



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 1

Introduction

1.1 Photosensing in biological systems

The absorption of visible light and its conversion to other forms of energy is at the core of some of the most fundamental processes in biology. Indeed, life on earth owes its existence to photosynthesis, a process through which sunlight is harvested and converted into chemical energy by plants, algae and photosynthetic bacteria. Another familiar example of light absorption initiating a biological response over several temporal and length scales is vision: light stimulates a conformational change of the photosensitive component in the retina, which is followed by a cascade of chemical reactions ultimately culminating in the stimulation of the optical nerve. In general, photosensing in a biological system occurs through a photoreceptor protein that hosts a chromophore (i.e. the molecule bound to the protein proper and responsible for light absorption and emission) which undergoes a photochemical reaction, such as photoisomerization, excitation transfer, electron or proton transfer upon photoexcitation. Deepening our physical understanding of the primary excitation processes and of the subsequent energy transfer in these photobiological systems is important both from a fundamental point of view and because of existing and potential applications in biology, biotechnology and artificial photosynthetic devices.

An important example of photosensitive biosystems is the family of autofluorescent proteins, a class of biological labels that has revolutionized cellular biology in the last decades. These molecules absorb light at one wavelength and emit (i.e. fluoresce) at a specific and longer wavelength. Since they can often be coexpressed with non-fluorescent proteins without affecting the latter's functions, autofluorescent proteins have been used in a multitude of applications, for example as fluorescent labels to visualize and track

proteins in living cells, to monitor protein-protein interactions, and as indicators of pH and calcium concentration in vivo. For certain applications, it is however desirable to modify, enhance or suppress the molecular mechanisms that modulate the response of the chromophore to external inputs. Understanding the relationship between the microscopic structure and the spectral properties of these biosystems permits the rational design of new photoactive systems with novel functions through selective mutations of existing autofluorescent molecules. Examples include shifting their excitation and emission spectra, or altering the sensitivity to external factors such as pH or past exposure to light.

Theoretical calculations of the optical properties of photoactive systems complement experimental spectroscopic data by providing an atomistic description of the dynamical response of the protein upon light activation. In order to attack these challenging problems, the computational approach must however meet several difficult requirements. First, it should provide an accurate quantum-mechanical description of the ground state and of the electronic excitations of the photoactive site. It should then include a dynamical description of ground state fluctuations and possibly of photo-induced dynamical effects. Finally, the calculations must be able to treat a realistically large model of the biosystem in order to understand how modifications of the protein environment affect the optical properties of the chromophore. It is far from trivial to satisfy all the above requirements. In most cases, ground state properties of large systems can be reliably and efficiently computed from first principles, in particular through density functional theory (DFT) approaches, and sufficient knowledge has also been accumulated to establish the reliability of a given calculation. However, the computation of excitation energies is proving to be more complicated, and there are serious problems with the approaches employed in the study of large photoactive biomolecules. In surveying the vast theoretical literature on photosensitive systems, one finds that the large spread of semi-empirical and first-principle approaches used for a particular system yields an equally large spread of results and predictions.

The most appealing approach for the computation of excitations in large molecular systems is certainly time-dependent density functional theory (TD-DFT) given its favorable scaling with system size. While generally reasonably accurate, conventional adiabatic TDDFT often fails to describe charge transfer excitations in extended conjugated systems and excitations characterized by two- and higher-electron excitations. As we will show in this thesis, these and other shortcomings may result in the poor description of the excitations of photoactive chromophores which usually are conjugate π -systems with electronic states often displaying multi-configurational character. The unre-

liability of density-functional-based approaches to accurately describe photoexcitations of biomolecules implies that the main researchers aggressively working in this field are employing conventional highly-correlated quantum chemical approaches as multi-reference configuration interaction (MRCI) and complete active space second-order perturbation theory (CASPT2). These approaches rely on expanding the wave function in Slater determinants and, as the system size increases and the energies of the single-particle orbitals become closely spaced, the space of orbitals which must be included to recover a significant fraction of electronic correlation grows enormously. Therefore, when these approaches are applied to large biomolecules, compromises must be taken as in the use of a small atomic basis or a reduced space of active orbitals. Consequently, while highly-correlated quantum chemical approaches are accurate for small systems where these techniques can be pushed to their limits, the same level of accuracy cannot in general be guaranteed when going to a large biosystem.

In this thesis, we employ a hierarchy of state-of-the-art computational methods to deal with the problem at different levels of accuracy. We believe that, for the description of the ground state properties of these photoactive biomolecules, conventional techniques are sufficiently accurate while, for excited states, we want to explore the performance of a different approach as new theoretical ways to handle excited states are needed. More specifically, ground state properties can be described using density functional theory in combination with ab-initio molecular dynamics to equilibrate the structures and study the thermal fluctuations of the chromophore and its immediate surroundings. The long-range protein-chromophore interactions can be included via hybrid quantum-classical simulation schemes, where the photoactive site is described quantum mechanically and the interaction with the rest of the macromolecule is treated using an atomic force field.

For the computation of excited states, on the other hand, we will use a different theoretical framework based on many-body quantum Monte Carlo techniques that has been developed over the last few years by Filippi and coworkers and has already yielded accurate excitations of a variety of small photoactive molecules. Moreover, to describe the long-range protein-chromophore interactions, we will combine for the first time quantum Monte Carlo with a molecular mechanics approach where the chromophore is treated quantum mechanically and the rest of the protein classically. The advantage of quantum Monte Carlo methods compared to highly correlated quantum chemical approaches is that they scale far more favorably with system size. While we already know that quantum Monte Carlo is competitive with highly-correlated quantum chemical approaches for small molecules, this study represents the first application of quantum Monte Carlo techniques to

the description of the excitations of a realistic complex biomolecular system.

Using this hierarchical combination of computational approaches, we study here the rich photophysical behavior of Green Fluorescent Protein (GFP), the prototype of the class of autofluorescent proteins and one of the most widely used fluorescent labels in cellular biology. In particular, we investigate the interplay between the spectral properties and the microscopic structural features of the chromophore-protein complex in its different forms. Beyond being extremely relevant in biotechnology, GFP represents a perfect playground for our theoretical investigation of photoactive biomolecules due to several reasons. First, this protein is experimentally very well characterized, serving as a stringent test for any approach aiming at describing excitation processes in biosystems. Then, GFP has already been the subject of a large number of theoretical semi-empirical and first-principle studies, none of which fully conclusive. Finally, despite the substantial body of literature, several issues which we will not touch in this thesis are still open and not convincingly addressed by theoretical calculations. These include the conformational changes in the chromophore and their relation to the so-called dark states which are reversibly accessible after photoexcitation during blinking and switching. For all these reasons, GFP is the ideal arena where to validate and possibly sharpen our proposed methodology while addressing the theoretical challenge to understand the nature of the excitations in this relevant autofluorescent protein.

1.2 The Green Fluorescent Protein

Green fluorescent protein (GFP) is the prototype of the class of autofluorescent proteins [1–3]. GFP is an intrinsically fluorescent protein and was first extracted [4] from the bioluminescent jellyfish *Aequorea victoria* of the Pacific Ocean, shown in Fig. 1.1. The *Aequorea* jellyfish bioluminesces (i.e. emits light as a result of a chemical reaction) at the rim of its bell and two proteins are involved in its bioluminescence, aequorin and the Green Fluorescent Protein. By a quick release of calcium ions, the jellyfish can induce the photoprotein aequorin to emit blue light, which is then transduced to green via radiationless energy transfer to a coupled Green Fluorescent Protein. The biological function of GFP in the jellyfish *Aequorea* is therefore to convert the blue emission of aequorin to green emission. Interestingly, it still remains unclear how and why these organisms use their bioluminescent capabilities as jellyfish do not flash at each other in the dark, nor glow continuously [5]. Moreover, it is not understood why these jellyfishes would synthesize a separate protein rather than mutate the chemiluminescent protein to shift its

wavelength, and why green emission should be ecologically superior to blue.



Figure 1.1: Two views of the hydromedusa *Aequorea victoria* from Friday Harbor, Washington [5].

Independently of the reason, evolution and natural selection has generated a very efficient optical device and this optimization through evolution is probably a reason for the success of GFP in biotechnology. Over the last decades, GFP has in fact become one of the most widely used markers in cellular biology. The most successful and numerous applications of GFP are as a genetic fusion partner to host proteins, which maintain their normal functions but are now fluorescent and can be dynamically visualized in living cells and organisms. This property is dramatically illustrated in Fig. 1.2, where the genetic code of a mouse has been modified to express Green Fluorescent Protein. Moreover, significant experimental efforts have gone in engineering mutants of the original *Aequorea victoria* GFP with different colors, enhanced fluorescence and photostability or specific sensitivity to external factors such as temperature or pH. These mutagenesis studies have resulted in new fluorescent probes that range in color from blue to yellow. The search for mutants with longer-wavelength emission has been motivated by the difficulty to distinguish GFP emission from the background cellular fluorescence, as well as the desire to develop fluorescent resonance energy transfer (FRET) partners with the required overlap between absorption and emission spectra, to tag different proteins and study protein-protein interactions in vivo.

Because the construction of red-shifted mutants from the *Aequorea victoria* jellyfish GFP beyond the yellow spectral region has proven largely unsuccessful, longer-wavelength fluorescent proteins emitting in the orange and red

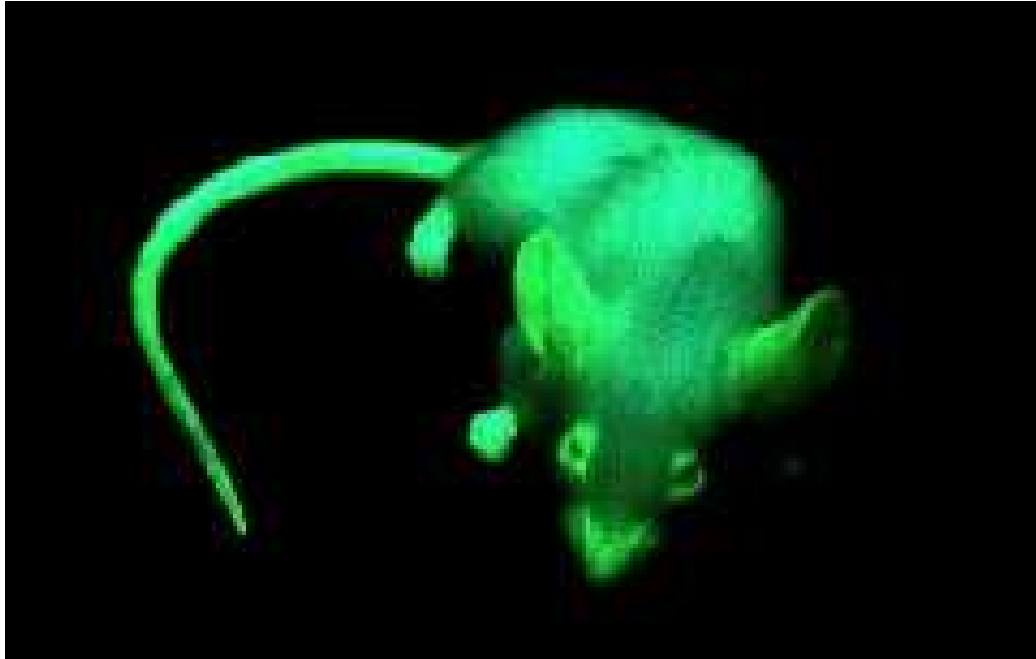


Figure 1.2: Mouse expressing Green Fluorescent Protein, illuminated under blue light [6].

spectral regions, have been extracted from other sea organisms as the marine anemone, *Discosoma striata*, and reef corals belonging to the class *Anthozoa*. Still other species have been mined to produce similar proteins having cyan, green, yellow, orange, and deep red fluorescence emission. Consequently, a broad range of fluorescent protein genetic variants is now available that feature fluorescence emission spanning almost the entire visible light spectrum. In the following, we will restrict our discussion to wild-type GFP, that is the original protein of *Aequorea victoria*.

The tertiary structure of wild-type Green Fluorescent Protein is shown in Figure 1.3. The fold comprises 11 β -sheets arranged in a barrel-like structure with a diameter of about 24 Å and a height of 42 Å. This structure forms the so-called β -can which is capped by short helical segments. The chromophore is well protected in the center of the barrel and is linked to the α -helical stretch which runs close to the central part of the barrel. This fold motif with minor variations is common to all proteins of the GFP family, including the fluorescent proteins extracted from other sea organisms. The correct folding of GFP in the β -can structure and the configuration of the residues around the chromophore are crucial to the formation and the fluorescence of the chromophore which is rigidly kept inside this chemically protective structure,

displaying high stability and quantum yield of fluorescence. In fact, the isolated chromophore is not fluorescent in aqueous solution, and denaturation yields a loss of fluorescence which is regained when the β -can structure is correctly reformed. The isolated chromophore is also shown in Fig. 1.3 and is a *p*-hydroxybenzylideneimidazolinone molecule formed autocatalytically by an intramolecular post-translational cyclization of three consecutive amino acids (Ser-65, Tyr-66, and Gly-67).

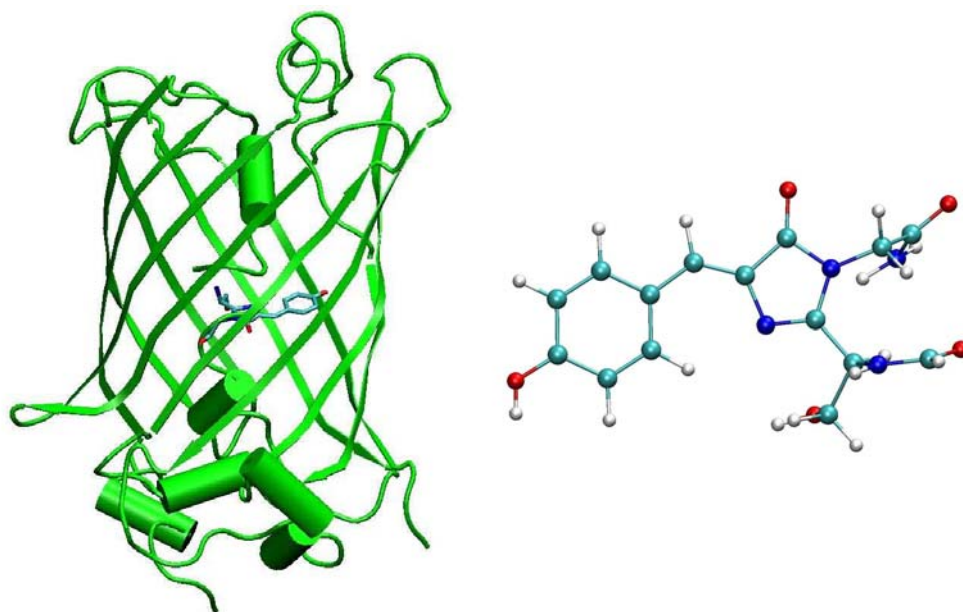


Figure 1.3: Tertiary structure of Green Fluorescent Protein represented in strand style (left). The β -can structure has a diameter of about 24 Å and a height of 42 Å. The chromophore is highlighted in the center of the protein cavity and is also shown isolated in vacuum (right).

The fluorescent mechanisms of wild-type GFP is prototypical of the GFP family. At thermal equilibrium, the absorption spectrum of wild-type GFP has two peaks at 398 nm (3.12 eV) and 478 nm (2.59 eV). Excitation at 398 nm results in an emission maximum in the region of 506 nm (2.45 eV) while irradiation at 478 nm yields emission with a maximum at 482 nm (2.57 eV) [7]. The absorption spectrum of wild-type GFP at room temperature is shown in Fig. 1.4. The two absorption bands at 398 and 478 nm were attributed early on to two interconvertible states of the protein with the chromophore in a neutral (protonated) A form and an anionic (deprotonated) B form, respectively. Upon photoexcitation of the neutral A form, the excited chromophore

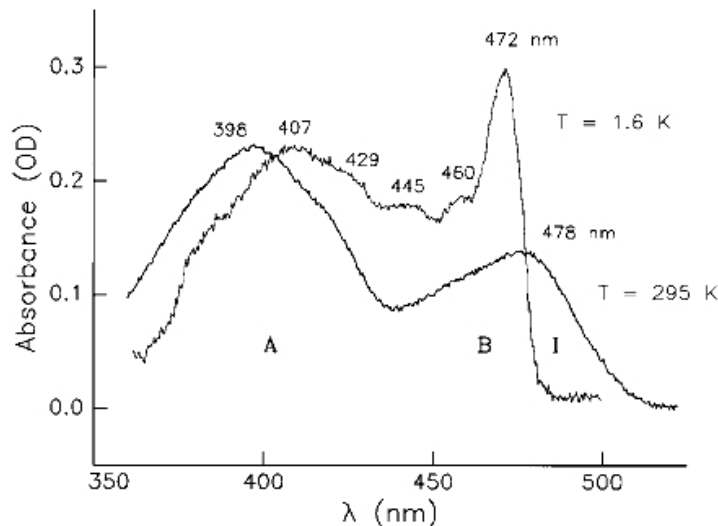


Figure 1.4: Absorption spectra of wild-type GFP at room temperature ($T=295$ K) and low temperature ($T=1.6$ K) from Ref. [7].

transfers a proton through a complex hydrogen-bond network to the residue Glu-222 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 excitation of the B state stems from direct decay of the excited B^* state. Therefore, both the I and the B states are characterized by an anionic (deprotonated) chromophore but the I form has a protein environment similar to the neutral A form while the environment of the B form is structurally different from the A and I forms with the Thr-203 residue being rotated and forming a hydrogen bond with the phenolic oxygen. The fluorescence mechanisms of wild-type GFP is summarized in Fig. 1.5 where a schematic representation of the neutral and anionic chromophores and the corresponding protein binding sites is also shown. This model for the photocycle of wild-type GFP was originally proposed after ultrafast excited-state dynamics measurements and rationalized on the basis of the resolved x-ray structures of the neutral A form and of the B form as stabilized in GFP mutants. We will return to a detailed analysis of the three forms of wild-type GFP and their protein environments in Chapter 4.

Finally, we report here few additional experimental observations which are relevant for our theoretical calculations. In particular, the absorption

1. Introduction

1.2. The Green Fluorescent Protein

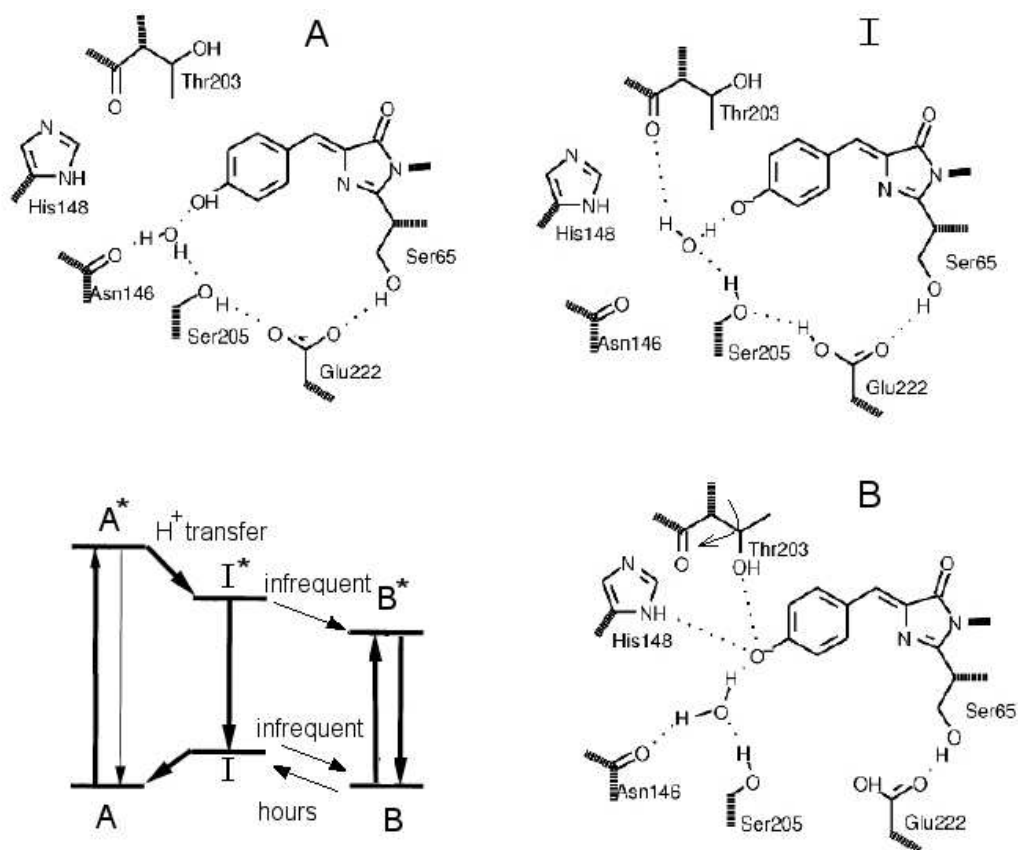


Figure 1.5: Scheme of the fluorescence mechanisms of wild-type Green Fluorescent Protein. The hydrogen bond network from the chromophore through the residues involved in the proton transfer is shown for the neutral A and the anionic I and B forms. Note the change in conformation of residue Thr-203 in going from the I to the B form. The figure is adapted from Ref. [3].

spectrum of wild-type GFP at 1.6 K is also shown in Fig. 1.4. At low temperature, the two maxima shift at 407 nm (3.05 eV) and 472 nm (2.63 eV), and the ratio of the absorbances of the A and B forms inverts with respect to room temperature indicating that the B form has a slightly lower ground state than the A form. 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. Finally, spectral hole-burning experiments have located the 0-0 transitions of the three forms and shown that the ground state of the I form is higher than the ground states of the A and B forms, and separated from them by energy barriers of several hundred wavenumbers. Moreover, the excited-state barrier between A* and I* is low while the barrier between

I^* and B^* is about 2000 cm^{-1} (0.25 eV), so the only possible interconversion is between the excited states of the A and I forms [7].

1.3 Previous theoretical work

The structural and optical properties of wild-type Green Fluorescent Protein have already been the subject of several theoretical investigations. We will not review the early semi-empirical and quantum chemical studies [8–10] since they were not able to unambiguously assign the charge states to the experimental absorption bands. Initially, a cationic and a zwitterionic form of the protein were even proposed as the protonated and the deprotonated state of the chromophore. Moreover, some early calculations yielded excitation energies for a particular charge state of the chromophore varying by more than 1 eV when slightly different quantum chemical approaches were employed [11]. We will focus instead on the most recent first-principle calculations of the excitations of wild-type GFP.

Particularly relevant is a recent first-principle study of the neutral and anionic forms of GFP by Marques *et al.* [12] who report a remarkably good agreement of the time-dependent density functional theory (TDDFT) spectra in the local density approximation (LDA) with experiments. These theoretical results are summarized in Fig. 1.6 where the TDDFT/LDA absorption peaks of 3.01 eV and 2.67 eV are compared with the experimental low-temperature maxima of 3.05 and 2.63 eV for the A and the B form, respectively. Few anomalous features characterize however these calculations and raise doubts about the definite and conclusive nature of this study. While the chromophore-protein structures are optimized in the presence of the protein environment using a DFT/LDA quantum mechanics in molecular mechanics (QM/MM) approach, the TDDFT excitation energies are then computed on the isolated chromophores without the surrounding protein environment. Therefore, possible polarization effects of the protein are not included in the calculation of the excitations of the chromophore. Moreover, the authors model the anionic I form and not the B form by deprotonation of the neutral A form, but erroneously state to be simulating the B form.

Highly-correlated quantum chemical calculations have been recently published for the I and B forms by Sinicropi *et al.* [13] using complete active second-order perturbation theory (CASPT2) for a large model chromophore of GFP in the presence of a classical protein environment. With respect to the original x-ray structure of the neutral A form, only the coordinates of the chromophore and three water molecules are relaxed within the complete active space self consistent field (CASSCF) approach. For the construction

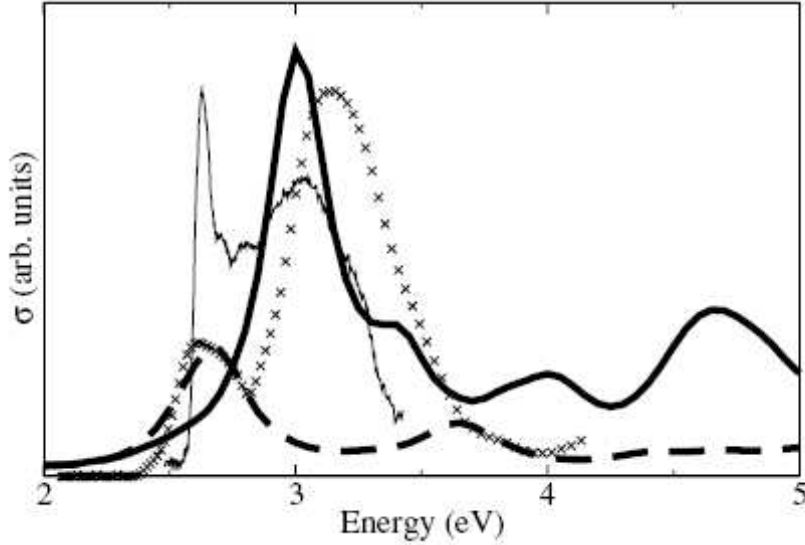


Figure 1.6: TDDFT/LDA spectra of the neutral (thin solid line) and anionic (thick solid line) chromophores of wild-type GFP as computed by Marques *et al.* [12]. The experimental low-temperature (thin line) and room-temperature (crosses) spectra are also shown. The TDDFT calculations are performed for the isolated chromophores whose structures were optimized in ground state DFT/LDA QM/MM calculations. We note that the spectrum for the computed anionic I form is erroneously attributed to the B form. The figure is adapted from Ref. [12].

of the I form, the neutral chromophore is deprotonated and some relevant residues are manually reoriented while, for the B form, residue Thr-203 is partially relaxed in its proper conformation. The CASSCF QM/MM embedding scheme is therefore very simple and lacks a complete relaxation of the chromophore-protein structure. Nevertheless, the CASPT2 absorption maximum of 2.81 eV for the B form appears to be reasonably close to the experimental value of 2.63 eV, while a better agreement with experiments is obtained for the emission maxima of both the I and B forms. Unfortunately, the authors do not report the excitation for the neutral A form of the protein, so it is not possible to access whether this approach is actually capable to correctly describe how the spectrum shifts with the protonation state of the chromophore.

1.4 This thesis

The main focus of this thesis is the computational study of the absorption properties of wild-type Green Fluorescent Protein in the neutral A and anionic I and B forms. We not only construct a series of model chromophores in the gas phase but also investigate how the spectral properties of the chromophore are modified by the protein environment using hybrid molecular mechanics in quantum mechanics approaches to account for the long-range chromophore-protein interactions. To compute the excitations of GFP, we employ both conventional time-dependent density functional theory as well as quantum Monte Carlo techniques. Since this thesis is the first application of mixed classical/quantum Monte Carlo methods to the computation of the excited states of a large biomolecule, it serves the dual purpose of both understanding the spectral tuning of the excitations of GFP by the protein proper as well as assessing the performance of quantum Monte Carlo to describe the excited states of a complex biosystem. This thesis is organized as follows.

In Chapter 2, we describe the computational methods we use in the thesis. We review highly-correlated quantum chemical approaches as well as density functional theory also in its time-dependent formulation. We discuss in depth quantum Monte Carlo methods, in particular the functional form of the trial wave function and the optimization scheme used to obtain the optimal parameters in the excited-state wave functions. We briefly describe molecular mechanics techniques and the hybrid quantum mechanics in molecular mechanics (QM/MM) scheme used for the study of Green Fluorescent Protein. The computational details conclude this Chapter.

In Chapter 3, we construct a set of models of the neutral and anionic chromophores of GFP in the gas phase to begin exploring the performance of adiabatic time-dependent density functional theory and quantum Monte Carlo approaches. The results are puzzling. TDDFT appears to be overestimating the excitations of a small anionic model chromophore as compared to photodissociation spectroscopy experiments and highly-correlated CASPT2 calculations while the experimental absorption maximum obtained with the same technique for a cationic model is reasonably well reproduced. If signatures of possible problems in the use of TDDFT can be found for the larger models that we have constructed, we are not able to rationalize the reasons for its apparent failure in the description of the smaller anionic model chromophore. Moreover, using quantum Monte Carlo techniques and sophisticated wave functions, we obtain excitations for the small anionic model in reasonable agreement with TDDFT. A significant difference with TDDFT is instead that QMC yields a large shift in the excitation when going from the

neutral to the anionic model of the GFP chromophore in the gas phase.

In Chapter 4, we construct the protein models of the neutral A and the two anionic I and B forms of wild-type GFP using a density functional theory QM/MM approach. The outcome of this ground state modeling is already surprising and 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. We carefully analyze the structures of our protein as well as of other models available in the literature and conclude that the DFT QM/MM calculations by Marques *et al.* [12] are incorrect due to what we believe is a wrong description of the binding site of the chromophore. Naturally, the incorrect description of the residues surrounding the chromophore affects its response to light and the perfect agreement of the TDDFT spectra for the corresponding isolated chromophore with experiments shown in Fig. 1.6 is in fact purely coincidental and due to the use of incorrect chromophore structures. Our TDDFT/MM calculations of our chromophore models in the presence of a classical protein environment yield an absorption maximum in agreement with experiments for the neutral A but not for the anionic I and B forms of GFP. The red-shift in excitation due to deprotonation of the chromophore is very badly underestimated by adiabatic TDDFT which sees almost no difference between the neutral and anionic excitations. We then explore for the first time the use of QMC in describing the excitations of a chromophore in its protein environment and perform QMC/MM calculations of the excitation energies of the three forms of wild-type GFP, using for the moment only a simple wave function. We find that the experimental shift between the different charge states of the chromophore-protein complex is well reproduced by QMC but the absolute excitation energies are overestimated as compared to experiments. We show some first steps to investigate the possible reasons for this error such as shortcomings in the QM/MM description of the chromophore-protein interaction, which, we believe, will resolve the issue in combination with the use of more sophisticated wave functions.

Chapter 5 is self-standing and outside the main thread of the thesis, and focuses on the cooperative effects of π - π and π -anion interactions, a relevant theme within supramolecular chemistry for the design of receptors of anionic species. In particular, we investigate the geometrical and energetic effects induced by π - π stacking on the anion- π system of the unusual triazine-triazine-nitrate complex recently observed experimentally, using semi-empirical dispersion corrected density functional theory and QMC methods. We reproduce and rationalize the highly asymmetrical features of the experimental structure, which are not imposed by the coordination of the anion- π - π subunit within the particular synthesized compound. We quantify the energetic

stabilization induced by π - π stacking and discuss ways to further enhance this cooperative effect in the design of anion-host architectures.