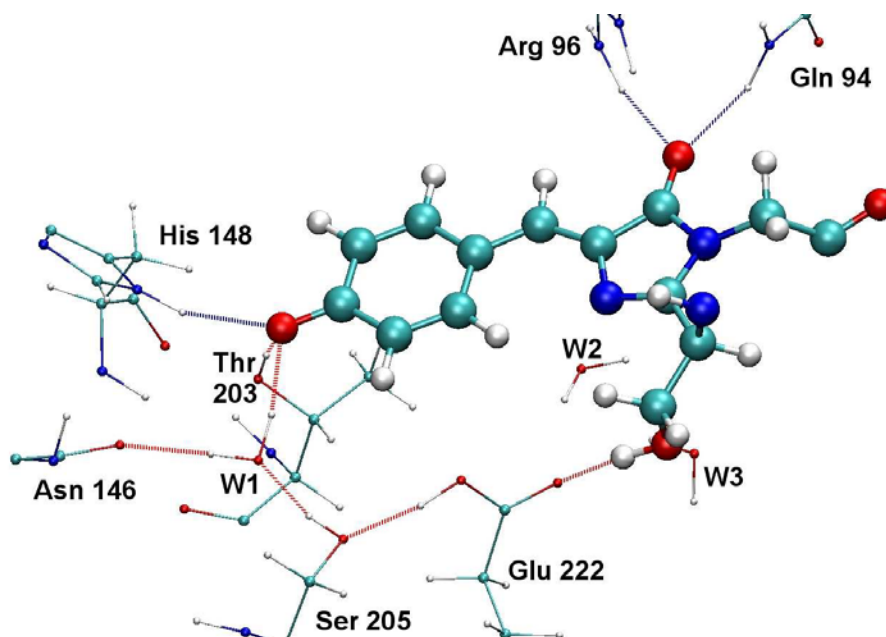


Figure 4.7: Binding site of the anionic B form of GFP. The residues closest to the chromophore are shown. The hydrogen bonds are drawn if the bond length is less than 3 Å and the donor-hydrogen-acceptor angle is less than 30°. The oxygen of the phenolic ring is deprotonated and the Glu-222 residue becomes neutral as the proton acceptor. The position of Thr-203 is different than in the I form and forms a hydrogen bond with the chromophore.

back to the chromophore. In addition, Thr-203 is now positioned to further stabilize the negative charge on the chromophore as it donates a hydrogen bond to the phenolic oxygen. Since this additional hydrogen bond further stabilizes the electronic ground state, the absorption maximum of the B form is expected to be blue shifted with respect to the I form.

4.2 Structural analysis of the models

To analyze the structural features of the models of the neutral and anionic forms of wild-type GFP obtained in our QM/MM calculations, we focus on the binding site of the chromophore as these local properties will predominantly determine its excitation spectrum. For the labeling of the atoms of the chromophore, we refer the reader to Fig. 4.8 where the heavy atoms are numbered starting from the phenolic oxygen along the top ridge of the

phenol, through the bridge and around the imidazole ring.

The internal structure of the chromophore is expected to play a very important role in tuning the excited state properties of the chromophore. In particular, the degree of bond-length alternation in the conjugate chain running through the chromophore is correlated to the size of the gap between the ground state and the lowest π -bonding to antibonding excitation with a stronger bond-length alternation yielding a larger gap. This fact can be easily understood by considering the limiting case of an infinite chain of carbon atoms, where the gap is zero in the absence of bond-length alternation and opens as the chain dimerizes. In the particular case of GFP, the existence of correlation between gap and bond-length alternation has been largely verified through the analysis of several GFP mutants for which the X-ray structure is available, and the construction of simplified theoretical models of the binding site of the GFP chromophore [69].

The protein environment and, in particular, the residues in the binding site of the chromophore can affect the spectral response of the chromophore in a dual manner. They can in principle tune the internal geometrical structure of the chromophore as well as act on the excitations more directly as some close residues may be charged or form hydrogen bond to the chromophore. The hydrogen-bond network in the active site is in fact rather different in the three forms of GFP where residues play different roles in stabilizing either the ground state (opening the gap) or the excited state (closing the gap). The role of the positively charged Arg-96 in close proximity to the GFP chromophore will also be discussed at length below.

We begin with a detailed analysis of the neutral A form of GFP as, only for this form, a comparison with the X-ray structure is possible [71]. The anionic B form is constructed by homology starting from the X-ray structure of the mutant S65T but a direct comparison with the crystallographic data of the mutant would be misleading as the positions of some close residues are rather different. In Table 4.1, we summarize the main structural features of the chromophore of the neutral A form of GFP within the protein and in vacuum, and compare with the X-ray structure at 1.90 Å resolution used as starting geometry. We first note that the structural properties of the chromophore optimized in the protein and in vacuum are remarkably similar. Both chromophores are essentially planar, and the presence of the protein only yields an average bond-length shortening of 0.01 Å without affecting the bond-length alternation of the conjugate chain of the chromophore. When comparing the internal bonds of the chromophore resulting from our QM/MM simulation to the initial X-ray structure, we see that the agreement is rather good with a root mean square deviation of 0.04 Å, and a maximum deviation of only 0.07 Å. The hydrogen-bond lengths between

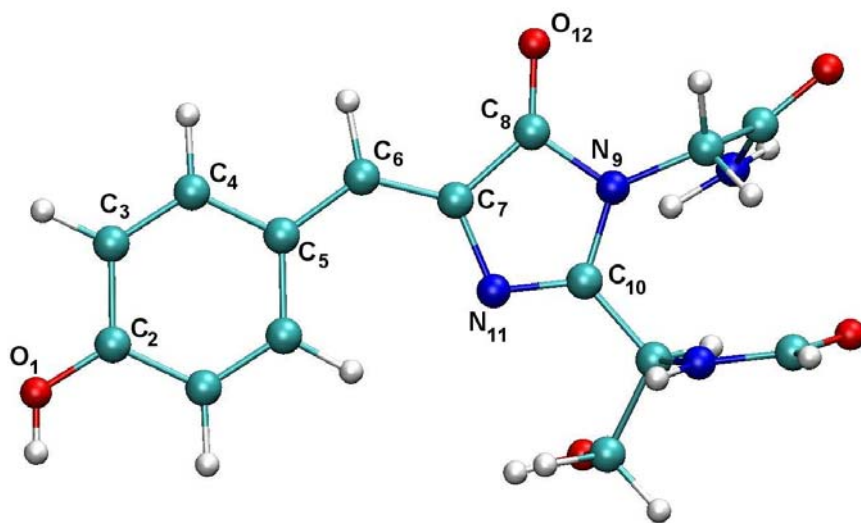


Figure 4.8: Atom numbering used for the chromophore of neutral GFP.

hydrogen donor and acceptor are shown for residues in close proximity to the chromophore and are also found in very good agreement with the X-ray structure.

In Fig. 4.9, we show the bond lengths along the chromophore of the neutral A and the anionic I and B forms of GFP obtained in our DFT/PBE QM/MM calculations and compare them with the structures optimized in vacuum. As in the case of the neutral chromophore, the bond lengths of the anionic I and B forms do not dramatically differ from the values we obtain when the anionic chromophore is optimized in vacuum. Slight differences are observed in the degree of bond alternation close to the central bridge as the difference between the two central bond lengths, C_5-C_6 and C_6-C_7 , is smaller in vacuum than in the protein. Overall, we can however conclude that, for all three forms of wild-type GFP, the protein environment acts to preserve an internal structure of the chromophore which is not dramatically altered with respect to the one in vacuum.

As in the comparison of the gas-phase neutral and anionic chromophores in Chapter 3, it is easier to understand the geometrical changes with the charge state of the chromophore if we show again the two resonant forms of the anionic chromophore in Fig. 4.10. 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

	X-ray	Protein	Vacuum
Bond length (Å)			
O ₁ —C ₂	1.40	1.36	1.37
C ₂ =C ₃	1.39	1.40	1.41
C ₃ —C ₄	1.37	1.38	1.40
C ₄ =C ₅	1.38	1.42	1.43
C ₅ —C ₆	1.40	1.44	1.45
C ₆ =C ₇	1.41	1.37	1.38
C ₇ —C ₈	1.49	1.47	1.50
C ₈ —N ₉	1.32	1.39	1.43
N ₉ —C ₁₀	1.35	1.40	1.39
C ₁₀ =N ₁₁	1.34	1.31	1.33
N ₁₁ —C ₇	1.45	1.40	1.41
C ₈ =O ₁₂	1.22	1.25	1.23
Dihedral angle (°)			
D(C ₄ C ₅ C ₆ C ₇)	173.9	176.8	179.0
D(C ₆ C ₇ N ₁₁ C ₁₀)	178.9	177.1	179.6
Hydrogen-bond length (Å)			
Arg-96(N) ... CHR(O ₁₂)	2.71	2.76	—
Gln-94(O) ... CHR(O ₁₂)	3.08	2.86	—
CHR(O ₁) ... W1(O)	2.64	2.56	—
His-148(N) ... CHR(O ₁)	3.27	3.21	—

Table 4.1: Structural parameters for the neutral A form of GFP obtained in our QM/MM simulations as compared to the X-ray structure used as initial configuration. We list the most representative bond lengths and dihedral angles of the chromophore, and the hydrogen-bond distances between the hydrogen donor and acceptor (D ... A) for close residues to the chromophore (CHR). We also list the values for the chromophore optimized in vacuum. See Fig. 4.8 for the labeling of the atoms of the chromophore. For reference, the bonds of the central carbon bridge are C₅—C₆ and C₆=C₇.

the imidazole oxygen.

When comparing the neutral and the anionic forms, we see that the largest difference occurs in proximity of the hydroxyl oxygen of the phenolic ring. After deprotonation of the hydroxyl group, the oxygen-carbon bond, O₁—C₂, loses its single-bond character and significantly shortens by about 0.1 Å when going from the neutral to the anionic form. The shortening is slightly

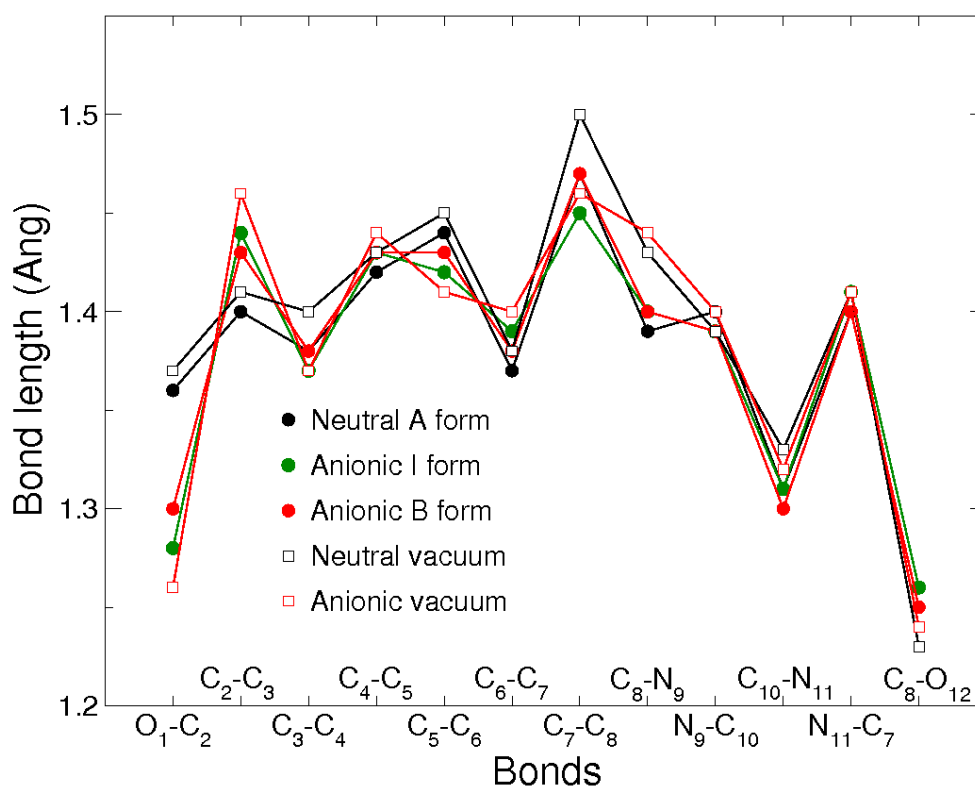


Figure 4.9: Structural properties of the chromophore of the neutral A, and the anionic I and B forms as obtained in our QM/MM simulations. We also show the neutral and anionic structures optimized in vacuum within all-electron DFT/BLYP with a cc-pVTZ basis. The bonds of the central carbon bridge are C_5-C_6 and C_6-C_7 .

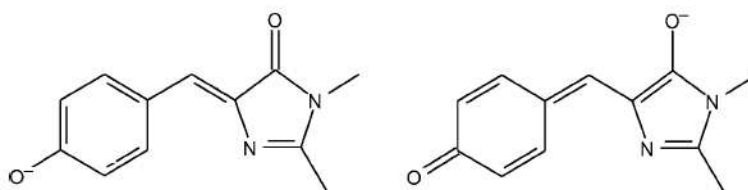


Figure 4.10: Scheme of the two resonant forms of the anionic chromophore: Benzenoid (left) and quinonoid (right).

smaller in the B form as compared to the I form as the negative charge on the oxygen is further stabilized in the B form by the position of Tyr-203, which donates an additional hydrogen bond to the oxygen. As a consequence, while the neutral model is characterized by a more marked aromaticity of the phenolic ring (with more similar bond lengths between all carbon atoms of the ring), the aromaticity of the phenolic ring of the anionic forms is reduced, yielding a quinoid structure. The effect of deprotonation is also visible for the bonds in proximity of the central bridge of the chromophore where the degree of bond alternation is reduced in the anionic as compared to the neutral form. Interestingly, we note that the difference between the neutral and the anionic bond lengths is overall larger in vacuum: For instance, the two central bonds of the carbon bridge in the anionic species have almost the same lengths, the difference being less than 0.01 Å, while in the neutral species, these two bonds differ by as much as 0.07 Å. In the protein, the same differences for the I and the neutral form amount to 0.04 and 0.07 Å, respectively. By inspecting the other bonds along the conjugated chain, it appears that the protein environment acts to partially compensate the change in protonation state, and keeps the chromophore in a more similar structural conformation in the two protonation states with respect to vacuum.

In Fig. 4.12, we compare the geometrical properties of the chromophore of the three forms of GFP with the results of other simulations available in the literature. We first focus on the comparison with the work by Marques *et al.* [12] who construct their protein models within a DFT/LDA QM/MM approach. It is important to understand the structural features of their models as they most significantly differ from ours as well as from other calculations reported in the literature. Moreover, we will show later that, using their DFT/LDA QM/MM structures, Marques *et al.* can claim an excellent agreement between TDDFT and the experimental absorption spectra of wild-type GFP.

For the neutral A form, the structural agreement of our model with the DFT/LDA calculations by Marques *et al.* [83] is reasonable while, for the anionic I form, the models are significantly different. In the neutral form, Marques *et al.* find a slightly more marked bond-length alternation along the chromophore, and the same structure is essentially preserved when moving to their anionic I form, with the exception of the oxygen-carbon bond of the phenolic group, O₁—C₂, which is significantly shorter in the I form in agreement with our calculations. In particular, the carbon bond alternation in the central carbon bridge is 0.1 Å in the neutral form and 0.09 Å in the I form, compared to 0.07 Å and 0.04 Å in our calculations. Another evident difference with our results is the bond length of the subsequent single carbon bond of the imidazole ring, C₇—C₈. In their LDA/DFT model, this bond

length is longer by 0.05 Å and 0.07 Å than the values we obtain in the neutral and anionic I form, respectively.

The difference between our model and the one by Marques *et al.* cannot be attributed to the use of LDA instead of the generalized gradient approximation PBE as the exchange and correlation functional employed when relaxing the chromophore geometry. We showed in our studies of model chromophores in vacuum that the use of LDA yields essentially the same structural parameters as PBE, BLYP or B3LYP (see Fig. 3.3). A closer analysis of the protein structures of Marques *et al.* reveals instead that the different geometry of the chromophore is likely due to particular choices made in the setup of the protein environment. The most striking feature in their models is that *all* histidine protein residues in both the neutral and I forms are protonated both at the δ and ϵ nitrogens, and carry therefore a positive net charge. As most histidines are far from the chromophore site, their charge state will not significantly affect the chromophore geometry. However, His-148 is close to the chromophore and even creates a hydrogen bond with the chromophore in the anionic state (Figs. 4.5 and 4.7). Therefore, since protonating both nitrogen of His-148 corresponds to placing an additional positive charge in close proximity to the chromophore, we may expect a noticeable effect on the structural parameters of the system. For clarity, we note a rather misleading feature of the work by Marques *et al.* as the structural coordinates made available by the authors do not correspond to the figure which appears in their paper where His-148 is drawn as only protonated at the δ nitrogen. We also note that Marques *et al.* construct the I form by deprotonation of the neutral form but always refer incorrectly to this structure as the B form.

To proceed, we need to motivate why we believe our model to be a closer representation of wild-type GFP in its various protonation forms than the one by Marques *et al.*. Our model corresponds to one accepted in the literature and its validity is mostly supported by the X-ray characterization of the structures of wild-type GFP and other mutants. The β barrel in GFP appears to be partially perturbed around the phenolic end of the chromophore in proximity to His-148. The β strand that covers the chromophore moves around His-148 and the backbone of one strand from residue 144 to 150 is not directly hydrogen bonded to the adjacent backbone between residues 165 and 170. The two backbones appear instead to be held together by forming hydrogen bond with the imidazole ring of His-148, with the backbone nitrogen of Arg-196 being a donor to nitrogen ϵ of His-148.

Accordingly, as shown in Fig. 4.11, we do not protonate the nitrogen ϵ of His-148 but allow it to be a hydrogen-bond acceptor for the backbone nitrogen of Arg-168. With this assumption, we obtain a structure for the neutral form with an Arg-168(N) $\cdots$ His-148(N) distance of 3.05 Å in close

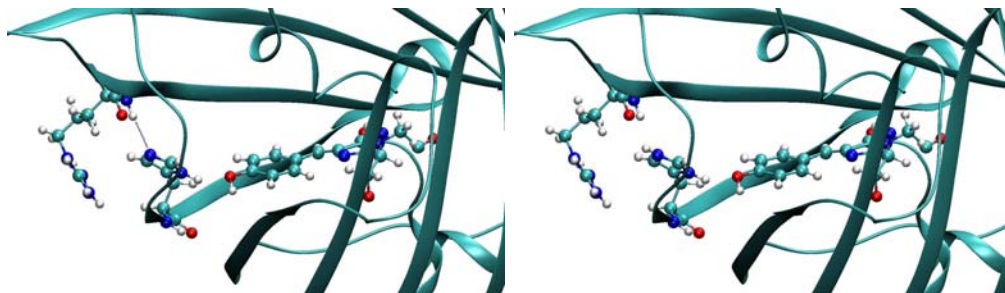


Figure 4.11: Particular of the structure of the I form for our model (left) and the one by Marques *et al.* [12] (right). The chromophore, the residues His-148 and Arg-168, and part of the β barrel are shown. Note the difference in the protonation of the nitrogen ϵ of His-148 which, in our model, is a hydrogen-bond acceptor for the backbone nitrogen of Arg-168.

agreement with the X-ray value of 2.94 Å. On the other hand, Marques *et al.* find a significantly larger nitrogen-nitrogen distance of 3.40 Å since their His-148 is protonated at both nitrogens and cannot be a hydrogen bond acceptor for Arg-168. Moreover, as shown in Table 4.2, the positive His-148 in the models by Marques *et al.* is significantly closer to the chromophore with a His-148(N) $\cdots$ CHR(O₁) distance which is 0.4 Å shorter than the value in our as well as the X-ray structure. Finally, the bond lengths along the chromophore are significantly closer in our model to the X-ray values than the ones computed by Marques *et al.* We therefore believe that the overall better agreement of our structure with the crystallographic data is strong evidence that our model is most likely a closer representation of the real structure of GFP than the one where His-148 is protonated at both nitrogens.

In Fig. 4.12, we also compare the structure of the chromophore that we obtain in the three forms of GFP with other models available in the literature. We find that the structures of our chromophore for the neutral A and anionic B forms are in close agreement with the simulations denoted as (MM + QM) by Laino *et al.* [81]. These authors construct the missing force field for the chromophore to perform a first MM relaxation of the protein and, subsequently, refine the structural parameters by performing a DFT/BLYP optimization of the chromophore binding site only keeping the close residues with fixed heavy atoms at the boundary. The structure of the chromophore in our model is remarkably similar to the one by Laino *et al.* [81] for both the neutral and the anionic B form, and the hydrogen-bond distances of the chromophore to the closeby residues are also in reasonable agreement (see Table 4.2). A detailed comparison with the hydrogen-bond distances of the (MM + QM) model is however not proper as only few residues were kept

4.2. Structural analysis 4. Treating the protein environment in QM/MM

in their QM optimization. Even though all distances listed in Table 4.2 do not significantly change in the QM refinement, an exception is the Thr-203(O)⋯CHR(O₁) distance which in their MM simulation is 2.80 Å but becomes significantly larger, 3.28 Å, in the subsequent partial QM optimization.

Finally, our results for the anionic I and B forms are in reasonable agreement with the CASSCF/MM calculations by Sinicropi *et al.* [13] who relax the chromophore structure within CASSCF together with the position and the orientation of three closeby classical waters. The structure of the remaining of the protein in the neutral A form is kept to the original crystallographic coordinates while the I form is obtained starting from the A form and manually reorienting the residues Ser-205 and Glu-222 to form the expected hydrogen bond network (see Fig. 4.5). The B form is derived from the I form by relaxing Thr-203 in a conformation to form hydrogen bond with the phenolic oxygen of the chromophore (see Fig. 4.7). Despite the simplicity of their QM/MM embedding scheme and lack of complete relaxation, the few available hydrogen-bond distances are in agreement with the values we find, as shown in Table 4.2. For the chromophore, we find that the degree of bond-length alternation is slightly larger in CASSCF as compared to the DFT calculations, but the overall agreement is rather good.

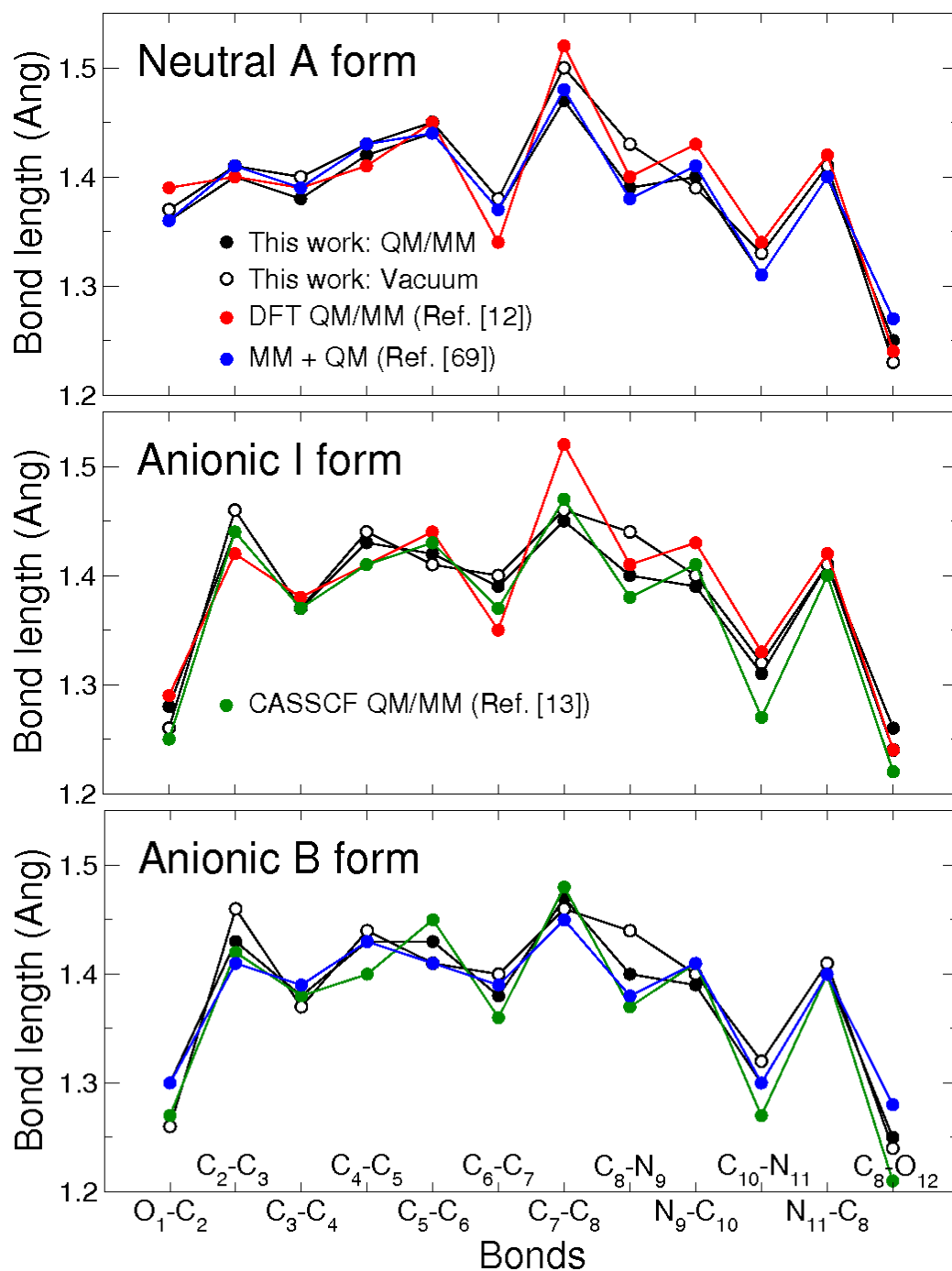


Figure 4.12: Bond lengths of the chromophore of the A, I, and B forms of GFP. We compare our QM/MM results to DFT/LDA [12] and CASSCF QM/MM [13] simulations, and to MM simulations followed by partial QM relaxation [69]. The neutral and anionic geometries optimized in vacuum are also shown. The bonds of the central bridge are C₅—C₆ and C₆—C₇.

	Arg-96(N) ⋮ CHR(O ₁₂)	Gln-94(N) ⋮ CHR(O ₁₂)	CHR(O ₁) ⋮ W1	His-148(N) ⋮ CHR(O ₁)	Thr-203(O) ⋮ CHR(O ₁)
Neutral A form					
This work	2.76	2.86	2.56	3.21	—
DFT QM/MM	2.74	3.18	2.77	2.81	—
MM + QM	2.71	3.05	2.61	N/A	—
X-ray	2.71	3.08	2.64	3.22	—
Anionic I form					
This work	2.77	2.85	2.78	2.96	—
DFT QM/MM	2.72	3.31	2.83	2.67	—
CASSCF QM/MM	N/A	N/A	2.71	N/A	—
Anionic B form					
This work	2.76	2.95	2.79	3.02	2.76
CASSCF QM/MM	N/A	N/A	2.69	N/A	N/A
MM + QM	2.63	3.00	2.75	2.80	2.80 ^a

^a 2.80 Å is the MM while 3.28 Å the (MM + QM) value [82].

Table 4.2: Most representative hydrogen-bond distances in Å between the hydrogen donor and acceptor (D ⋯ A) for close residues to the chromophore (CHR). We compare our DFT/PBE QM/MM results to other models available in the literature as the DFT/LDA [12] and CASSCF QM/MM [13] simulations and the MM simulations followed by partial DFT/BLYP relaxation (QM+MM) [69]. Note that, for the bond CHR(O₁) ⋯ W1, the chromophore is a donor in the neutral A form and an acceptor in the I and B forms, and that Thr-203(O) only forms a hydrogen bond to the chromophore in the B form.

4.3 TDDFT/MM absorption spectra

We compute the vertical low-lying singlet excited states of the chromophore of GFP in the protein environment for the structures of the A, I and B forms presented in the previous Section using a TDDFT/MM approach and the Gaussian03 code [53]. We perform all-electron linear-response TDDFT calculations in the presence of the external field of the classical protein environment. Within Gaussian03, the MM atoms are treated as unscreened point charges which we place at the positions corresponding to the optimal geometry that we obtained in our QM/MM simulations. For the MM atoms, we use the same partial charges as in the QM/MM simulation. We employ the BLYP and B3LYP exchange-correlation functionals and a cc-pVTZ Gaussian basis set. We always compute at least the lowest five solutions as the state with the largest oscillator strength is not always the lowest.

The linear-response TDDFT results for wild-type GFP in its three forms are summarized in Table 4.3. We compute two sets of excitation energies corresponding to the chromophore in the protein environment (Protein) and to the same chromophore without the surrounding protein (Vacuum). Comparing the excitation energies with and without the protein environment allows us to access the polarization effects of the protein on the electronic states of the chromophore. Moreover, a comparison with the excitations for the chromophore geometry optimized in vacuum (see Chapter 3) reveals how the changes in the structure of the chromophore due to the presence of the protein affects the excitation energies. In all cases, we list the lowest two singlet excitations, minus the Kohn-Sham energy of the highest occupied molecular orbitals (HOMO) as the DFT ionization continuum begins at minus the value of the HOMO orbital energy [84] and the Kohn-Sham gap between the lowest unoccupied and the highest occupied molecular orbitals.

Comparison with experiments

We first give a general look at the results focusing on the comparison with the experimental absorption maxima [85,86]. For all three forms and both exchange-correlation functionals, the excitation energy with the largest oscillator strength is the lowest singlet state with a dominant $\pi \rightarrow \pi^*$ (HOMO $\rightarrow$ LUMO) character. For the neutral A form, the BLYP and the B3LYP excitation energies in the protein nicely bracket the experimental value with BLYP giving an energy 0.03 eV lower and B3LYP 0.18 eV higher than the experiments. The oscillator strength (about 0.6) and the HOMO $\rightarrow$ LUMO contribution to the character of the excitation is comparable when the BLYP and B3LYP functionals are used. The ionization threshold for both function-

als is significantly higher than the excitation energies, not to pose a problem. The effect of polarization by the protein environment is small and only amounts to a red shift of about 0.04 eV in the excitation energy. We may therefore regard this result as a success of linear-response TDDFT in describing the experimental spectrum of the neutral form of GFP but we will return to this point later when interpreting the TDDFT results for all three forms of wild-type GFP.

The TDDFT excitations for the anionic forms of GFP are instead not in good agreement with the experimental values. For the I form, BLYP and B3LYP overestimate the experimental excitation energy by 0.37 and 0.55 eV, respectively, yielding excitation energies not too dissimilar to the neutral case. By comparing with the excitations computed in the absence of the protein, we note that the inclusion of the protein environment yields a blue shift of 0.05 and 0.18 eV for the BLYP and B3LYP functionals, respectively, that is, a shift in the opposite direction than for the neutral form. More importantly, we observe that the inclusion of the protein significantly raises the DFT ionization threshold so that the excitations in the protein are 1.06 and 1.68 eV lower than the BLYP and the B3LYP ionization threshold, respectively. Therefore, while the meaning of the excitation energies in vacuum is questionable as they lie above the ionization threshold, the problem represented by the general underestimation of the ionization threshold in DFT disappears when the protein is included.

For the B form, the agreement of TDDFT with experiment is equally poor as for the I form. TDDFT/BLYP and B3LYP overestimate the experimental value by 0.30 and 0.49 eV, respectively. Also for the B form, the inclusion of the protein environment yields a blue shift 0.14 and 0.45 eV for the BLYP and B3LYP functionals, respectively. Differently from the case of the I form, only the use of the B3LYP functional yields an excitation in the protein well below the excitation threshold while BLYP gives an excitation in the protein which is 0.42 eV higher than the BLYP ionization threshold. Moreover, within TDDFT/BLYP, the two lowest singlet excitations are nearly degenerate with a comparable oscillator strength (0.43 and 0.32) and character of the excitations. This degeneracy is lifted when using the B3LYP functional which moves the second singlet 0.57 eV higher than the lowest singlet. For both anionic forms, the use of the B3LYP functional always yields a lowest state with a higher oscillator strength and a more definite single-excitation (HOMO $\rightarrow$ LUMO) character than when the BLYP functional is employed.

On the positive side, we note that TDDFT correctly predicts the expected blue shift in going from the I to the B form. In the B form, the ground state is further stabilized with respect to the excited state by the additional hydrogen bond of the phenolic oxygen of the chromophore with Thr-203, and one would

expect a larger excitation energy than in the I form. In particular, BLYP and B3LYP predict the excitation energy of the B form to be higher than the I form by 0.06 and 0.07 eV, respectively. With the available experimental data, we can only estimate an upperbound (0.13 eV) to the difference between the excitations of the B and the I form: For the I form, the location of the absorption maximum is not available and we therefore use the energy of the 0-0 transition, which is smaller than the absorption maximum by an amount equal to the Stokes' shift.

Finally, we note that the mixed agreement we find between TDDFT and experiments is at odds with the good performance observed by Marques *et al.* [12] who obtain 3.05 eV and 2.65 eV as the TDDFT/LDA absorption maximum for the A and the I form, respectively. We believe that the positive picture emerging from their work is coincidental and due to the fact that their model for the I form is different from our structure as discussed in Section 4.2. For both the neutral A and the anionic I form, they compute the TDDFT excitations of the protein chromophore without the protein environment. As their chromophore for the A form is not too dissimilar from our structure, we know from our calculations that computing the excitation without the protein environment will indeed yield a TDDFT excitation very close to the experimental value. On other hand, the chromophore of their I form is significantly different from our geometry due to how they protonate all histidine residues in the protein. Apparently, this geometry in vacuum gives a TDDFT excitation close to the experimental value for the I form. For clarity, we point out again that even though they build the I form by deprotonation of the neutral form, they refer to their structure as the B form and therefore find a particularly good agreement with experiments as they compare with the absorption maximum of the B form (2.63 eV).

Geometrical and electrostatic effects of the protein environment

In Table 4.4, we summarize the TDDFT excitation spectra for the chromophore optimized in the gas phase and in the protein environment for the three forms of wild-type GFP, and compute the spectral shift, $\Delta E_{\text{protein}}$, due to the interaction of the chromophore with the protein. The spectral shift can be decomposed in two contributions due to the geometrical change in the chromophore geometry and to the electrostatic effect of the environment on the excitation:

$$\Delta E_{\text{protein}} = \Delta E_{\text{geom}} + \Delta E_{\text{elect}} .$$

The geometrical shift (ΔE_{geom}) is computed as the difference of the excitation of the protein chromophore without the protein environment and the

excitation of the chromophore optimized in the gas phase. The electrostatic shift (ΔE_{elect}) is obtained as the difference of the excitation energies of the protein chromophore with and without the protein environment.

For the neutral A form, BLYP and B3LYP yield a total red shift due to the chromophore-protein interaction of -0.08 and -0.19 eV, respectively. For both functionals, the shift due to the electrostatic interaction with the MM atoms is very small (less than -0.04 eV) while B3LYP sees a larger red shift due to geometrical changes than the one obtained with the BLYP functional. As we have seen in the previous Section, the degree of bond alternation for the neutral and the anionic forms is more marked for the chromophore optimized in vacuum than in the protein. As expected, this translates in a red shift in the excitation energies when going from the gas-phase chromophore to the protein chromophore, a shift which is significantly larger (-0.16 eV) when using the B3LYP functional.

For the anionic I form, BLYP and B3LYP predict no protein shift (0.01 eV) as the geometrical red shift and the electrostatic blue shift cancel. The reference protein shift of -0.09 eV is estimated as the difference between the energy of the 0-0 transition in the protein [7] and the absorption maximum of a smaller chromophore in the gas phase [58]. Therefore, it represents a lowerbound to the true shift, as the excitation of a larger chromophore in the gas phase (with the same dimensions as the one optimized in the protein) will be smaller than the excitation of the smaller model chromophore, while the absorption maximum of the I form is larger than its 0-0 transition. In the anionic B form, BLYP and B3LYP yield comparable small blue shifts of 0.07 and 0.08 eV, respectively. For both functionals, the geometrical red shift is very close to the values obtained for the anionic I form since the I and B chromophores have very similar geometries. However, due to the further stabilization of the ground state in the B form discussed above, the electrostatic blue shift is larger in the B than the I form, yielding a net protein blue shift. The experimental estimate of the shift is 0.04 and again represents a lowerbound as we are using the absorption maximum for a smaller model chromophore to estimate the excitation in the gas phase.

Understanding the mixed performance of TDDFT

The mixed success of TDDFT, which appears to be accurately describing the excitation of the neutral form but not of the anionic forms of wild-type GFP, may be due to several factors which we now analyze separately. The first obvious culprit is the use of TDDFT in the adiabatic approximation and we will therefore try to understand if any of its commonly known problems can also affect our results. However, we need to keep in mind that, whether

there is a problem with TDDFT, it must only affect the anionic and not the neutral form as the neutral appears to be correctly described by adiabatic TDDFT.

We begin with the underestimation by DFT of the ionization threshold and, in Table 4.3, list for all systems minus the Kohn-Sham energy of the HOMO orbital, which indicates the start of the TDDFT continuum. We observe that for the neutral chromophore with and without the protein environment, the ionization threshold is well above the lowest excitation energy and therefore does not pose a problem. On the other hand, as in the case of the anionic model chromophores in the gas phase (Chapter 3), the TDDFT excitation energies of the I and B anionic chromophores in vacuum lie above the ionization threshold and their meaning is therefore questionable. The inclusion of the protein environment significantly raises the ionization threshold of the I form above the lowest singlet excitation energy while, for the B form, only the use of the B3LYP functional brings the ionization threshold above the relevant excitation energy. However, even though the TDDFT continuum is now higher than the excitation, TDDFT overestimates the experimental excitation energies of the I and B forms by 0.4-0.5 eV. We also note that the TDDFT error in the protein is comparable to the error in the gas phase for the smaller anionic chromophore, where TDDFT overestimates the excitation by 0.26 and 0.46 eV when using BLYP and B3LYP, respectively. Therefore, the position of the ionization threshold does not appear a relevant factor in understanding why the performance of TDDFT is significantly poorer in describing the anionic than the neutral form of GFP.

Adiabatic TDDFT is also known to perform rather poorly if the excitations are characterized by significant charge transfer. To understand whether this is a problem, we can compare the TDDFT excitations and the Kohn-Sham eigenvalue differences since these two quantities will become equal if the excitation is characterized by charge transfer. We note that this analysis holds when no exact exchange is used in the functional as only in this case the eigenvalue differences are a close representation of excitation energies since all orbitals see the same potential and therefore a constant number of electrons. In the presence of exact exchange, the virtual orbitals see a different potential and a different number of electrons than the occupied ones, and for instance the HOMO-LUMO gap will be closer to the difference between ionization potential and electron affinity.

In the three forms of GFP, the singlet TDDFT excitations with the largest oscillator strength are the lowest in energy and have a dominant HOMO $\rightarrow$ LUMO character, and should therefore be compared with the Kohn-Sham HOMO-LUMO gaps which are also listed in Table 4.3. For the neutral A and anionic I forms, the BLYP HOMO-LUMO gap is always smaller by about

1 eV than the lowest TDDFT excitation. Therefore, since TDDFT significantly corrects the Kohn-Sham gap, we can infer that the excitations are not characterized by strong density transfer. For the B form, the comparison with the BLYP Kohn-Sham gap is not as straightforward since the two lowest excitations are nearly degenerate at 2.93 and 2.94 eV, and dominated by the (HOMO-3) $\rightarrow$ LUMO contribution even though the HOMO $\rightarrow$ LUMO transition is rather large: The HOMO-LUMO BLYP gap is 1 eV smaller than the lowest excitation while the (HOMO-3)-LUMO eigenvalue difference is equal to 2.90 eV and therefore very close to the excitation energy. This could be a sign of potential problems of the TDDFT description of the B form. In Fig. 4.13, we also show the difference between the ground state density and an estimate of the TDDFT excited state density for the neutral A and the anionic I form with and without the protein environment. The anionic excited state density computed without the protein environment appears to extend in the tails of the chromophore as compared to the ground state one. However, this slight charge transfer to the tails disappears when the protein environment is included and no striking difference can be observed between the neutral and the anionic case.

In summary, the indicators given by the position of the excitation with respect to minus the HOMO Kohn-Sham eigenvalue and to the Kohn-Sham eigenvalue difference are rather similar in the neutral and the anionic forms (in particular the I form) of wild-type GFP and it is therefore not evident why TDDFT should give an excellent agreement with experiments for the neutral form but not the anionic forms. A possible reason for the failure of TDDFT could be that the excitation has a significant double- or higher-excitation character in the case of the anion but not in the neutral form. We come back to this point when analyzing the quantum Monte Carlo results in the next Section as these results seem to indicate that the character of the excitation is rather similar in the two cases. Finally, it is important to stress that, when computing the excitation energies with a QM/MM scheme, the MM atoms only affect the electronic states of the QM part via an electrostatic polarization field. Therefore, all the residues which are hydrogen bonded to the chromophore cannot in reality forming a proper “bond”. In Section 4.5, we analyze the effect of extending the QM part of the simulation to include relevant closeby residues.

Figure 4.13: Difference between the DFT/B3LYP ground and excited state densities for the protein chromophore of the neutral A form (top) and the anionic I form (bottom) without (left) and with (right) the protein environment. The isosurface corresponds to a value of -0.001 in red and $+0.001$ in blue. A negative (red) value corresponds to the excited state density being larger than the ground state one.

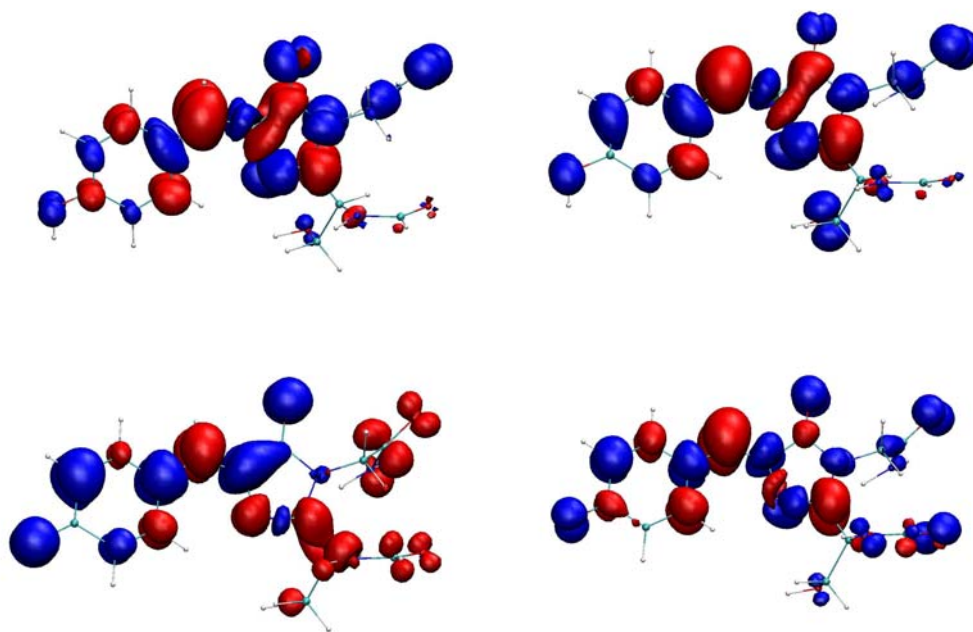


Table 4.3: TDDFT/BLYP and B3LYP excitation energies (eV) and oscillator strengths (in parenthesis) for the A, I, and B forms of GFP with (Protein) and without (Vacuum) the protein environment, computed using a cc-pVTZ basis. The geometries are optimized in the presence of the protein. The dominant electronic transitions and their contributions in parenthesis (if > 0.1) are also listed. Only the lowest two singlet excitations are given together with minus the Kohn-Sham energy of the highest occupied molecular orbitals ($-\epsilon_{\text{HOMO}}$), which corresponds to the ionization threshold in DFT. Also the Kohn-Sham gap between the lowest unoccupied and the highest occupied molecular orbitals ($\Delta\epsilon_{\text{HL}}$) is listed.

	Vacuum		Protein	
	BLYP	B3LYP	BLYP	B3LYP
Neutral A form (Expt. 3.05 eV)				
$S_0 \rightarrow S_1$	3.07(0.63)	3.26(0.58)	3.02(0.65)	3.23(0.64)
	H $\rightarrow$ L(0.47)	H $\rightarrow$ L(0.58)	H $\rightarrow$ L(0.45)	H $\rightarrow$ L(0.58)
	H-1 $\rightarrow$ L(0.30)	H-1 $\rightarrow$ L(0.19)	H-1 $\rightarrow$ L(0.23)	H-2 $\rightarrow$ L(0.20)
	H-4 $\rightarrow$ L(0.12)	H-2 $\rightarrow$ L(0.12)	H-3 $\rightarrow$ L(0.21)	
$S_0 \rightarrow S_2$	3.19(0.02)	3.66(0.22)	3.14(0.01)	3.48(0.18)
	H-3 $\rightarrow$ L(0.69)	H-2 $\rightarrow$ L(0.67)	H-4 $\rightarrow$ L(0.69)	H-2 $\rightarrow$ L(0.66)
	H-4 $\rightarrow$ L(0.12)	H $\rightarrow$ L(0.15)		H $\rightarrow$ L(0.17)
$\Delta\epsilon_{\text{HL}}$	2.10	3.36	2.05	3.29
$-\epsilon_{\text{HOMO}}$	5.10	6.01	5.57	6.47
Anionic I form (Expt. ≈ 2.50 eV)				
$S_0 \rightarrow S_1$	2.82(0.60)	2.87(0.67)	2.87(0.71)	3.05(0.96)
	H $\rightarrow$ L(0.43)	H $\rightarrow$ L(0.54)	H $\rightarrow$ L(0.52)	H $\rightarrow$ L(0.58)
	H $\rightarrow$ L+2(0.29)	H $\rightarrow$ L+1(0.31)	H-2 $\rightarrow$ L(0.20)	
	H $\rightarrow$ L+1(0.27)			
$S_0 \rightarrow S_2$	3.03(0.12)	3.19(0.28)	3.08(0.12)	3.48(0.001)
	H $\rightarrow$ L+2 (0.63)	H $\rightarrow$ L+1(0.62)	H-2 $\rightarrow$ L(0.60)	H-1 $\rightarrow$ L(0.68)
	H $\rightarrow$ L(0.16)	H $\rightarrow$ L+2(0.15)	H-4 $\rightarrow$ L(0.19)	H-2 $\rightarrow$ L(0.14)
	H $\rightarrow$ L+4(0.13)		H-5 $\rightarrow$ L(0.19)	
$\Delta\epsilon_{\text{HL}}$	1.71	2.79	1.82	2.92
$-\epsilon_{\text{HOMO}}$	1.05	1.79	3.93	4.73

Continues on the next page

Continues from the previous page

	Vacuum		Protein	
	BLYP	B3LYP	BLYP	B3LYP
Anionic B form (Expt. 2.63 eV)				
$S_0 \rightarrow S_1$	2.79(0.57)	2.87(0.75)	2.93(0.43)	3.12(0.90)
	H $\rightarrow$ L(0.43)	H $\rightarrow$ L(0.55)	H-3 $\rightarrow$ L(0.46)	H $\rightarrow$ L(0.59)
	H $\rightarrow$ L+2(0.30)	H $\rightarrow$ L+1(0.25)	H $\rightarrow$ L(0.39)	
	H $\rightarrow$ L+1(0.26)		H-1 $\rightarrow$ L(0.14)	
	H $\rightarrow$ L+5(0.10)		H $\rightarrow$ L+1(0.11)	
$S_0 \rightarrow S_2$	3.01(0.13)	3.23(0.18)	2.94(0.32)	3.69(0.04)
	H $\rightarrow$ L+2(0.62)	H $\rightarrow$ L+1(0.65)	H-3 $\rightarrow$ L(0.53)	H-1 $\rightarrow$ L(0.67)
	H $\rightarrow$ L (0.17)	H $\rightarrow$ L(0.16)	H $\rightarrow$ L(0.34)	H-3 $\rightarrow$ L(0.10)
	H $\rightarrow$ L+3(0.15)		H-1 $\rightarrow$ L(0.11)	
$\Delta\epsilon_{\text{HL}}$	1.69	2.76	1.92	3.05
$-\epsilon_{\text{HOMO}}$	1.06	1.81	2.51	3.31

Table 4.4: TDDFT/BLYP and B3LYP excitation energies in eV computed for the protein chromophore without ($\text{CHR}_{\text{protein}}$) and with (GFP) the inclusion of the protein environment, and for the chromophore optimized in the gas phase (CHR_{gas}). The spectral shifts due to the geometrical changes in the chromophore ($\Delta E_{\text{geom}} = \text{CHR}_{\text{protein}} - \text{CHR}_{\text{gas}}$) and to the electrostatic effect of the environment ($\Delta E_{\text{elec}} = \text{GFP} - \text{CHR}_{\text{protein}}$) are listed together with the total protein shift, $\Delta E_{\text{protein}} = \Delta E_{\text{geom}} + \Delta E_{\text{elec}}$.

	CHR_{gas}	$\text{CHR}_{\text{protein}}$	GFP	ΔE_{geom}	ΔE_{elec}	$\Delta E_{\text{protein}}$
Neutral A form						
BLYP	3.10	3.07	3.02	-0.03	-0.05	-0.08
B3LYP	3.42	3.26	3.23	-0.16	-0.03	-0.19
Expt.	—	—	3.05	—	—	—
Anionic I form						
BLYP	2.86	2.82	2.87	-0.04	0.05	0.01
B3LYP	3.04	2.87	3.05	-0.17	0.18	0.01
Expt.	≈ 2.59	—	2.50	—	—	-0.09
Anionic B form						
BLYP	2.86	2.79	2.93	-0.07	0.14	0.07
B3LYP	3.04	2.87	3.12	-0.17	0.25	0.08
Expt.	≈ 2.59	—	2.63	—	—	0.04

4.4 QMC/MM excitation energies

We now compute the vertical excitations of the three forms of wild-type GFP in quantum Monte Carlo. As for the chromophore models in vacuum, we only compute the bright $\pi \rightarrow \pi^*$ (HOMO $\rightarrow$ LUMO) transition, which has the largest oscillator strength and therefore corresponds to the maximum absorption of GFP. The QMC excitations are given by the difference of the excited state S_1 and the ground state S_0 total energy, which are now computed in the presence of the MM protein environment. To perform a quantum Monte Carlo calculation in the presence of the protein environment, we need to include the environmental effects in the actual QMC calculation as well as in the construction of the starting many-body wave function.

The inclusion of a classical MM environment in the QMC calculation is straightforward as it amounts to include an additional external potential which we simply represent on a grid encompassing the QM component. The potential is computed using the positions of the MM atoms as obtained in the DFT/PBE QM/MM calculation to setup the protein model, and the partial charges of the MM atoms are screened as in the CPMD code. For the setup of the determinantal component of the QM wave function, we employ the GAMESS code which allows us to include the MM atoms as screened point charges via the effective fragment potential module of GAMESS. We only consider the MM atoms which lie within $R_c = 12 \text{ \AA}$ of the QM/MM boundary and leave the positions of the MM atoms at the optimal coordinates of the DFT/PBE QM/MM calculation. As GAMESS only allows a particular way to screen the point charges which is different than the one in the CPMD code, we verify below that the use in QMC of the MM potential compatible with the screening in GAMESS yields equivalent QMC/MM excitation energies than when using the CPMD screening.

All QMC calculations are performed with Hartree-Fock semi-relativistic energy-consistent pseudopotentials specifically constructed for use in QMC, and the corresponding cc-pVDZ basis sets [67]. We use Jastrow-Slater wave functions where the determinantal component is determined within GAMESS in the presence of the MM atoms and the Jastrow correlation factor is subsequently optimized by energy minimization using the Hamiltonian with the additional potential due to the MM environment. In the Jastrow factor, we only include electron-electron and electron-nucleus correlation terms as the additional electron-electron-nucleus terms would significantly slow down the calculations. Moreover, in Chapter 3, we have seen for the small anionic model chromophore that neglecting the three-body terms in the Jastrow factor does not significantly affect the DMC excitation energies. For the determinantal components of the ground and excited states, we use the four

Table 4.5: Vertical excitation energies (eV) for the neutral A and the anionic I and B forms of wild-type GFP computed within VMC, DMC, and TDDFT/BLYP and B3LYP. The experimental energies correspond to the absorption maxima for the A and B forms and the 0-0 transition for the I form. The CASPT2 results are from Ref. [13]. For the QMC results, the statistical error on the last figures is given in parenthesis.

	A form	I form	B form
Excitation energy (eV)			
Expt.	3.05	2.50	2.63
VMC	3.96(09)	3.34(09)	3.71(9)
DMC	3.66(11)	3.16(11)	3.15(12)
TDDFT/BLYP	3.02	2.87	2.93
TDDFT/B3LYP	3.23	3.05	3.12
CASPT2 [13]	–	2.65	2.81

determinants resulting from a two-state SA-CASSCF(2,2) calculation with equal weights. Again, this choice for the wave function is both motivated by the necessity to reduce the computational demands of the calculation, and by the observation for the small anionic model chromophore that the inclusion of a larger active space as well as the reoptimization of the orbitals in the presence of the Jastrow factor only affects the DMC excitation energy by at most 0.1-0.2 eV.

In Table 4.5, we summarize the VMC, DMC and TDDFT vertical excitations for the neutral A and the anionic I and B forms of wild-type GFP and compare them with experiments as well as CASPT2 results available in the literature [13]. We observe that the DMC excitation energies are always lower than the VMC values as in the case of the model chromophores in vacuum, but remain too high as compared to experiments. For all three forms, DMC overestimates the excitation energies by as much as 0.5-0.6 eV while the available CASPT2 results appear to be in significantly better agreement with experimental values.

For the anionic I and B forms, the DMC excitation is higher than experiments by roughly the same amount as in vacuum where, for consistency, we are considering the gas-phase excitation computed with the same type of wave function, that is, a SA-CASSCF(2,2) determinantal component with no reoptimized orbitals. Based on our experience in the gas phase, we may

expect that reoptimizing the determinantal component in the presence of the Jastrow factor will affect the excitation energy by a small amount of the order of 0.1-0.2 eV and therefore not sufficient to bring the DMC excitations in agreement with experiments. Again similarly to case of the gas-phase anion, we find that the DMC excitation is reasonably close to the linear-response TDDFT excitation. For the neutral form, the DMC excitation energy is also overestimated as compared to experiments by roughly the same amount as in the anionic forms and therefore in disagreement with the the TDDFT value which instead agrees perfectly with the experimental absorption maximum.

On the positive side, the shift in the excitation energy following deprotonation of the neutral form is well reproduced by DMC while TDDFT fails in describing how the energy correlates with the charge state of the chromophore. The DMC difference between the neutral and the anionic I form is roughly (0.50 ± 0.16) eV, as compared to the experimental value of 0.50 eV. The DMC difference between the neutral and the anionic B form is instead (0.49 ± 0.16) eV while the experimental shift is 0.42 eV.

Understanding the mixed performance of quantum Monte Carlo

Since this work represents the first application of quantum Monte Carlo methods to the computation of the excitations of a large biosystem via a mixed quantum/classical approach, we are still in the process of understanding the reasons of its mixed performance. It is certainly positive and extremely important that DMC can well reproduce the difference between the excitation energies of the various forms of GFP, especially since TDDFT fails to see significant differences between the different charge states of the chromophore. However, an absolute error of 0.5-0.6 eV is certainly too large and at variance with the good performance we have previously observed with the present QMC approach when describing the excitation energies of small prototypical gas-phase molecules [61–64].

There are essentially three possible errors in a DMC calculation, the time-step error, the localization error in treating the non-local pseudopotential and, finally, the fixed-node error. None of this errors is actually small and we always rely in their cancellation when computing the difference of total energies. As the shift between the neutral and anionic forms are well reproduced, the source of error must be affecting all three forms in the same manner by raising the absolute value of all excitations. The impact of the time-step and localization errors have been already analyzed in detail for the small model chromophore in vacuum. The system is now slightly larger and we are using the same time step of 0.055 a.u. as for the smaller model chromophores. However, the algorithms we are using for the DMC calculation

is characterized by small time-step errors and in going to the larger system, we have added the “same” type of atoms (C, N, O and H) and consequently the “same” type of electrons. Therefore, if the time-step extrapolated excitation energy is indistinguishable within statistical error from the excitation at $\tau = 0.055$ a.u. for the smaller chromophore, we expect the same to be true for the protein chromophores. The same analysis holds for the impact of the localization error on the excitations. For the smaller chromophore, the use of a three-body Jastrow factor does not significantly affect the DMC excitation energies within statistical error, so the effect of localizing the non-local potential appears small. Moreover, the DMC excitation energy of the small chromophore is also unchanged when a different algorithm beyond the localization approach is employed. Therefore, as the protein chromophore is rather similar to the smaller chromophore, we would also expect that time-step error and localization error are under control.

The most serious source of error is the fixed-node error which can only be controlled by using better many-body wave functions. In the small anionic chromophore, all attempts to use more sophisticated wave functions have only marginally lowered the DMC excitation energy by 0.1-0.2 eV with respect to the result obtained using the simplest unoptimized SA-CASSCF(2,2) wave function. For the protein system, we find that using the same simple ansatz for the many-body wave function yields excitation energies which are all overestimated with respect to experiments by 0.5-0.6 eV. If the source of error were the fixed-node error, the fact that the error is the same for both the anionic and the neutral forms would likely mean that the multi-configurational nature of the excitation is the same for both charge states and that we are missing some important static correlations in our wave function. This is however at odds with the observation that linear-response adiabatic TDDFT appears to be able to describe the neutral but not the anionic form. In the previous Section, when trying to understand the mixed performance of TDDFT, we attributed the problems of TDDFT with the anionic form to the possible double- and higher-excitation character of this excitation as charge-transfer or underestimation of the ionization threshold do not seem to play a role. Now, from the DMC results with the SA-CASSCF(2,2) wave function, we can infer that, whatever the multi-configurational nature of the excitation, it should be rather similar for both charge states, a fact which leaves unexplained the mixed success of TDDFT. Therefore, what is the source of error in the QMC calculations? Certainly, the most likely candidate remains the fixed-node error and, also in the protein, additional extensive (and computationally costly) studies must be performed with larger active spaces in the determinantal component of the wave functions to really ensure that the CASSCF(2,2) ansatz is not too simplistic.

Another obvious source for our problems is the QM/MM model itself, not being representative of the real protein structure of wild-type GFP and/or of the chromophore-protein interaction. We have already seen how small variations in such a complex model can yield very different excitation energies as in the study by Marques *et al.* who find perfect agreement of adiabatic TDDFT with experiments by protonating (we believe) incorrectly a closeby residue to the GFP chromophore. To understand whether our protein structure is realistic, one could study different mutants as the availability of more theoretical data would allow us to better understand correlations between models and performance of TDDFT and QMC. This rather challenging and long route has not been undertaken in this thesis.

It is instead conceptually simpler even though computationally rather costly to check whether the protein-chromophore interaction is poorly described in our QM/MM model. In the calculations of the electronic states of the QM chromophore in the MM environment, the interaction between the QM and the MM part is purely electrostatic and the environment is not polarizable in response to the excitation of the quantum chromophore. Moreover, as already pointed out above, the complete quantum nature of hydrogen bonds cannot be described in terms of electrostatic interactions only. To understand the limitation of our QM/MM description of the protein-chromophore interaction, we can simply enlarge the QM part to include close residues. One first step in this direction is described in the next Section.

4.5 Enlargement of the QM part

We focus here on the effect of enlarging the QM part of our QM/MM calculation of the I form of wild-type GFP. In Fig. 4.14, we show how we enlarge the QM part of our system by including the residues Arg-96 and Gln-94 which are hydrogen bonded to the oxygen of the imidazole ring. We choose to include these two residues as their presence stabilizes the excited state more than the ground state since, in the excited state, charge is displaced from the phenolic to the imidazole ring. We therefore expect that, if their approximate MM treatment is responsible for the poor agreement of the theoretical excitations with experiments, a correct inclusion in the QM part of these two residues should lower the excitation in the direction of the experimental numbers. Also we note that residue Arg-96 is positively charged and has therefore an important role in stabilizing the negative charge of the anionic chromophore.

The structure of the enlarged QM/MM system is again equilibrated within DFT/PBE QM/MM with the CPMD code. Starting from the structure of the I form obtained in Section 4.1.2, we perform the following steps:

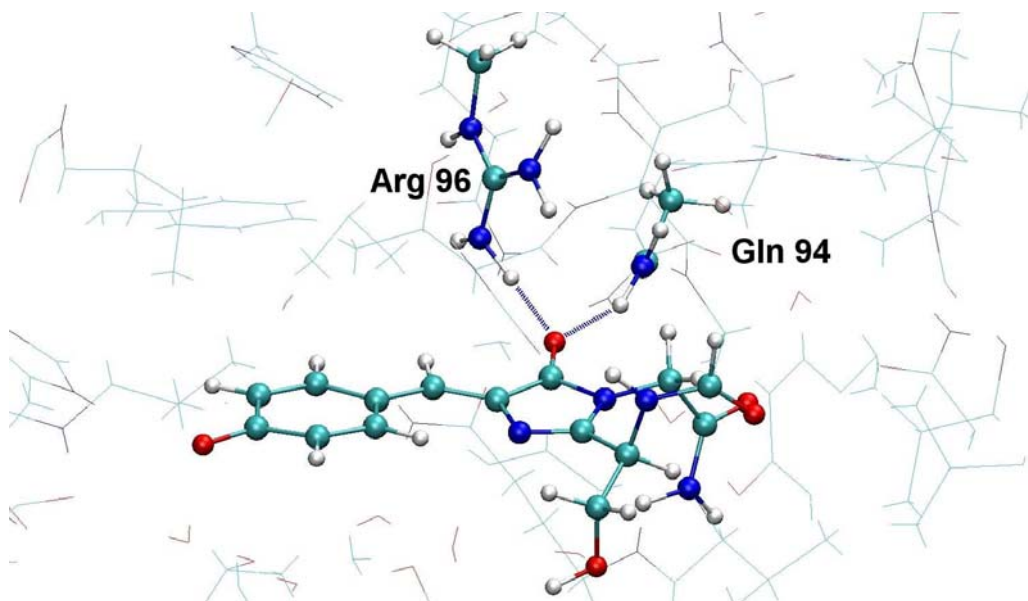


Figure 4.14: Enlarged QM structure of the I form of wild-type GFP. The original chromophore as well as residues Arg-96 and Gln-94 are emphasized and are now included in the QM calculation.

- 1) 0.7 ps of molecular dynamics at constant energy (NVE). The temperature monotonically grows and the QM energy becomes lower.
- 2) 1 ps of molecular dynamics with $f = 0.99$ of the MM system only with the QM atoms at fixed positions.
- 3) 0.07 ps of molecular dynamics with $f = 0.999$.
- 4) 0.14 ps of molecular dynamics with $f = 0.995$.

At the end of the simulation, the bond lengths along the chromophore are oscillating less than $0.0005 \text{ \AA}$, and the hydrogen-bond distances between the imidazole oxygen and hydrogens of Arg-96 and Gln-94 vary less than $0.005 \text{ \AA}$. In figure 4.14, we show the behavior of the hydrogen-bond distances between the imidazole oxygen and the hydrogen donated by Arg-96(N) and Gln-94(N) during the first 0.05 ps of the equilibration process. The starting structure is the optimal structure obtained with the two residues treated as classical, where the two bonds lengths determined via an effective electrostatic interaction between the QM oxygen and the MM hydrogens are rather similar. Interestingly, as the two residues are included in the QM part and the hydrogen bonds treated fully quantum mechanically, the two bonds immediately start to differ. The bond of the oxygen with the hydrogen from the

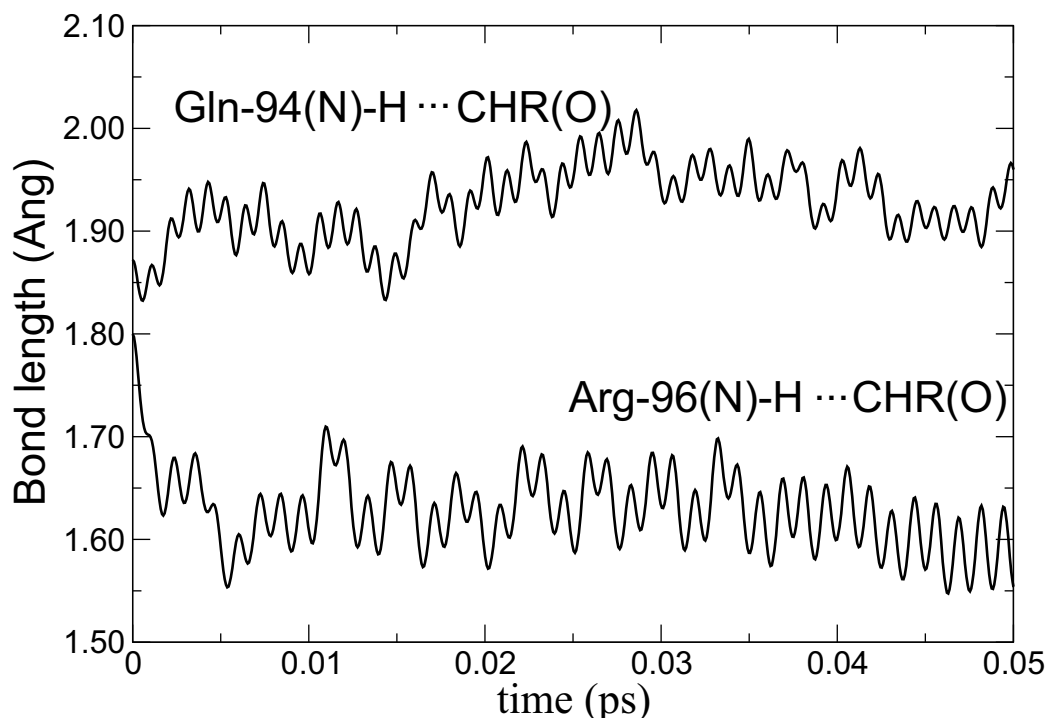


Figure 4.15: Bond lengths in Å between the imidazole oxygen and the hydrogen of the hydrogen-donor nitrogens of Arg-96 and Gln-94 which are included in the QM system. The first 0.05 ps of molecular dynamics are shown. The starting configuration is the optimal QM/MM structure obtained with the two residues treated as classical.

positive Arg-96 is stronger and the corresponding distance becomes shorter while the distance to Gln-94 lengthens. At the end of the dynamics run (not shown in the figure), the oxygen distance from the hydrogen of Arg-96 and Gln-94 settles at 1.66 and 1.88 Å, respectively. The Arg(N)···O₁₂ and Gln(N)···O₁₂ distances become 2.71 and 2.86 Å, respectively, that is, only different by -0.06 and 0.01 Å than the optimal values obtained when the two residues are treated classically. The distances along the chromophore are less affected by the inclusion of the two residues in the QM part of the system. For instance, the central bonds of the carbon bridge become 1.41 and 1.40 Å, as compared to the starting distances of 1.42 and 1.39 Å.

In Table 4.6, we show the TDDFT/MM excitation energies using the BLYP and the B3LYP functionals which should be compared to the corresponding TDDFT energies of Table 4.3. The reduction in the excitation energy is very small and equal to 0.07 and 0.03 eV for the BLYP and B3LYP

Table 4.6: TDDFT/MM excitation energies in eV for the I form of wild-type GFP where the QM part is enlarged to include Arg-96 and Gln-94. The results obtained with the BLYP and B3LYP functionals are shown.

	BLYP	B3LYP
$S_0 \rightarrow S_1$	2.80(0.57)	3.02(0.97)
	H $\rightarrow$ L(0.46)	H $\rightarrow$ L(0.58)
	H-2 $\rightarrow$ L(0.39)	
$S_0 \rightarrow S_2$	3.00(0.26)	3.43(0.0023)
	H-2 $\rightarrow$ L(0.53)	H-1 $\rightarrow$ L(0.69)
	H $\rightarrow$ L(0.28)	
	H-3 $\rightarrow$ L(0.17)	
	H-4 $\rightarrow$ L(0.16)	
$\Delta\epsilon_{\text{HL}}$	1.82	2.91
$-\epsilon_{\text{HOMO}}$	4.14	4.93

functional respectively. For this enlarged QM/MM setup, we also perform quantum Monte Carlo calculations using a SA-CASSCF(2,2) wave function and including the MM environment. We find a DMC excitation energy of 3.2(1) eV which is compatible with the DMC result of 3.16(11) obtained with the smaller QM system. Therefore, even though enlarging the QM subsystem to include the residues Arg-96 and Gln-94 reduces the corresponding hydrogen bond distances to the imidazole oxygen, this geometrical changes are not sufficient to bring a significant shift in the TDDFT and DMC excitation energies.

4.6 Conclusions

We constructed the protein models of the neutral A and the two anionic I and B forms of wild-type GFP using a DFT QM/MM approach. For the neutral form, we started from the available X-ray structure while we obtained the I form by deprotonating and further relaxing the optimal neutral A form. The starting structure of the B form was built by homology from the X-ray structure of the mutant S65T of wild-type GFP.

We first performed a thorough structural analysis of the neutral and anionic forms comparing to available crystallographic data as well as other theoretical studies available in the literature. While all evidence points at

the correctness of our protein models, we find that the DFT QM/MM calculations by Marques *et al.* [12] yield structures significantly different from ours since these authors have incorrectly described the binding site of the chromophore by wrongly protonating a closeby residue. Since the incorrect description of the residues surrounding the chromophore affects its geometry and consequently its response to light, it follows that the perfect agreement of the corresponding TDDFT spectra for the isolated chromophore with the experiments is purely coincidental. This surprising outcome shows how difficult it is to correctly describe a complex biosystem and how easy to be misled in believing the correctness of a given model when comparing to relatively few experimental numbers.

Starting from our optimal structures, we have computed the vertical TDDFT excitations of the chromophore of GFP in the protein environment for the neutral A, and the two anionic I and B forms. We find that adiabatic TDDFT appears to be accurately describing the excitation of the neutral form but significantly overestimates the excitations of the anionic forms by 0.3-0.4 eV depending on the functional. Consequently, the TDDFT shift in the excitation upon deprotonation is not correctly reproduced since TDDFT gives very similar excitations energy for all three forms of GFP. We tried to analyze the reasons of this mixed performance but it is not evident why TDDFT should give an excellent agreement with experiments for the neutral but not the anionic forms, in particular the I form. The excitations of these systems do not appear to be characterized by particular charge transfer and the underestimation of the DFT ionization threshold is not an issue once the chromophore is embedded in the protein environment. A possible explanation is that the excitation of the anionic form has a significant double- or higher-excitation character as compared to the neutral case. However, QMC calculations obtain a comparable description of the excitations of all three protein forms using the same, simple form of wave function, which indicates that, whatever the multi-configurational nature of the excitation, it should be rather similar for both charge states.

Finally, we have explored for the first time the use of QMC in describing the excitations of a chromophore in its protein environment and performed QMC/MM calculations of the excitation energies of the three forms of wild-type GFP using for the moment only a simple wave function. While the DMC excitation energies of both the neutral and the anionic forms are significantly higher than experiments by as much as 0.5-0.6 eV, the experimental shift between the different charge states of the chromophore is well reproduced by DMC. Even though it is reassuring that DMC can describe relative energies correctly, the DMC overestimation of the excitation energies is rather large and we have only begun to investigate the possible reasons for this error

such as shortcomings in the QM/MM description of the chromophore-protein interaction.