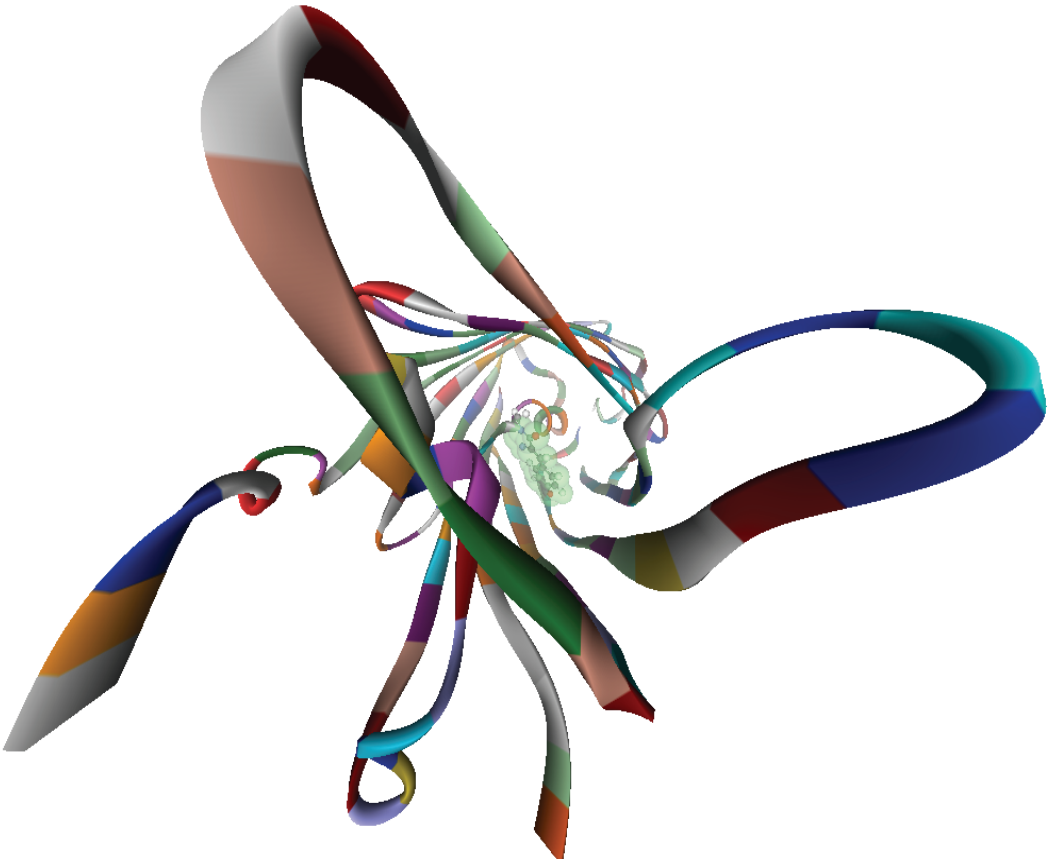
UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

DOTTORATO IN SCIENZE CHIMICHE

XXV CICLO

# Optical properties and photoreactivity of the Green Fluorescent Protein. A theoretical study.

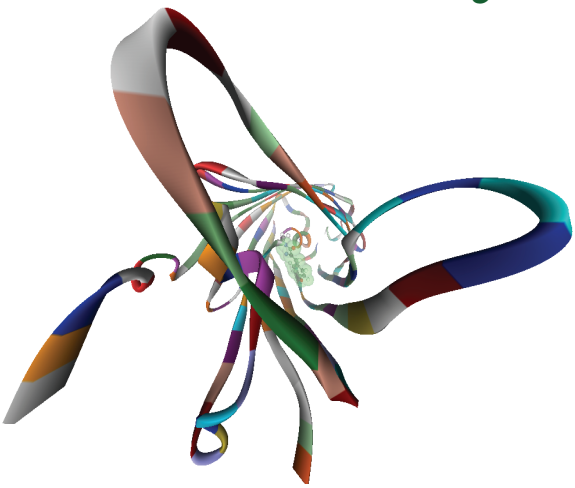


*Tutore Prof. Nadia Rega* *Candidata Silvia Tenuta*

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

DOTTORATO IN SCIENZE CHIMICHE
XXV CICLO

# Optical properties and photoreactivity of the Green Fluorescent Protein. A theoretical study.



*Candidata Silvia Tenuta*

*Tutore Prof. Nadia Rega* *Relatore Prof. Filomena Sica*

*Coordinatore Prof. Lucio Previtera*

*2009-2012*

# L'esperienza non è ciò che accade all'uomo ...

# Contents

**Abbreviations index** **3**

**Preface** **7**

**Abstract** **19**

**1 Introduction** **25**

**2 Methods and Models** **55**

2.1 Methods . . . . . 56

2.1.1 Time independent and dependent density functional theory . . . . . 56

2.1.2 Quantum Mechanics/Molecular Mechanics hybrid methods . . . . . 67

2.1.3 Linear Response and State Specific Polarizable Continuum Model . . . 75

2.2 Models . . . . . 85

2.2.1 Wild type neutral GFP . . . . . 85

2.2.2 Anionic GFP . . . . . 96
2.2.3 Proton Transfer . . . . . 106
2.2.4 Electronic level of theory . . . . . 120

**3 Results and discussion 125**
3.1 Neutral GFP . . . . . 125
3.1.1 The isolated chromophore . . . . . 126
3.1.2 The environment contribution . . . 128
3.1.3 Validation tests . . . . . 139
3.2 Anionic GFP . . . . . 143
3.2.1 TD-DFT accuracy check . . . . . 143
3.2.2 Protein effect . . . . . 145
3.2.3 HBDI anion in solution . . . . . 147
3.2.4 HBDI anion/solvent clusters in solution . . . . . 150
3.2.5 Linear Response versus State Specific solvation . . . . . 157
3.3 Proton transfer mechanism analysis . . . . 160
3.3.1 Approximated simple models analysis 163
3.3.2 GFP structure analysis . . . . . 170
3.3.3 IRC analysis . . . . . 177

**4 Conclusions 197**

**Bibliography 205**

# Abbreviations index

**A GFP**, the neutral form of GFP

**B GFP**, the anionic stable form of GFP

**CC**, Coupled Cluster

**CI**, Configuration Interaction

**CPCM**, Conductor-like Polarizable Continuum Model

**DFT**, Density Functional Theory

**ESPT**, Excited State Proton Transfer

**FP**, Fluorescent Protein

**GFP**, Green Fluorescent Protein

**HBI**, 4-hydroxybenzylidene-imidazolidin-5-one

**HBDI**, 4-hydroxybenzylidene-2,3-dimethyl-imidazolinone - GFP neutral chromophore

**HBDI$^-$**, 4-hydroxybenzylidene-2,3-dimethyl-imidazolinone in the anionic form - GFP anionic chromophore

**HF**, Hartree-Fock

**I GFP**, anionic intermediate GFP form, obtained as result of the proton transfer reaction

**IRC**, Intrinsic Reaction Path

**LR-PCM**, Linear Response Polarizable Continuum Model

**ONIOM**, N-layered Integrated molecular Orbital and Molecular mechanics

**PCM**, Polarizable Continuum Model

**PT**, Proton Transfer

**QM**, Quantum Mechanics

**QM/MM**, Quantum Mechanics /Molecular Mechanics

$\mathbf{S}_0$, singlet ground state

$\mathbf{S}_1$, singlet excited state

**SS-PCM**, State Specific Polarizable Continuum Model

**TD-DFT**, Time-Dependent Density Functional Theory

**TS**, Transition State

**VEE**, Vertical Excitation Energy

**wt-GFP**, wild type Green Fluorescent Protein

# Preface

*Originality does not consist*
*in saying what no one else has ever said before,*
*but in saying exactly what you think yourself.*
*(J. Stephens)*

**Why optical properties and photoreactivity?**

Analyzing the optical properties and the photoreactivity of a system means to describe its photophysics and photochemistry. The *photophysics* of a system is the ensemble of the processes related to the absorption of one or more photons and involves the physical phenomena induced by the interaction of the electromagnetic radiation and the substance. In particular, as it is possible to schematize through Jablonski diagram, the photophysics of a system includes photoexcitation, emission processes (luminescence), non radiative relaxation phenomena, photoinduced energy transfer and charge transfer (see Figure 1).

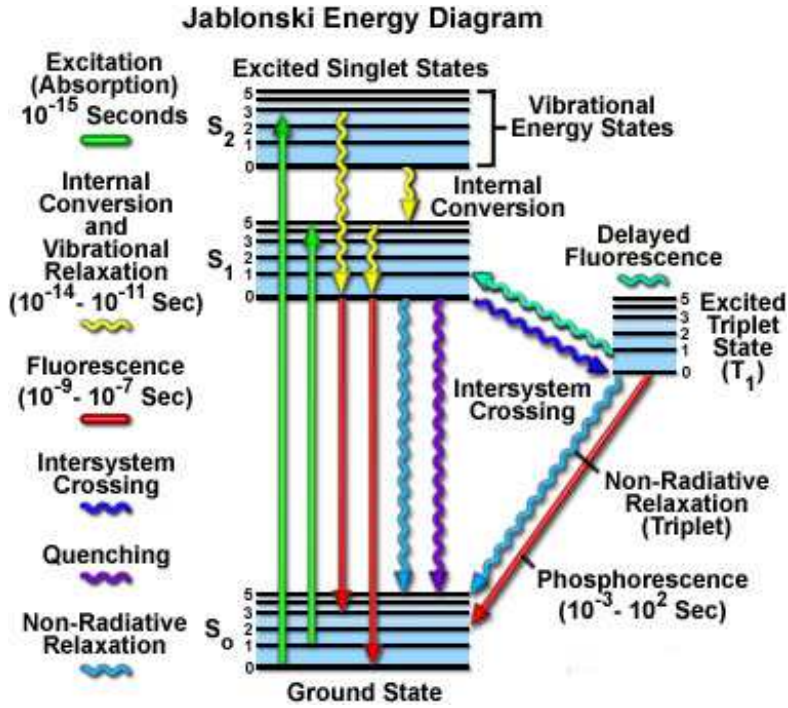



Figure 1: A Jablonski diagram illustrates the electronic states of a molecule and the transitions between them. The states are arranged vertically by energy and grouped horizontally by spin multiplicity. Radiative transitions are indicated by straight arrows and nonradiative transitions by squiggly arrows. The vibrational ground states of each electronic state are indicated with thick lines, the higher vibrational states with thinner lines.

Photophysical processes do not produce chemical modifications of the system studied. The *photochemistry* of a system involves all the chemical reactions occurring after the excitation process in the excited states. As it is possible to understand, photoinduced reactions involve the excited states of a system, whose properties could be really different from the ones of the ground state. In order to have knowledge of these properties and predict the photochemical behaviour of a molecule it is necessary to accurately understand the physical bases of the absorption and emission processes. The study of photophysics and photochemistry has recently supported innovation in many fields of the scientific research, including the photonics, the molecular electronics, the sensors development and the design of new devices and molecular machines.

In this work we are going to focus on optical processes occurring in biological systems. The interaction between radiation and biological molecules can control, damage or promote mechanisms which are the basis of the organisms functionality or existence itself. The comprehension of the paths employed by the system of interest to use the energy provided by external sources allows a better exploitation of the system properties and the design of similar, but more efficient systems.

**Why fluorescent proteins?**

In recent years it has been developed an increasing interest on biological molecules interacting with radiation because they are suited for many advanced applications: for example, photosensitive molecules are often used in association with complex and advanced experimental techniques, such as some innovative spectroscopies which have the capability of following pico or femtoseconds phenomena.

We have chosen Fluorescent Proteins (FPs), in particular, Green Fluorescent Proteins (GFPs) as prototype in order to engine protocols applicable to the analysis of every protein photochemistry, breaking it up in its different components and focusing in particular on the protein environmental influence on the isolated chromophore properties.

FPs are members of a structurally homologous class of proteins that share the unique property of being self-sufficient to form a visible wavelength chromophore from a sequence of amino acids within their own polypeptide sequence. FPs are quite versatile and have been successfully employed in almost every biological discipline from microbiology to systems physiology. These ubiquitous probes have been extremely useful as reporters for gene expression studies in cultured cells and tissues, as well as living

animals. In live cells, FPs are most commonly employed to track the localization and dynamics of proteins, organelles and other cellular compartments. *Aequorea victoria* GFP was the first whose gene was cloned in 1992. Since these early studies, GFP has been engineered to produce a vast number of variously colored mutants, fusion proteins, and biosensors that are broadly referred to as FPs. More recently, FPs from other species have been identified and isolated, resulting in further expansion of the color palette. Development of existing FPs, together with new technologies, such as insertion of unnatural amino acids, will further expand the color palette. Finally, the tremendous potential in FPs applications for the engineering of biosensors is just now being realized. The number of biosensor constructs is rapidly growing. By using structural information, development of these probes has led to improved sensitivity and will continue to do so. The success of these endeavors certainly suggests that almost any biological parameter will be measurable using the appropriate FP-based biosensor.

In this preface I want to mention at least one of the latest technical improvement using FPs, dated 28 november 2012. A newly-engineered strain of mice whose dividing cells express a FP has been developed: it could open the door to new methods of regulating cell proliferation in hu-

mans (see Figure 2). Cell proliferation plays a key role in degenerative diseases, in which specific cells do not replicate enough and in cancers, in which cells replicate too much. Cells in the human body grow and multiply during body growth or during tissue regeneration after damage. However most mature tissues require only rare cell divisions. Scientists who wish to study these rare populations of replicating cells face a serious obstacle: most current methods for labeling and identifying replicating cells involve procedures that kill the cells and destroy sensitive biological material. This limits the ability of scientists to examine important functions of these cells, for example the genes active in such cells. This obstacle could be overcome using FPs.

This is only an example of a fluorescent-cell technique which may speed diagnosis of some human diseases.

**Why theoretical analysis?**

Among the various techniques suited to study the photophysics and the photochemistry of a biological or a chemical system there are the computational methodologies, which have had an increasing development in recent years in a wide field, going from biotechnology or molecular biology to medicine, pharmacy or materials chemistry. Com-

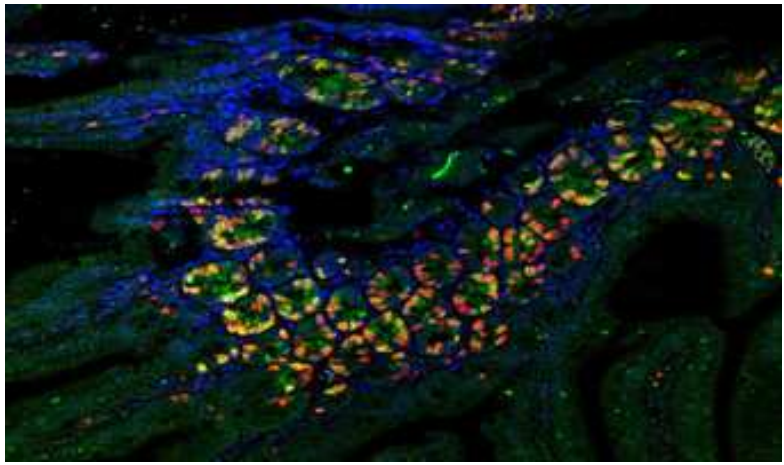


Figure 2: Fluorescence microscopy section of a mice intestine. Green/yellow areas mark sites of cell proliferation. Cell nuclei marked in blue and nuclei with DNA replication marked in red (photo: Agnes Klochendler)

putational chemistry is the branch of chemistry, which uses computer simulation to solve different kinds of problems. It applys theoretical calculation developments in order to simulate structure, properties and reactivity of molecules in several environments. Its quick evolution has made computational chemistry a useful tools of research in many areas, including the spectroscopic characterization or the design of new high-tech molecules. Calculations have the role of supporting and complete the experiments, providing a key predictive and interpretative back-up:

- it is a powerful help in evaluating ambiguous, cryptic or vague experimental data and in rationalizing the aspects which influence a specific behaviour or process.

- It is a predictive tool because it allows both to guide and optimize experimental procedures which require by itself high costs or elevated dangerousness and to study systems which could not be studied experimentally or in another way for technical or economical reasons.

- It provides the possibility to foresee the existence of unknown or not yet synthesized biological or chemical species.

Exemplifying molecular properties possible to be determined through computational methods are structure, molecular energy or interaction energy, electronic properties (i.e. electrostatic or dipole etc.), spectroscopic properties (such as IR, NMR, EPR, UV, Raman etc.) and features related to the reactivity of the system taken into account (kinetic constants, reaction rates, activation thresholds etc.)
Computational modeling is a crucial complement to experiments in deepening our understanding of optical processes occurring in photosensitive biological systems. In fact, theory allows to

- describe the ground and the excited electronic states, their properties and fate, the transitions which can happen among them, absorption, fluorescence or phosphorescence, quenching or relaxation processes arisen with heat dispersion.

- Evaluate the possible reaction occurring in the excited states.

- Analyze the relation between the structure of the photoactive biosystem treated and its spectroscopic properties, giving the possibility to focus on the specific site related to the optical property or to examine the entire protein environment influence.

**Which theoretical approach to describe GFP?**

Looking at the vast theoretical literature produced on GFPs it is usual to find that the large spread of approaches used to describe GFP system brings to an equally large spread of results and predictions. Despite significant theoretical progress in electronic structure methods, it is far from trivial to accurately compute excitation properties of even relatively small photoactive molecules. The difficulties arise from the rather strict requirements the theoretical approach must meet to provide a predictive description of a photoactive biosystem. It should give an accurate and balanced description of the ground and the excited states of the photoactive site and also be able to treat a realistically large model of the biosystem. Post Hartree-Fock (HF) methods have established themselves as important theoretical tools to compute excitations of photosensitive small systems. In fact, they allow an accurate description of limited size systems, while they are not performing when scaling the system extent, because the computational cost they require to treat a large system is very high and usually not compatible with research times. Density Functional Theory (DFT) [1–3] and Time-Dependent Density Functional Theory (TD-DFT)[4–7] are the most appealing approach to efficiently compute the excitations

of large molecular complexes. In fact, DFT permits to treat the electronic problem in an effective way, maintaining a comparable accuracy with respect to the most sophisticated post HF methods and requiring accessible computational costs, if charge transfer excitations [8, 9] or electron transfer processes are not involved into the system. The best choice to describe large-sized systems is the use of hybrid methods (e.g. the quantum mechanics/ molecular mechanics treatment- QM/MM) through which it is possible to define more "layers" within the molecular structure that are treated at more levels of theory. These methods represents a good compromise between accuracy and computational cost. We have used the density functional theory (DFT) and its time-dependent version (TD-DFT) [1–7] to describe the electronic excitation and hybrid quantum mechanical/molecular mechanics methods (QM/MM) [10–13] to analyze the entire GFP and find a correlation between the structural and the optical properties of this protein.

# Abstract

*Everyone has a duty: happyness!*
*(P. Brukner)*

Green Fluorescent Proteins (GFPs) show a lot of peculiar properties, such as the high stability to a wide range of conditions, the resistance to denaturant agents, the possibility to be fused with a broad variety of proteins without altering their functions and the extreme flexibility. These characteristics make them and their mutants ideal noninvasive markers in living cells [14–16], suited for numerous applications, like reporter of gene expression or cell lineage tracer, in many fields, from molecular biology to drug discovery. Among the most promising uses of GFP there is the employment in medical research, for example as marker for tumor cells or neurons identification to obtain a scan of cerebral tissue. In order to maximize the potentiality of these proteins and use them in a more effective way to improve bioimaging techniques and nanotechnol-

ogy it is fundamental to have a deeper comprehension of their photochemistry and of the relation between their structure and properties. Although the huge number of recent works, GFP photochemistry has not often been revealed in its deeper and hidden features [17].

The theoretical-computational approach can be very useful to analyze the photochemistry of a biological system. In fact, the processes involved are typically ultrafast and so they are often not completely understandable through the experiment, but need the support of the calculations.

Our work has been aimed at analyzing the optical behaviour and the photoreactivity of the GFP, using computational tools as time-dependent density functional theory (TD-DFT) [1–7] and hybrid quantum mechanical/molecular mechanics methods (QM/MM) [10–13].

The main species involved into GFP photochemistry are three: the neutral form (A), the anionic intermediate (I) and the anionic stable form (B). In the ground state there is an equilibrium between A and B forms governed by changes in protein concentration, ionic strength, pH and temperature. This equilibrium has been extensively investigated both in the ground [18–20] and in the first singlet excited state[21–23], where A gives origin to B form passing through an unstable intermediate, I. I form is the result of the proton release from the neutral chromophore to

a glutammic residue of the protein and corresponds to the anionic chromophore in a nonequilibrium environment.

One of the major issues in literature is that the chromophore does not show regular trends of absorption bands with polarity or proticity of the solvent, as it is possible to verify looking at the experimental trend [24]. In addition, the trend in solution seems to be of different nature when compared to the protein. To reach a deeper comprehension of the GFP anionic chromophore properties we have analyzed its optical absorption in gas phase and several solvents. We calculated at the same level of theory the vertical excitation energy of the anionic chromophore in several environments, reproducing the experimental results with 0.02 eV as standard deviation of the accuracy. This systematic error allows us to reproduce and to understand, for the first time, the observed trend of the chromophore absorption peak in the protein, ethanol, dioxane, methanol and water. As a consequence, we have been able to analyze with confidence the relative weight of different solvation effects affecting the vertical excitation energy. As indirect and remarkable result, the present analysis suggests that the optical absorption of the chromophore in the gas phase is close to the value of 2.84 eV extrapolated by Dong [24], and confirms the still debated hypothesis that the protein environment induces a red-shift of 0.23

eV.

Another matter of discussion is the role of the protein environment on the chromophore photochemistry, especially for the anionic GFP [24–30]. In fact, it is difficult to accurately define how the bulk influences the intrinsic properties of a fluorescent protein, such as the position of absorption and emission peaks and the possible excited state reactions. A deep comprehension of the photophysics changes induced by the molecular environment is necessary to reach a fine control on GFPs absorption and a highly modulated design of fluorescence.

We have well reproduced the experimental red shifts in both the neutral and anionic GFP (as experimental reference for the optical absorption we have adopted the value extrapolated by Dong et al.[24]). Then we have analyzed in detail the contribution of the protein environment on the absorption band: we have identified three different effects. Nearly a third of the entire protein contribution is a *structural* effect, which is due to the different geometry (i.e. bond distances, angles, dihedrals) imposed by the protein to the chromophore. A second effect is purely *electrostatic*, and is due to the hydrogen bond network surrounding the chromophore, and to the interactions it engages with a water molecule and several residues, mainly the Ser205 and the Glu222. Most of the remaining shift

is due to a modification of a $\pi$ orbital mainly localized on the chromophore. This contribution is *quantum mechanical* in nature.

Another matter of discussion is the proton transfer, occurring in GFPs in the excited state after the excitation of the neutral form, whose mechanism has to be defined yet.

We have conducted a qualitative analysis of the proton transfer for both a QM model and the entire QM/MM protein using an approximated static model. We have hypothesized two extreme mechanisms, a linear synchronous path and a step-wise mechanism, and analyzed the reaction energy profiles for the ground and the excited state of both the QM model and the entire protein. We have obtained the qualitative results that the transfer could occur only into the excited state and that the protein has a fundamental role to let the transfer happen. In fact, without the environment the barrier to overcome for the transfer reaction is too high.

Starting from these preliminary results, we have conducted a deeper analysis of the proton transfer mechanism. First of all to handle a model which could accurately describe the effect of the protein environment on the transfer reaction, we have obtained QM/MM GFP structures for the neutral species and the anionic intermediate in both the

ground and the excited state involved into the absorption. Then we have extrapolated reduced models from these structures and proposed a mechanism for the proton transfer within these models: we have followed the proton transfer reaction by integrating the intrinsic reaction coordinate (IRC) on both the $S_0$ and $S_1$ potential energy surfaces. The whole reaction is more favoured in the excited state. The reaction in $S_0$ presents a synchronous and concerted path, while in $S_1$ appears to be concerted, but the three proton movements are asynchronous. In particular the proton transfer mechanism appears to be driven by the protonation of the Glu222, rather than the proton detachment from the chromophore.

# 1 Introduction

*Change starts when someone sees the next step.*
*(W. Drayton)*

Although the study of Fluorescent Proteins (FPs) has been developed a lot in recent years because of their numerous applications in a wide field, going from biochemistry and biotechnology to medicine and biosensoristics [15, 16], their photochemistry has not often been revealed in its deeper and hidden features.

We chose GFP as a prototype in order to engine protocols applicable to the analysis of every proteins photochemistry, breaking it up in its different components and focusing in particular on the protein environmental influence on the isolated chromophore properties.

**What GFPs are?**

Figure 1.1: *Aequorea victoria* jellyfish

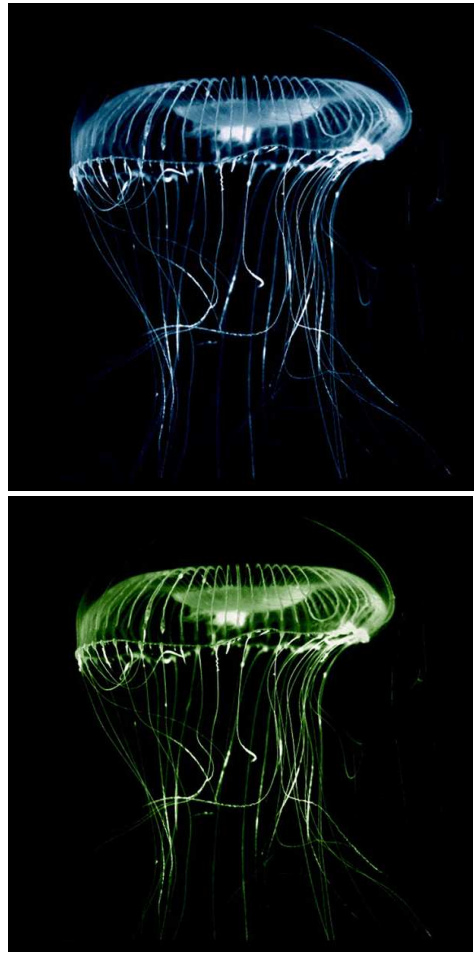
Green Fluorescent Proteins (*GFPs*) are spontaneously fluorescent proteins, found in numerous organisms, living in a variety of environments.

One of the most studied species is the *Aequoreavictoria* (see Figure 1.1) jellyfish [31], that lives in the North-West part of the Pacific Ocean and whose GFP was the first for which the gene was cloned [32] and expressed [33].

Two proteins are involved in Aequorea chemiluminescence process, *Aequorin* and *GFP*. Aequorin is the enzyme that catalyzes the oxidation of the substrate, coelenterazine, through the binding of three calcium ions. A $Ca_3$-apoaequorin-coelenteramide complex results from this interaction. This complex in vitro gives a blue emission [34–36], while in vivo it undergoes radiationless energy transfer to GFP, which fluoresces green [31, 37, 38]. So GFP role in its natural environment is to transmute blue chemiluminescence from a distinct primary protein into green fluorescence (see Figure 1.2).

It is interesting to note that the first report about this type of chemiluminescence was written by Pliny the Elder in the first century AD, after the observation of the bright glow of certain jellyfish present in the Bay of Naples. The study of fluorescence has been continued for centuries, since in 1962 Shimomura, Chalfie and Tsien have found out the GFP, revealing the mysteries of *Aequorea victo-*

Figure 1.2: The natural role of GFP is to transmute Aequorin blue chemiluminescence into green fluorescence.



*ria*'s chemiluminescence and allowing to use this protein and its mutants for different applications. Unconsciouss of the important developments this discovery would have brought, the three researchers started their studies on *Aequorea victoria* attracted by the fact that, when it is irradiated with ultraviolet light, it fluoresces green [39]. Thanks to their scientific contribution they won the nobel prize for chemistry in 2008.

### GFPs properties and applications

GFP has been object of investigation because it shows a lot of peculiar properties. It is highly stable to a wide range of conditons: it is resistant to denaturant agents, such as 1% sodium dodecyl sulphate or 6M guanidinium chloride (denaturation only occurs under very hard conditions, eg. 6M guanidine hydrocloride at 90°C or pH $<$ 4 or pH $>$ 12). It tolerates heat, alkaline pH (in fact, it remains fluorescent up to 65°C or pH 11), detergents and most proteases for many hours [40–43]. It is possible to fuse GFP with a broad variety of proteins without altering their functions [32, 33, 44–46]. In fact, many studies proved that the color changes observed in GFPs emission are determined by processes involving the proteins to which GFP is linked and do not arise in GFP itself.

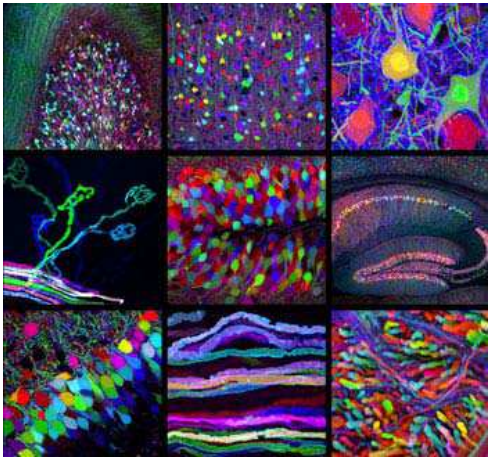


Figure 1.3: Mosaic of a section of cerebral tissue, in which it is possible to distinguish different neurons and their extensions due to the several markers colours.

These characteristics, along with GFP large flexibility, make it and its mutants ideal noninvasive markers in living cells, suited for numerous applications, like reporter of gene expression, cell lineage tracer, measure of protein-protein interactions, signaling and trafficking in cellular systems [44–51].

Moreover, the molecular cloning of its cDNA and the possibility to express it as a functional transgene make GFP a useful tool of investigation in many fields, from molecular biology, cell biology [15] and biotechnology to drug discovery [16]. One of the most promising uses of GFP is the employment in medical research, for example as

marker for tumor cells to follow tumor progression and allow to detect metastases [52–54] or as marker for neurons identification in order to obtain a scan of cerebral tissue [55]: modifying different GFP amino acids, Tsien succeeded in obtaining a brighter green fluorescence and also light blue, cyan and yellow fluorescence. The variety of markers, deriving from GFP, allows to get biological multicoloured pictures full of important information. For example, the mosaic shown in Figure 1.3 presents a section of cerebral tissue, in which it is possible to distinguish different neurons and their extensions due to the several markers colours.

There are also some techniques of analysis which can be enhanced by GFP or other fluorescent proteins. Among them one of the most successful is the fluorescence resonance energy transfer (*FRET*) based on a nonradiative exchange of energy from an excited donor fluorophore to an acceptor fluorophore that is located within 10-100 Å from the donor. GFP-based FRET could be used to determine calcium concentrations, to analyse protein-protein interactions [56–59] or conformational changes [60].

### *Aequorea victoria* GFP structure

*Aequorea victoria* GFP is composed of 11 $\beta$-strands, which

form a $\beta$-barrel with a diameter of about 24 Å and a height of 42 Å. It is crossed by an $\alpha$-helix and displays fragments of $\alpha$-helix on the ends of the cylinder (see Figure 1.4a). This structure is called $\beta - can$ [61]. In the center of the can, located on the $\alpha$-helix, there is the chromophore, well-protected from the external environment, as it is possible to see in Figure 1.4b (this structure allows the chromophore to avoid quenching by oxygen [62], attack by hydronium ions [63] or whichever action made by any solvent).

GFP can be both monomer or dimer, depending on the crystal growth conditions; in fact, at high ionic strength and high protein concentration GFP monomers tend to dimerize [64]. It is worth of note that Aequorin binds the dimer, but not the monomer and that dimers or multimers predominate in the light-emitting organelles of the marine jellyfish.

The cromophore (see Figure 1.4c) is composed of three amino acids, Ser 65, Tyr 66 and Gly 67, which give origin to a 4-(p-hydroxybenzylidene)-imidazolidin-5-one structure through an autocatalytic post-translational internal cyclization. The detailed mechanism for the formation of this cycle is still unknown, even if Tsien et al. [45, 65] proposed the one shown in Figure 1.5, that is in accordance with the available experimental data.

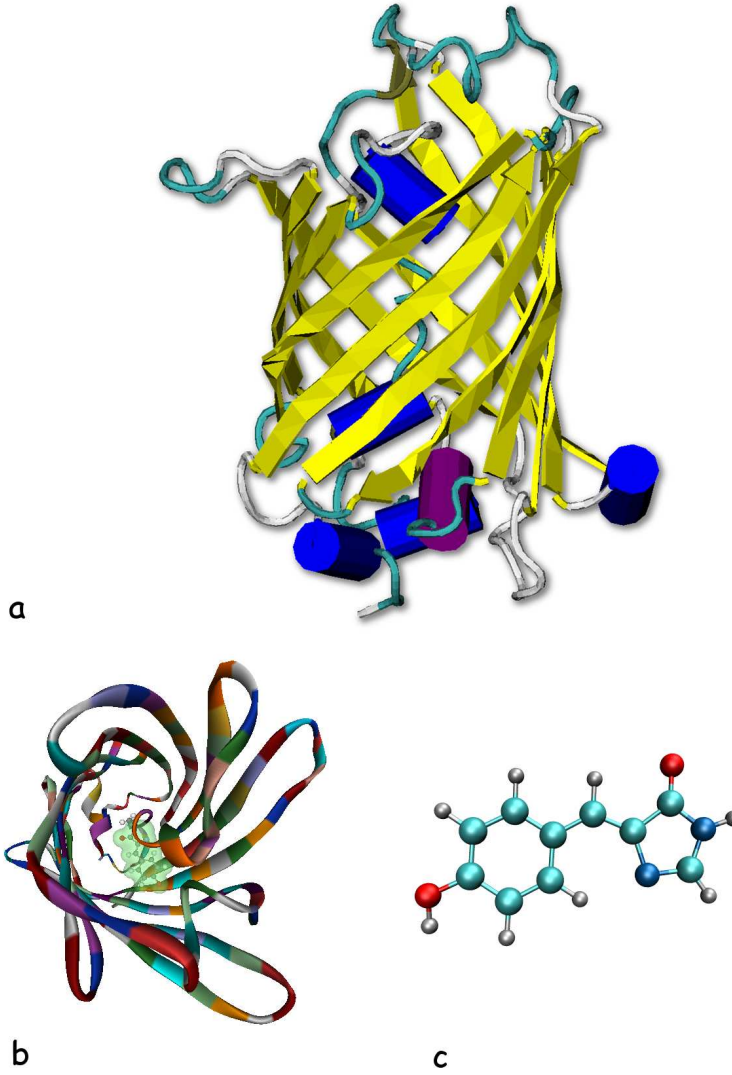



Figure 1.4: a) *Aequorea victoria* GFP $\beta - can$ structure. b) The chromophore (highlighted through a green surface) located in the center of the can, on the $\alpha$-helix. c) Sketch of the chromophore structure.

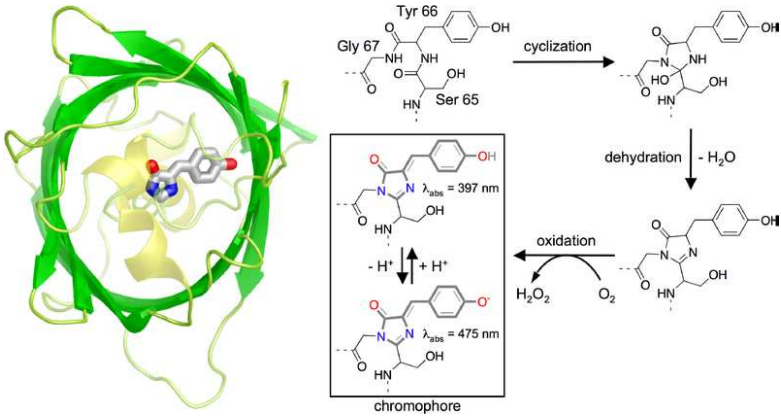



Figure 1.5: Proposed mechanism for GFP chromophore formation: the first step is a rapid cyclization between Ser 65 and Gly 67 to form the intermediate imidazolidin-5-one. Afterwards there is the dehydrogenation between $C_\alpha$ and $C_\beta$ of Tyr 66 made by $O_2$ to form a conjugated chromophore.

The stable and rigid pocket [66], in which the fluorophore is located, contains a large number of charged residues in the immediate environment and is formed by both polar and apolar amino acids side chains. Water molecules and several polar residues form a wide network of hydrogen bonds, which stabilize and jam the chromophore.

Deletion mapping experiments showed that the chromophore needs the entire protein structure to be formed and to fluoresce (denatured GFP is non-fluorescent). The role of the protein environment on GFP photochemistry is a challenging issue of scientific interest. It seems that the potential energy surface of the excited states is modified by the cavity and that the structural relaxation of both neutral and anionic species is influenced by the protein backbone [67].

### GFP photophysics and photochemistry

Photophysics and photochemistry study is one of the most interesting aspects within GFP's investigation [14] and recently it has attracted much attention. Both experimental [68–72] and theoretical [67, 73–78] works have been conducted in order to shed light on this complex matter. Two thermodynamically stable states (shown in Figure 1.6) of the chromophore are photoactive in wild-type GFP, one

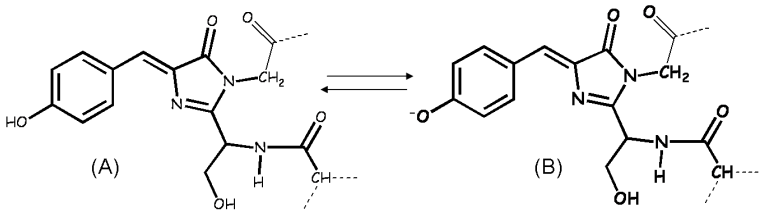



Figure 1.6: The two thermodynamically stable states of GFP chromophore: in *A* form Tyr66 carries a phenol, while in *B* form it loses the proton and becomes a phenolate

carring a phenol (A form) and the other a phenolate (B form), which are responsible for the two peaks of absorption at 3.12 eV (ultraviolet radiation), attributed to the neutral A form, and at 2.61 eV (visible light) [41, 70], corresponding to the anionic B form [79].

There is an equilibrium between ground state A and B forms governed by changes in protein concentration, ionic strength, pH and temperature. This equilibrium has been extensively investigated both in the ground [18–20] and in the first singlet excited state [21–23].

Emission occurs at 2.44 eV (corresponding to 3.12 eV absorption) and 2.46 eV nm (corresponding to 2.61 eV absorption) [65]. The evidence that excitation of both neutral and anionic form determines similar emission spectra is probably due to the fact that the phenolic oxygen of Tyr 66 becomes more acidic in the excited state, so a proton transfer occurs, originating a common anionic excited

state responsible for the emission [65, 68, 69].

When GFPs are excited through visible light they fluoresce in the UV part of the spectrum and they undergo an internal photoconversion process. To have a complete view of the knowledge of GFP photochemistry state-of-art it is necessary to analyse the models proposed to describe GFP interaction with radiation during this photoisomerization.

The first proposed was the *three-state photoisomerization model* [68, 80], shown in Figures 1.7 and 1.8(upper panel). According to this process, the neutral species gives origin to the anionic species passing through an unstable intermediate state, apparently corresponding to the anionic chromophore in a nonequilibrium protein environment (see Figure 1.7). The first step is the passage from the neutral form to the anionic intermediate (proton transfer - PT), characterized by the release of Tyr66 phenolic proton, which is shuttled to Glu222, passing through a hydrogen bonds network, including a water molecule (water 25 of 1GFL PDB structure) and Ser205 (see Figure 1.9). The second step is the passage from the intermediate to the anionic form, characterized by a conformational change of the protein environment.

As some experimental data do not confirm this model, it has been proposed the *four-state photoisomerization*

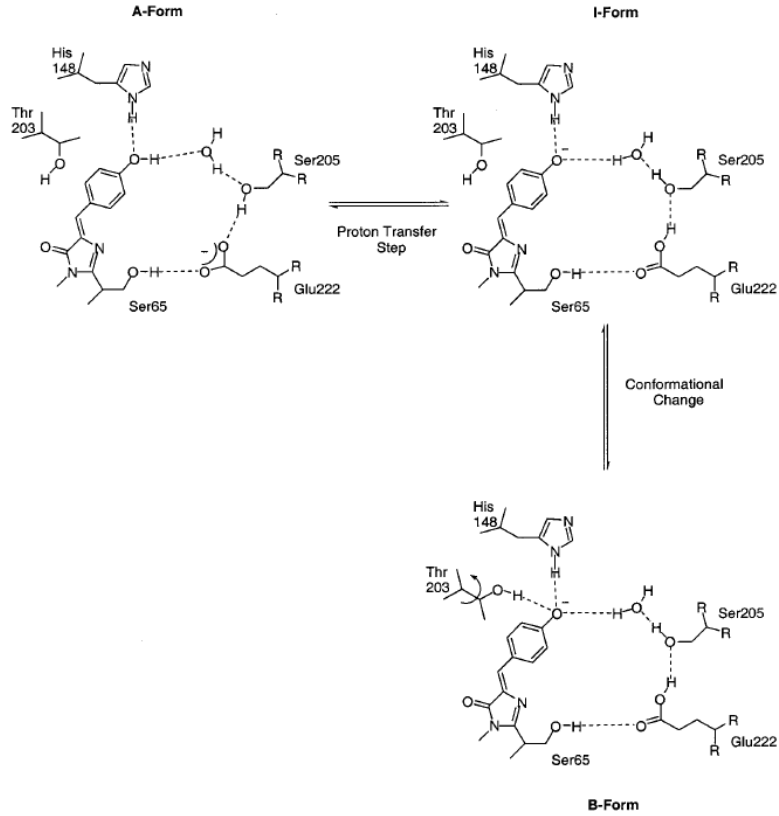



Figure 1.7: Proposed mechanism for the photoisomerization of wild-type GFP. The neutral form of the chromophore (A) can convert to the anionic stable species (B) by going through the anionic intermediate state (I).

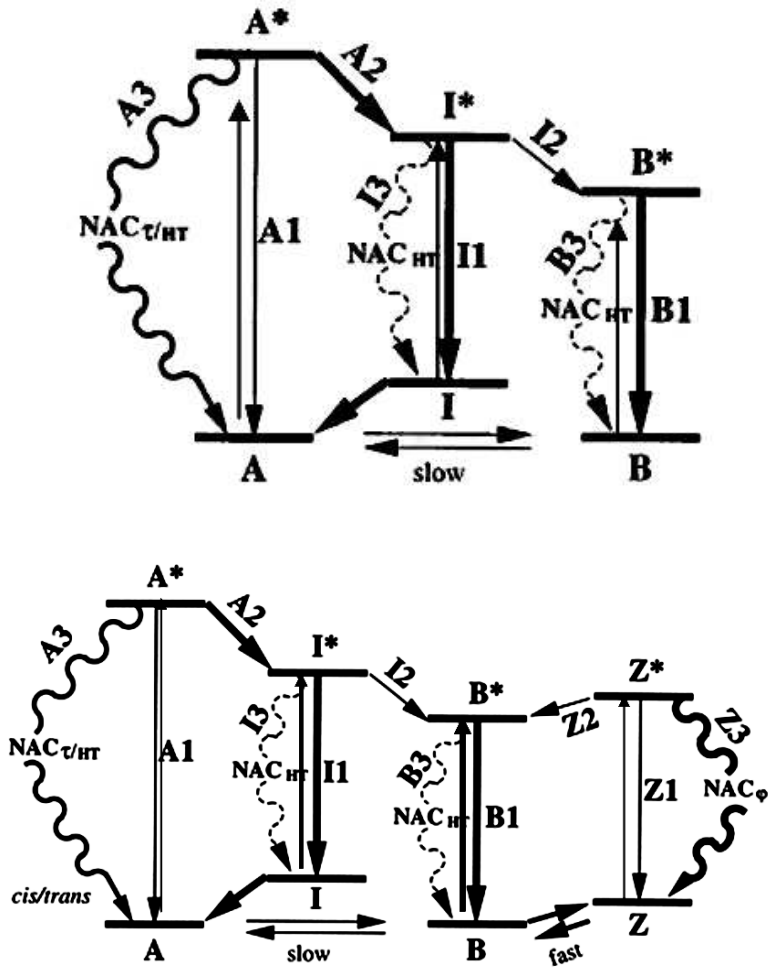



Figure 1.8: Three-state photoisomerization model (upper panel): A is the neutral form of the chromophore, B is the anionic species and I is the intermediate. Excited states are labeled by asterisks. Four-state photoisomerization model (lower panel) for the photophysical behaviour of GFP [74].

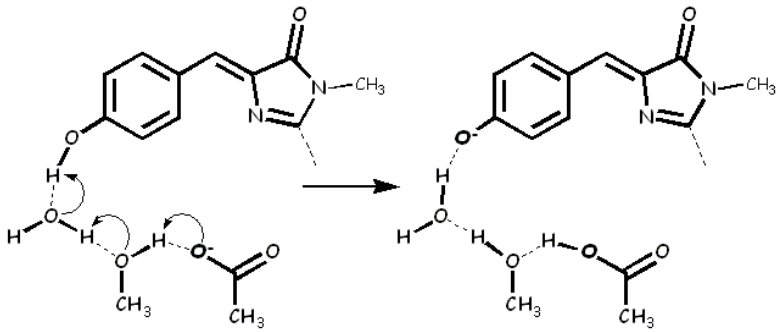


Figure 1.9: Proton shuttle mechanism: the proton belonging to Tyr66 is shuttled to Glu222 passing through a hydrogen bonds network, including a water molecule and Ser 205.

*model* [74], which considers the presence of an additional zwitterionic form, Z, which in its excited state, $Z^*$, has a preferential tendency to give radiationless decay (see Figure 1.8). The introduction of this additional species might explain the peculiar behaviour revealed by spectroscopic measurements on single GFP molecules: on/off blinking on the seconds (about 1 s) time scale could be attributed to an equilibrium between the emitting anionic form and a dark intermediate and the switching behaviour on the minutes (about 5 min) time scales, could be attributed to the existence of this dark state [81–85]. Furthermore, a shoulder recorded on the red edge (475 nm) of the absorption spectra could be attributed to the presence of the zwitterionic species [70]. The presence of a cationic

species has been also proposed, obtained by the protonation of the heterocyclic nitrogen close to the $C_\alpha$ of Tyr66 [73, 86]. However, a Raman study on GFP [87] and other experimental evidences [88] do not support the hypothesis of the existence of this form.

However the more corroborated photoisomerization model is the three-state one and so the main species involved into GFP photochemistry are: the neutral form (A form), an anionic intermediate (I form) and the anionic stable form (B form). Here we briefly summarize the main knowledge about these species. The neutral form is the only whose structure is well known (1GFL PDB code). For the other two forms there is still uncertainty on the structure, but the main doubts are related to the hypothesis of the excited state evolution. There are different structures of GFPs and of their mutants which are considered to be in the anionic form or in equilibrium between the neutral and the anionic form (1EMA, 1EMG, 2WUR, 1HCJ PDB codes). For the anionic intermediate there are only structural hypothesis, not supported by the lack of experimental data. Theory can give a significant contribution to the definition of the structure of each GFP form and to find the relation between structural and optical properties. Among the many theoretical works bound to define the structure of the different forms of GFP there is the

study conducted by Tozzini et al. [89] in 2001, in which they accurately studied structural, electronic and vibrational properties of the four forms (neutral, anionic, zwitterionic and cationic) of GFP isolated chromophore in the ground state and in the first excited state through DFT [1, 2] based techniques and ab-initio Car-Parrinello molecular dynamics (CPMD) [90]. In addition, in a following work Tozzini et al. [91] found a quantitative relationship between the optical and the structural properties of the neutral chromophore of GFP. This was accomplished by defining general rules which create a relation between structural parameters of the chromophore and optical response, depending on different environments.

**Anionic GFP**

The anionic form of GFP is the most studied because that is the fluorescent form and the comprehension of the molecular mechanisms determining the fluorescence and their control allow to project new and advanced systems. The chromophore fluorescence depends on the environment. In fact, it does not fluoresce if the protein is denaturated or if put in solution without the protein bulk. This suggests that the fluorescent state undergoes a barrierless deactivation process if in solution, but presents an

high barrier if it is in protein. This behaviour is probably caused by environmental specific electronic or steric effects. Have a deep comprehension of this ultrafast decay photochemical mechanism will determine a huge enhancement in using GFPs as biosensors. Understanding the photochemistry of the isolated GFP chromophore in the anionic form ($HBDI^-$) is mandatory to control the photochemical properties in the protein. The anionic chromophore could be described by two main resonance formulas which have similar influence on determining the real structure of the system. This generates the delocalization of the negative charge on the imidazole and the phenolate rings and a huge mixing of the bond order. Unfortunately, the $HBDI^-$ absorption measurements in vacuum have been very challenging because of the low photodetachement energies, laying near the $S_0 \rightarrow S_1$ VEE [92]. The fact that the lowest bright excited state lies above the electron-detachment threshold in the gas phase make these bound excited states metastable with respect to electron detachment, which has implications in the computation and interpretation of the $HBDI^-$ excitation energies [92, 93]. The lack of a certain experimental reference for the chromophore absorption in vacuum is due to this metastability. Pioneering action spectroscopy experiments carried out on $HBDI^-$ anion in the gas-phase by

Andersen [94, 95] and Jockusch [96] revealed a maximum in the absorption value around 2.58 eV, which is very close to the absorption maximum of the protein (2.61 eV). In our opinion is quite speculative that the GFP scaffold offers a near-vacuum environment for the chromophore, in fact the absorption of $HBDI^-$ anion in DMSO and DMF is found to be similar to that in protein and, as already noted, the chromophore in GFP is red shifted compared with $HBDI^-$ in solution but interacts strongly with the protein through H-bonding. In contrast, the $HBDI^-$ anion spectrum is strongly blue-shifted in H-bonding solvents [24, 97]. The spectroscopic behaviour of $HBDI^-$ anion has been also extensively experimentally studied in several solvents showing a wide absorption range (from 2.57eV to 2.91eV) and short fluorescence life times in solution, without a neat dependence on solvent polarity and suggesting strong stereo-electronic protein influences on its electronic properties [24]. The group of Dong presented a more energetic gas-phase value, 2.84eV, extrapolated from solution-phase experiments [24] and recently photodestruction experiments of Zenobi et al. asserted that the gas-phase absorption value, promoted by a single-photon, is $> 2.61$ eV, blue shifted compared to earlier results obtained with pulsed lasers [98]. The metastability makes a system difficult to be described not only exper-

imentally, but also theoretically. A lot of computational studies have been conducted to shed light on this matter. Sophisticated Post-HartreeFock (post-HF) methods have presented several excitation values [74, 99–103], above all, CASPT2 2.67 eV [75], State-Averaged CASSCF 2.69 eV [104] and more recent studies: CASPT2 of Filippi et al. 2.76 eV [30] and MRMP2 based on state average CASSCF of Krylov et al. 2.61 eV [105]. Several studies based on Time Dependent Density Functional Theory [4–6, 106] (TD-DFT) have also been proposed obtaining excitation energies in 2.94 - 3.10 eV energy range [30, 107, 108] . DFT and TD-DFT have difficulty in the accurate description of metastable states and charge-transfer processes. Post HF methods are less invalidated by this feature of the chromophore, but require a very high computational cost and are not suitable when it is necessary to introduce the solvent or the protein effects on a small model. The $HBDI^-$ anion has also been theoretically studied in water solution by implicit and explicit models [107, 109–111] and in other solvents by implicit models by Nemukhin et al. [107]. Having considered the anionic chromophore it is important to understand what it is known about the anionic form of the protein. The GFPs Class 2 (EGFP - according to Tsien classification of the green fluorescent proteins [112]) that stabilises mainly the B form, presents

an enhanced red shifted (2.54 eV) [113] anionic band without any absorption in the A form range; therefore it represents an optimal experimental structural GFP benchmark model in which the chromophore is the anionic form.

### Protein influence on the isolated chromophore optical properties

For all GFP forms it is evident that even if the majority of the contribution to the optical behaviour is due to the chromophore, it is really difficult to accurately define how the protein environment influences the intrinsic properties of the fluorescent protein, such as the position of absorption and emission peaks and the possible excited state reactions. The $\beta$-barrel, presumably reduces the rate of radiationless decay by restricting the conformational space of the chromophore [77], leading to ca. $10^4$ higher fluorescence quantum yields. At the same time, the $\beta$-barrel can apparently tolerate a high degree of disorder within the protein, as judged by the analysis of the numerous crystal structures, as well as a variety of $\pi$ systems . The $\beta$-barrel obviously plays an enormous role in the photophysics of GFP. Many experimental works have been conducted in order to precisely identify the contribution of the environment, for example Dong et al. [24]

have introduced the chromophore into different solvents of various polarity, trying to reproduce the effect of the $\beta$-barrel. Their results shows that no solvent is able to reproduce the effect of the barrel and no solvent effect serves to turn on the fluorescence, which remains subject to stereo electronic effects. The contrast between solution and protein-constrained behaviour supports our idea that the protein environment plays a fundamental role in GFP photochemistry. Theoretical studies can provide an invaluable contribute in order to clarify these issues. In fact, they allow us to analyze the electronic detail of the system and find the relation among the electronic, structural, dynamical properties and photochemistry.

**Theoretical approaches to describe GFP system**

Looking at the vast theoretical literature produced on GFPs it is usual to find that the large spread of approaches used to describe GFP system brings to an equally large spread of results and predictions. Despite significant theoretical progress in electronic structure methods, it is far from trivial to accurately compute excitation properties of even relatively small photoactive molecules. The difficulties arise from the rather strict requirements the theoretical approach must meet to provide a predictive description

of a photoactive biosystem. It should give an accurate and balanced description of the ground and the excited states of the photoactive site and also be able to treat a realistically large model of the biosystem. Therefore, despite their less favorable scaling with system size, highly correlated quantum chemical methods, such as complete-active-space second-order perturbation theory (CASPT2) [114], have established themselves as important theoretical tools to compute excitations of photosensitive [30] small systems. In fact, post Hartree-Fock methods allow an accurate description of limited size systems, while they are not performing when scaling the system extent. In fact, the computational cost (time to spend and performance of the machines) they require to treat a large system is very high and usually not compatible with research times. Sometimes for relatively large systems it could be taken a compromise in the choice of the size of the active space or of the basis set. Density Functional Theory (DFT) [1–3] and Time-Dependent Density Functional Theory (TD-DFT)[4–7] are the most appealing approach to efficiently compute the excitations of large molecular complexes. In fact, DFT permits to treat the electronic problem in an effective way, maintaining a comparable accuracy with respect to the most sophisticated post Hartree-Fock methods and requiring accessible com-

putational costs. However, on the other hand, DFT and TD-DFT lack of accuracy when considering systems involving charge transfer excitations[8, 9] or electron transfer processes. The best choice to describe large-sized systems is the use of hybrid methods (e.i. Quantum Mechanics/ Molecular Mechanics methods - QM/MM) through which it is possible to define more "layers" within the molecular structure that are treated at more levels of theory. These methods represents a good compromise between accuracy and computational cost.

It is worth noting that the majority of theoretical works in literature are focused on isolated chromophores [73, 75, 108, 115–117], ignoring the effect of the protein environment. Some works investigated the importance of the environment on GFP photochemistry, for example, Patnaik et al. [77] carried out a molecular dynamics simulation both for neutral and anionic forms of the chromophore. They have made a detailed analysis of structural modification of the chromophore, using a ONIOM potential: DFT for the high level of theory (chromophore) and the universal force field (UFF) for the entire molecule. Their data have shown that the protein shell has a double function: on one hand it mostly reduces the chromophore flexibility, because of its rigid nature, on the other it stabilizes a wide network of hydrogen bonds between the chro-

mophore and the neighbouring residues. Although the employment of quantum mechanics/ molecular mechanics (*QM/MM*) based techniques is a useful tool of detection in this field, a UFF potential is not suitable for the description of a protein environment.

In order to evaluate the protein contribution Tozzini et al. [91] have monitored the effect of each functional group in close contact with the chromophore. In particular, they identified the important role of Arg96 and Glu94, attributing them a red shift of 30-40nm with respect to the isolated chromophore.

The experimental value of wt-GFP neutral absorption (3.12 eV) [41, 70] has been theoretically reproduced in different ways. For example, Rubio et al. [67] adopted a QM/MM technique to extract a realistic structure for the GFP chromophore for which the absorption spectrum was calculated at the time dependent - density functional theory (TD-DFT) level [4–7], giving a peak at 3.01 eV.

For the anionic GFP the main theoretical effort has been employed to model the vertical transition in protein mimic environment in order to validate and explain the wt-GFP effects, obtaining a good accuracy in the prediction of wt-GFP B form absorption value. Since the first studies of Zimmer and al. [118, 119], it has been highlighted that the protein matrix has a crucial role in excited state chro-

mophore behavior and relaxing. The first post HF [25] and DFT [91] studies have showed the importance to investigate the effect of the protein on transition. Finally, the more recent works of Krylov [92] and Filippi [30] have analyzed the different vicinal residue effects on excitation and vertical detachment energies (the first) or have showed that the conditions GFP chromophore in the protein are not close to those in vacuum (the last).

Here we focused to model the vertical $S_1 \leftarrow S_0$ transition of the anionic $HBDI^-$ in the gas phase, in several solvents, and the protein.

**Proton transfer mechanism**

The proton transfer is the passage from the neutral form to the anionic intermediate, characterized by the release of Tyr66 phenolic proton, which is shuttled to Glu222, passing through a hydrogen bonds network, including a water molecule (water 25 of 1GFL PDB structure) and Ser205 (see Figure 1.9).

According to several studies [120], this latter is only an intermediate structure, which through a conformational change of the protein environment, gives the anionic form responsible of the fluorescence.

The characterization of both the intermediate and the an-

ionic species lifetime is a crucial step to understand the GFP photochemestry. In this respect, the elucidation of the PT mechanism and kinetics is a prerequisite, and the debate about the occurrence of a concerted or a step-wise mechanism is still open. As regards to the study of proton transfer mechanism, many hypothesis have been taken into account [68, 121–126]. One of the most recent works by Gelabert et al. [76], focalized the attention on the first step of this process, the proton transfer to a nearby water molecule. The authors proposed the existence, along the proton transfer coordinate, of a $\pi\sigma^*$ electronic state that intersects with the photoactive $\pi\pi^*$ electronic state. This state could constitute a possible non-radiative deactivation pathway of the photoexcited neutral form of the chromophore.

On the basis of the considerations above, it is easily understandable that a complete picture of GFP spectral properties is still a challenging goal. The detailed photochemical behaviour is not revealed yet, as there are experimental data contrasting. There is an open debate about the best theoretical method able to treat GFP, in particular regarding the capability of TD-DFT method to accurately describe GFP system, the influence of the environment on the isolated chromophore optical properties, the pro-

ton shuttle mechanism, the residues involved in it, the time scales of each step and of the whole process.

To treat GFP system we use a ONIOM [10–13] scheme, which allows us to accurately consider the influence of the protein environment, even because we use DFT[1–3]/Amber[127] and TD-DFT[4–7]/Amber[127] to respectively describe the ground and the excited state. We have analyzed the anionic chromophore optical absorption in gas phase and in several environments, reproducing and understanding for the first time, the observed trend of the chromophore absorption peak in the protein, ethanol, dioxane, methanol and water. As a consequence, we have been able to analyze with confidence the relative weight of different solvation effects affecting the vertical excitation energy and of the protein.

We have well reproduced the experimental red shifts in both the neutral and anionic GFP and analyzed in detail the contribution of the protein environment on the absorption band, identifying three different effects, structural, electrostatic and quantistic.

We have conducted a deep analysis of the proton transfer mechanism. First of all to handle a model which could accurately describe the effect of the protein environment on the transfer reaction, we have obtained QM/MM GFP structures for the neutral species and the anionic interme-

diate in both the ground and the excited state involved into the absorption. Then we have extrapolated reduced models from these structures and proposed a mechanism for the proton transfer within these models: we have followed the proton transfer reaction by integrating the intrinsic reaction coordinate (IRC) on both the $S_0$ and $S_1$ potential energy surfaces.

This thesis is organized as follows: in the section of *Methods and Models* we are going to describe the methods we have adopted to obtain an accurate representation of GFP system and its optical properties and the models built. We have also reported the validation tests performed to obtain the best description of the GFP electronic energy for both the ground and the excited states. In the section of *Results and discussion* we are going to report our detailed analysis and discuss our main goals regarding the chromophore optical behaviour, the protein environment effect on the isolated chromophore and the proton transfer mechanism. In the *Conclusions* we are going to contextualize our main results and express future perspectives.

# 2 Methods and Models

*If you can trust yourself, when all men doubt you,*
*but make allowance for their doubting too...*
*yours is the Earth and everything that is in it*
*and, which is more, you' ll be a Man.*
*(R. Kipling)*

In this section are briefly described the methods we have adopted to obtain an accurate representation of GFP neutral and anionic structures, to carry out energetic calculations on GFP both the ground and the excited states, to analyze the solvation behaviour of the chromophores and evaluate hypothesis for the proton transfer mechanism. All the calculations were carried out by using the density functional theory (DFT) and its time dependent version (TD-DFT) formalism [4–7]. Implicit solvent effects were taken into account by the conductor-like version of the polarizable continuum model (CPCM) [128–130]. The protein environment were modeled recurring

to the N-layered integrated molecular orbital and molecular mechanics [10–13] (ONIOM) scheme, in combination with both the TD-DFT [131, 132] and the PCM [133, 134] formalism.

## 2.1 Methods

### 2.1.1 Time independent and dependent density functional theory

In this work all the quantum mechanical calculations were carried out by using density functionals. Methods based on the TD-DFT [4–7, 106, 135] were chosen for the quantum mechanical study of the excited states.

DFT [1, 2] is the study of the one-to-one correspondence between an interacting many-electron system and a fictitious non-interacting analog, the Kohn-Sham system, whose equations are much easier to solve numerically. This mapping is exact in principle, but must be approximated in practice. Over the last several years, there has been an explosion of interest in DFT, driven largely by its applications in quantum chemistry. This is due to the recent progresses made to reach a higher accuracy of available approximations and to the wide number of chemical problems that can be tackled with such a computationally in-

expensive tool. Accurate DFT applications yield predictions of atomic energies, reaction energies in chemistry, cohesive energies in solids, vibrational energies, phonon spectra, activation barriers, rotational energies, etc. A huge number of quantum mechanical methods for computing the properties (energy, oscillator strengths, eventually the minima, etc.) of the excited electronic states of medium-size molecules is currently available. Very schematically, we can distinguish between wavefunction-based methods and electron-density-based methods. In some methods of the former family, such as single-reference configuration interaction (CI) or multiconfigurational-based ones, multiconfigurational SCF (MCSCF), for example, the CASPT2 [114], or multireference CI [136], any electronic state is described as the combination of several Slater determinants corresponding to different electronic configurations (i.e., different occupation schemes of the molecular orbitals, MOs). The expansion coefficients of the different Slater determinants and, in multiconfigurational SCF approaches, the expansion coefficients of the MOs in the Slater determinants are then variationally computed. A different approach is followed by other wavefunction-based methods as those belonging to the coupled-cluster (CC) family [137, 138]. CC models are based on the single-reference wavefunction and allow us

to compute excitation energies within the equation-of-motion and linear response CC formalisms (EOM-CC and LR-CC, respectively). Accuracy of CC results depends on the level of truncation in the CC expansion. In this respect, extensions of CC theory for excited states to include triplet excitations within both iterative and noniterative schemes allow for very accurate computations of excited-state properties. However, due to their high computational cost, these methods cannot be applied to large molecules. A second class of methods is instead based on the knowledge of the electron density in a theoretical framework similar to that of the DFT. TD-DFT is a quantum mechanical theory used in physics and chemistry to investigate the properties and dynamics of many-body systems in the presence of time-dependent potentials, such as electric or magnetic fields. TD-DFT is hence an extension of DFT, and the conceptual and computational foundations are similar: the time-dependent wave function is equivalent to the time-dependent electronic density, and an effective potential of a fictitious non-interacting system can be derived which returns the same electronic density of the real (interacting) one. The TD-DFT is an exact quantum mechanical theory in which the time-dependent density is the fundamental variable and the exchange-correlation (XC) potential describes the many-body in-

teractions [139]. For small changes in the time-dependent XC potential, the linear response approach can be applied to solve the TD-DFT equations. In this way it is possible to obtain the relative energies of the excited states (the excitation energies) as poles of the frequency dependent ground state linear response function [140]. TD-DFT provides a fast and reliable approach to obtain potential energy surfaces for the excited states as a function of the molecular geometry by simply adding the ground state DFT energy to the excitation energy of the selected state. In addition, the excited state first order properties (e.g. forces on the nuclei, electric multipole moments) can be expressed via the Helman-Feynman theorem, as the first derivatives (gradients) of the excited state energy with respect to suitable external perturbations [141]. In this framework the availability of analytical TD-DFT gradients plays a strategic role in reducing the computational effort required by the exploration of the excited states potential energy surface (PES). Van Caillie and Amos pioneered this field [6, 7] formulating the theory of analytical geometrical derivatives for TD-DFT using local density approximation (LDA), gradient corrected and hybrid functionals. More recently, Furche and Ahlrichs [142] have presented the theory, the implementation, and the validation of several excited state properties obtained

from the TD-DFT. Their approach is based on a fully variational formulation for the excited state energy functional which allows for a very compact derivation of the first order properties, geometrical derivatives, and electronic density multipolar expansion. Benchmark results using hybrid functionals showed that structural properties of excited states are almost as accurate as those of the ground state calculations, at a comparable computational cost, even for fairly large molecules. For the excited states the TD-DFT recently emerged as a very effective tool, since, when coupled to suitable density functionals, it often reaches an accuracy comparable to that of the most sophisticated (but expensive) post Hartree-Fock methods, with a much more limited computational cost. As a consequence in the last years an increasing number of TD-DFT applications have appeared in the literature TD-DFT has often been criticized for not being a first-principle nonempirical method (in analogy with its "parent" DFT), not showing uniform accuracy in treating electronic transitions with different characters, and delivering a qualitatively wrong description of the crossing region between different electronic excited states. On the other hand, although wavefunction-based methods have known impressive methodological advances, significantly increasing their range of applicability, they suffer from

very high computational costs. As a consequence, some compromises concerning the basis set or the active space employed are often necessary, not only decreasing the expected accuracy but also introducing a significant degree of arbitrariness in formally rigorous methods. Furthermore, semiempirical parameters are often also present in sophisticated calculations, such as CASPT2. TD-DFT calculations can incorporate environmental effects [143] and quickly give UV/vis spectra for most organic and inorganic dyes. It has been proved to be really useful for the description of the electronic excitation of first and second row elements transition, noble and early transition metal systems and excited-state biological processes [67]. However, there are several systems/processes for which TD-DFT has often shown significant failures. TD-DFT is a mono-determinantal method, and thus it cannot be applied to electronic states with an intrinsic multireference character [144]. Analogously, TD-DFT can exhibit deficiencies in treating electronic transitions with substantial contributions from double excitations [8, 9, 145, 146], although interesting attempts to overcome the above limitations have been proposed. In several cases, however, an electronic transition exhibits a multireference character (or a significant contribution from double excitations) just because of a poor description of the ground-state

MO by HF orbitals, and such features are not present when using MOs computed at the DFT level. Another traditional failure of TD-DFT concerns the treatment of long-range charge transfer (CT) transitions between zero-overlap donor-acceptor pair. Standard functionals significantly underestimate the transition energy and fail to reproduce the correct 1/R trend when the donor/acceptor distance (R) increases [147]. However, in the last five years new functionals have been developed, which are able to deliver a correct estimate also of the long-range CT transitions [148–151]. Furthermore, it is worth noting that for electronic transitions which involve only partial CT character, the underestimate of the excitation energies by TD-DFT may be controlled by the use of hybrid functionals, whereas the performances of pure functionals are much poorer. Hybrids include a fraction (R) of exact exchange that is computed with the Hartree-Fock (HF) exchange formula [152–157]. Despite their countless successes, hybrids also encounter problems that seem (mostly) independent of the functional selected. Typical troublesome properties include van der Waals forces [158], bond length alternation (BLA) in semiconducting polymers [159, 160] nonlinear optics (NLO) properties of long - conjugated chains [161, 162] and charge-transfer electronic transitions [146, 163–165]. In these four cases, no single R

value provides a small (or consistent) error for increasingly large/ spaced compounds. In fact, these DFT limitations have a common origin: the so-called shortsightedness of DFT functionals. In other words, the density is not influenced by a change in the nearby electronic distribution [146, 161, 164]. To circumvent these shortcomings, several strategies have been designed and applied to the problems listed above: the correction(s) of the self-interaction error [166, 167], the inclusion of the current-density in the formalism [168], the addition of empirical dispersion terms [169] and the use of optimized effective potential for exact exchange [170] as well as the explicit consideration of long-range effects [171, 172]. This latter scheme leads to the range-separated hybrids that use a growing fraction of exact exchange when the interelectronic distance increases. In contrast, the hybrids in which the amount of HF exchange is constant all over the space will be referred to as global hybrids in the following (conventional or full-range exchange hybrids have also been used in the literature). It has been demonstrated that range-separated hybrids are very efficient for calculating BLA [172] or NLO [171, 172] properties in conjugated polymers as well as for determining properties of weakly bond complexes [173] or charge-transfer states in large molecular systems [173]. Therefore, some of the deficiencies ascribed

to TD-DFT are not intrinsic features of the method but depend on the choice of the functional. Also long-range corrected functionals are not able, however, to accurately treat other classes of compounds such as cyanines, especially when the length of the ethylenic bridge connecting the two $NH_2^+/NH_2$ moieties increases. This failure is likely due to the intrinsic multideterminantal character of the electronic transitions in such molecules. When treating systems of biological and technological interest, such as GFPs, a method coupling accuracy and computational feasibility is necessary. We think that TD-DFT represents one of the best compromise between accuracy and computational cost for describing the excited state behavior in medium/large-size molecules. Regarding absorption, TD-DFT is known to lead systematic errors in the excitation energy [8, 145, 174]. However, the performance of TD-DFT has been assessed in several studies on many chromophores representative of fluorescent proteins [67, 175]. In these cases the results encourage the employment of TD-DFT approachs combined with QM/MM techniques in order to investigate the photochemistry of more realistic systems.

**Theoretical basis of Time-Dependent DFT**

The formal basis of TD-DFT is the Runge-Gross theorem [106], the time dependent analogue of the Hohenberg-Kohn theorem. The Runge-Gross theorem shows that, for a given initial wavefunction, there is a unique mapping between the time-dependent external potential of a system and its time-dependent density. This implies that the many-body wavefunction, depending upon 3N variables, is equivalent to the density, which depends upon only 3. Therefore all properties of a system can be determined from knowledge of the density alone.

Given the Runge-Gross theorem, a computationally useful method has to involve a fictitious non-interacting system which has the same density as the physical interacting system of interest. As in DFT, this is called the (time-dependent) Kohn-Sham system. According to the TDKS procedure, a linear response of the initial electronic density in the ground state can be calculated by a self consistent scheme. Excitation energies are characterized as the poles of the response functions.

The recent implementation of TD-DFT analytical gradients allows for the determination of the excited-state stationary points and their properties (e.g., the multipole moments). Harmonic frequencies can be obtained by performing numerical differentiation of the analytic gradients, enabling us to perform the same kind of vibrational

analysis performed in the ground electronic state. TD-DFT calculations thus allow for determining the energy and the properties of the excited states with a limited computational cost. As a consequence, it is usually not necessary to impose any symmetry constraint, very large basis sets can be used, and no ad hoc choice (see, e.g., the active space in CASSCF/CASPT2 calculations) is usually necessary, also when dealing with large-size systems. These are important features: On the one hand, it is possible to treat different systems (e.g., a supramolecular system and its component) and different kinds of transitions (e.g., n$\pi^*$ and $\pi\pi^*$) with a similar degree of accuracy, putting the analysis of the computational results on firmer ground. As a consequence in the last years the number of TD-DFT studies has continuously increased, allowing important advances in our knowledge of the potentialities and the limitations of this method. It is nowadays well assessed that TD-DFT, when employing a suitable density functional, can provide fairly accurate results (with 0.2-0.3 eV the experimental results) in several classes of systems, despite the limited computational cost. It is important, however, to remember that the accuracy of TD-DFT results depends on several factors (functional, basis set, etc.) which have to be properly addressed.

### 2.1.2 Quantum Mechanics/Molecular Mechanics hybrid methods

One of the main themes in computational chemistry is to find a balance between the accuracy of the results and the computational cost. However, the computational cost of accurate methods (generally based on quantum mechanical ($QM$) models) scales very unfavorably with the size of the system. Molecular mechanics ($MM$) methods are, on the contrary, quite inexpensive and attractive, but inadequate to describe phenomena which are 'quantum' by nature, such as the chemical reactivity, and require specific procedures in order to obtain parameters 'ad hoc' for the system considered.

The so-called *hybrid* methods offer a good compromise between the desired accuracy and the computational costs. The basic idea starts with the understanding that various regions of the molecular system play very different roles in the process under study. Examples of such cases are many enzimatic reactions, where the bond breaking and forming takes place only in the active site, and the effects of the protein environment is usually only steric or electrostatic. With hybrid methods each region is treated at a different level of theory. More specifically, expensive computational methods (QM) are only required for the

part of the system 'where the action takes place', while less expensive methods (MM) can be employed for the remaining region. In this regard, very accurate results can be obtained employing only a fraction of the computational cost of conventional methods.

QM/MM potential was presented as [176–178]:

$$E^{QM/MM-EE} = E^{v,QM} + E^{MM} + E^{QM-MM} \qquad (2.1)$$

$E^{v,QM}$ is the QM energy of the QM region, in the field v generated by the partial charges of the MM region and $E^{MM}$ is the MM energy of the MM region (containing all the bonded and no-bonded MM terms that involve esclusively centers from the MM region. $E^{QM-MM}$ describes the interaction between the two regions and has two components. First, if there is covalent bonding between the QM and MM region, it contains the "border crossing" bonded MM terms that involve both QM and MM centers. Second, $E^{QM-MM}$ contains all the MM van der Waals terms that involve one QM center and one MM center. Because $E^{v,QM}$ already includes it, $E^{QM-MM}$ does not contain the electrostatic interaction between the QM and MM region.

Kollman [177] also suggested a simplified potential, which

was further explored by Thiel [179].

$$E^{QM/MM-ME} = E^{MM} + E^{QM} + E^{Q,QM-MM} \qquad (2.2)$$

The QM energy, $E^{QM}$, no longer involves the potential from the MM region. Instead, the electrostatic interaction between the regions is calculated in $E^{Q,QM-MM}$, by assigning partial charges to the QM atoms and using the regular expressions for point charge interactions from the MM force field. Thiel referred to the QM/MM potential using equation 2.1, $E^{QM/MM-E}$, as electronic embedding QM/MM, and to the potential using eq. 2.2, $E^{QM/MM-ME}$, as mechanical embedding QM/MM. The advantages of electronic embedding are that the wave function can be polarized by the charge distribution from the MM region, and that it provides a more accurate description of the electrostatic interaction between the two regions. On the other hand, it appears that in many cases the accuracy of the mechanical embedding version is sufficient and the semplified expression facilitates the implementation of methods to explore the potential surfaces. When covalent interaction exists between the QM region and MM region, the dandling bonds need to be capped in the QM calculation.

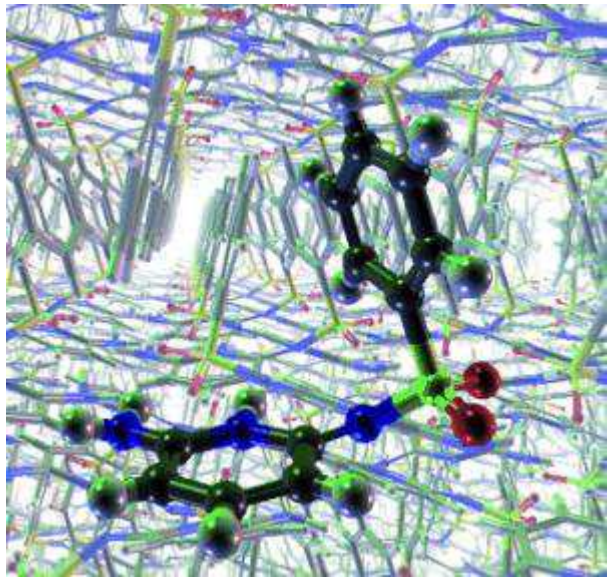


Figure 2.1: Sketch of a QM/MM method: the part of the molecule that is treated at quantum mechanical (QM) level is drawn in ball and bonds style, while the remaining part of the structure (drawn in finer bond style) is described with molecular mechanics (MM).

Over the years a variety of hybrid methods has been presented. These are conceptually very similar, but differ in a number of details.

A general scheme of the QM/MM phylosophy is reported in Figure 2.1.

In particular we adopted one of the most popular and robust hybrid model, namely 'Our own N-layered Integrated molecular Orbital- molecular Mechanics' (ONIOM) method

[10–13].

The energy of a system treated by ONIOM is expressed as

$$E^{ONIOM} = E^{model,QM} + E^{real,MM} - E^{model,MM} \quad (2.3)$$

where 'real' and 'model' refer to the full system and the QM region, respectively.

Besides some details concerning the bonded MM terms that involve both QM and MM atoms, the ONIOM expression (eq. 2.3) is essentially the same as the QM/MM-ME expression (2.2). $E^{model,QM}$ is equivalent to $E^{QM}$ and $E^{real,MM}$ - $E^{model,MM}$ in eq 2.3 describes both the MM region and the interaction between the two regions, similar to $E^{MM}$ + $E^{Q,QM-MM}$ in equation 2.2.

An important realization was that, unlike equations 2.1 and 2.2, all three terms in eq. 2.3 involve "chemical realistic systems".

From eq. 2.3 it is clear that the interaction between the layer is included at the low level method and that the method defaults to mechanical embedding for QM/MM combinations. However, ONIOM formalism was extended to include the electronic embedding too [11, 131]. Because the model system needs to be identical for both the QM

and MM calculation, we include the environment charges in both and do not change the real system calculation from eq. 2.3:

$$E^{ONIOM(QM:MM)-EE} = E^{v,model,QM} + E^{real,MM} - E^{v,model,MM} \tag{2.4}$$

To avoid overpolarization of the wave function, we may scale charges close to the QM region. Because these charges will then be scaled in both the $E^{v,model,QM}$ and $E^{v,model,MM}$ terms, the balance will not change. The charge interactions that are overcounted or undercounted at the QM level in the $E^{v,model,QM}$ will be balanced at the MM level in the $E^{v,model,MM}$ term.

The ONIOM method can be used also for more than two layers.

When covalent interactions between the QM region and the MM region exists, the not satured valence needs to be capped in the QM calculations. The simplest solution adopted by ONIOM is to use hydrogen atoms, which are, then, referred to as *link* atoms . A schematic use of H link is reported in Figure 2.2.

A further complication of covalent interaction is that there may be partial charges from the MM region very close to the QM region. Since in MM force field the interaction

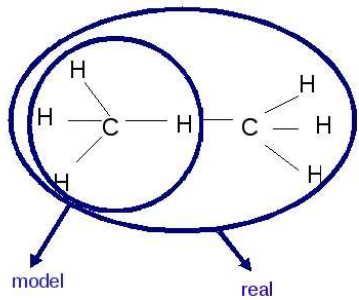



Figure 2.2: Link atoms into ONIOM calculations: The real system is the entire molecule, $C_2H_6$. The interface between the QM and the MM regions is located on a covalent bond, so it is necessary to saturate the valence. For this reason we use an H link, introducing in the QM region a C-H bond instaed of a C-C bond; therefore, the model is the $CH_4$ molecule.

between partial charges are scaled when they are less than three bonds apart, full inclusion of the partial charges in the boundary regions in $E^{v,QM}$ may lead to overestimation of the electrostatic interaction, while also the polarization of the wave function can be unphysical. Kollman [177] zeroed the charges that are less than three bonds away from the QM region. Although this will avoid overpolarization, it is rather arbitrary and also may lead to underestimation of the electrostatic interaction between the regions. Alternatives to zeroing the charges are to use delocalized (gaussian) charges instead of point charges [180, 181] or to redistribute the charges close to the boundary [182].
In Figure 2.3 we report the GFP structure represented

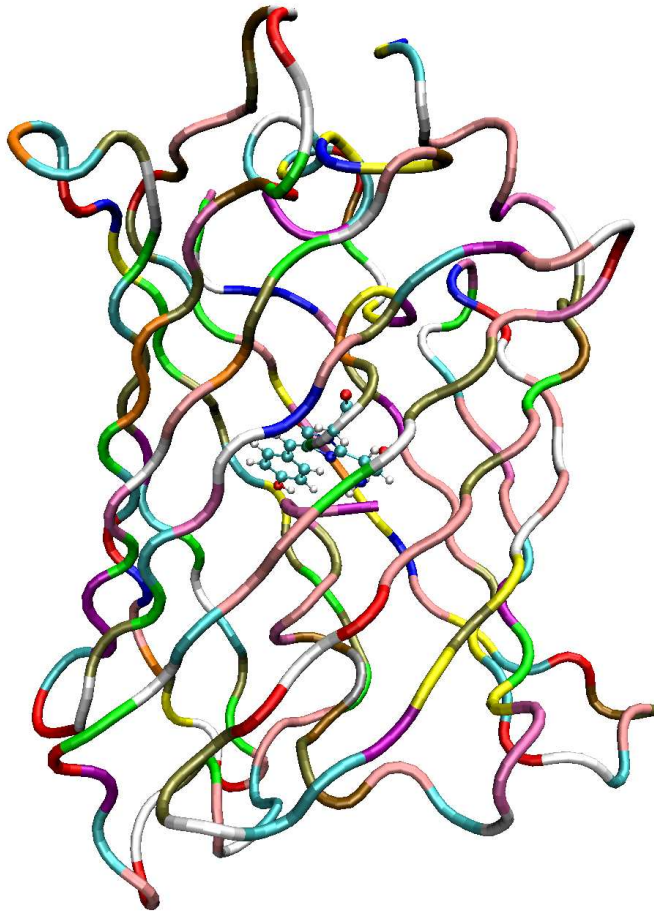


Figure 2.3: GFP structure obtained by employing ONIOM method: the core system (QM region) is drawn by ball and stick, while the remaining part of the real system (MM region) is drawn as tube

according to an ONIOM partition.
The QM/MM method used in this work is ONIOM as it is implemented in Gaussian 09 [183].

### 2.1.3 Linear Response and State Specific Polarizable Continuum Model

The polarizable continuum model [129, 130], PCM, describes the solvent as a structureless continuum, characterized by its dielectric permittivity, $\epsilon$, in which a molecular-shaped empty cavity is dug to host the solute which is fully described by its QM charge distribution. The dielectric medium is polarized by the solute charge distribution and acts as the source of a reaction field which in turn polarizes back the solute. This mutual effect is evaluated by solving the proper electrostatic problem described by the Poisson and Laplace equations with the proper boundary conditions at the cavity surface [184–186]. Nowadays, the acronym PCM indicates a family of methods which can be distinguished on the basis of the boundary conditions and the numerical approach used to solve the Poisson-Laplace electrostatic problem. In all formulations of PCM, the polarization of the medium is represented by the solvent reaction field expressed in terms of a potential defined through an apparent charge distribution, $\sigma$, spread on the

cavity surface.

For this thesis we used the CPCM [128, 187], which performs a PCM calculation using the CPCM polarizable conductor calculation model.

Inclusion of the solvent effect within a continuum model in time-dependent processes as absorption or emission poses several problems. On the one hand, the solution of the electrostatic problem is highly nonlinear. Indeed, the solvent reaction field should be variationally determined together with all the other parameters in the electronic method used (MO coefficients, CI coefficients, excitation amplitudes, etc), but all those parameters do depend on the solvent reaction field. Different approaches have been envisaged to tackle the above problem. The most used can be classified in two classes, that is, state-specific (SS) and linear response (LR) approaches [188–192]. In SS methods a different effective Shrodinger equation is solved for each state of interest, achieving a fully variational formulation of solvent effect on the excited-state properties. In the methods exploiting the LR formalism the excitation energies are "directly" determined without computing the exact excited electron density. We can consider in particular two aspects of the study of excited states in solution, regarding the definition of the energies of the excited states, thus affecting also their derivatives and properties.

- The electronic excitation is a process involving not only the solute but the entire solute-solvent system. As a consequence, the definition of the excited states of molecular solutes requires also the characterization of the solvent degrees of freedom. The difference of the characteristic time scale of the electronic degrees of freedom of the solute and the composite degrees of freedom of the solvent may lead to different excited state regimes, with two extreme situations the *nonequilibrium regime* in which the slow degrees of freedom of the solvent are not equilibrated with the excited state electronic redistribution upon excitation (vertical excitation processes), and the *equilibrium regime* in which the solvent is allowed to equilibrate (i.e., reorganize) all its degrees of freedom including the slow ones. The different regimes may dramatically influence the properties of the solute excited states.

- The second aspect concerns the status of the excitation energies of a solvated system. The state specific approaches, which are based on the explicit calculation of the excited state wave function, describe in a more satisfactory way the variation of the solute-solvent interaction accompanying the change of the electronic density during an electronic excita-

tion. On the other hand, the linear response methods introduce only effects related to the corresponding transition density.

Sometimes these differences have little effect on the excitation energies computed within the two different schemes (e. g. the calculations performed by Scalmani et al. [193] on the models of p-nitroaniline and 4-(dimethyl)aminobenzonitrile), while in other cases SS and LR methods provide very different estimates of the solvent effect on the excited-state properties and dynamical solvation effects. For this reason this point should be treated very carefully when using PCM in excited-state calculations.

**SS-PCM/TD-DFT**

The solvent reaction field contribution to the solute free energy (G) can be expressed in PCM as

$$G = \frac{1}{2}\mathbf{V}^{\dagger}\mathbf{q} \tag{2.5}$$

where vector $\mathbf{V}$ collects the values of the solutes electrostatic potential and $\mathbf{q}$ is the apparent surface charge placed at the center of the surface *tesserae* (i.e., the small tiles which the cavity surface is finely subdivided in),

where also $\mathbf{V}$ is computed. The polarization charges [186] depend on the solutes electrostatic potential and, thus, on its density through a general relationship of the form

$$\mathbf{q} = -\mathbf{D}\mathbf{V} \tag{2.6}$$

where the square matrix $\mathbf{D}$ is related to cavity geometric parameters and the solvent dielectric constant $\varepsilon$.

Consider a generic excited electronic state (2) together with the corresponding ground state (1). As discussed by Improta et al. [191], in SS methods the excited-state equilibrium free energy in solution, $[\mathtt{G}_{eq}{}^{(2)}]$, explicitly depends on the excited state (2) density:

$$\mathtt{G}_{eq}^{(2)} = \frac{1}{2}\sum_i q_i^{(2)} V_{i,\rho}^{(2)} = \frac{1}{2}\sum_i q_{i,f}^{(2)} V_{i,\rho}^{(2)} + \frac{1}{2}\sum_i q_{i,S}^{(2)} V_{i,\rho}^{(2)} \tag{2.7}$$

The nonequilibrium ($\mathtt{G}_{neq}{}^{(2)}$) free energy in solution involves an explicit dependence on the density of the ground state (1):

$$\mathtt{G}_{neq}^{(2)} = \frac{1}{2}\sum_i q_{i,f}^{(2)} V_{i,\rho}^{(2)} + \left(\sum_i q_{i,s}^{(1)} V_{i,\rho}^{(2)} - \frac{1}{2}\sum_i q_{i,s}^{(1)} V_{i,\rho}^{(1)}\right) + \\ + \left(\frac{1}{2}\sum_i q_{i,s}^{(1)} V_{i,f}^{(2)} - \frac{1}{2}\sum_i q_{i,s}^{(1)} V_{i,f}^{(1)}\right) \tag{2.8}$$

In the above equations $q_f$ /$q_s$ and $V_f$ /$V_s$ are the polarization charges and the corresponding potentials relative to the "fast" and "slow" solvent degrees of freedom. The potential generated by the density of state (n) is given as ${V_\rho}^{(n)}$. The absorption process is ruled by nonequilibrium solvation, and thus the solvent contribution to the vertical excitation energy can be computed by using the following relationship:

$$\Delta G_{abs} = \mathtt{G}_{neq}^{(2)} - \mathtt{G}_{eq}^{(1)} \tag{2.9}$$

Soon after the electronic transition has occurred, the system starts evolving on the excited-state potential energy surface (PES) toward its energy minimum. At the same time, slow solvent degrees of freedom start equilibrating on the excited-state electron density. These two processes cannot be rigorously decoupled, especially when they exhibit similar time scales, and we cannot thus expect that

a single strategy is suitable to all the possible emission processes. In many cases, however, this complex scenario can be simplified by means of qualitative considerations on the properties of the solvent and/or the excited potential energy surface. The equilibration of intramolecular degrees of freedom is faster than solvent equilibration except for very flat PESs or when many low-frequency motions are involved (large-amplitude torsional motions, inversions, etc.). This is especially true in polar solvents and for electronic transitions involving significant variations of the excited-state electron density. In such cases, indeed, time-resolved experiments suggest that the equilibration of the slow solvent degrees of freedom occurs on the picosecond time scale. This time should be long enough to assume that the excited electronic state has reached its minimum. A simple limiting case is that of ultrafast excited-state decay when only fast solvent degrees of freedom are expected to be in equilibrium with the excited-state density. In this limit, $\Delta \mathrm{G}_{em}$ can be computed exactly in the same way as $\Delta \mathrm{G}_{abs}$ [191]:

$$\Delta G_{em} = \mathtt{G}^{(2)}_{neq} - \mathtt{G}^{(1)}_{eq} \tag{2.10}$$

Of course, in this case excited-state geometry optimizations should also be performed in the nonequilibrium limit. Another simple limit is that of *very long* excited-state

lifetimes, which characterize, for instance, strongly fluorescent species. In this case, we can assume that all the solvent degrees of freedom are in equilibrium with the excited-state density. The ground state ($\mathtt{G}_{neq}{}^{(1)}$) nonequilibrium free energy in solution describing the emission process can thus be obtained from Eq. 2.8, interchanging labels 1 and 2. The fast solvent degrees of freedom are equilibrated with the ground-state electron density, whereas the slow ones are kept frozen at the value obtained in the equilibrium calculation of the excited state. In this limit, excited-state geometry optimizations should be performed with the solvent equilibrium limit, and the solvent contribution to the fluorescence energy ($\Delta$ G$_{em}$ ) is given as

$$\Delta G_{em} = \mathtt{G}_{eq}^{(2)} - \mathtt{G}_{neq}^{(1)} \tag{2.11}$$

The above relationship is the most suitable for treating phosphorescence, where states (2) and (1) correspond to $T_1$ and $S_0$, respectively.

The computation of the quantities involved in Eqs. 2.10 and 2.11 is straightforward using a generalization of the SS-PCM/TD-DFT method presented by Improta et al. [191], where the nonlinear problem of determining the polarization charges corresponding to the excited-state density is solved by using a self-consistent iterative procedure.

Starting from a TD-DFT calculation, a first approximation to the state-specific reaction field is computed using the electron density of the state of interest by solving Eq. 2.6. In the next step, a TD-DFT calculation is performed in the presence of this first set of polarization charges, providing an updated excited-state density and, consequently, a new set of polarization charges. This iterative procedure is continued until convergence on the reaction field is achieved. In the cases examined until now, 4/5 iterations are usually sufficient to reach a convergence $\leq$ 0.0001 a.u. on the final energy. The final equilibrium and nonequilibrium energies of the state of interest are easily determined by adding the corrections obtained by Eq. 2.7 or 2.8 to the excited-state energy provided by the TD calculation.

### LR-PCM/TD-DFT

In LR-PCM/TD-DFT the excitation energies are "directly" determined without computing the excited-state density by plugging in a PCM contribution in the TD-DFT equations [143].

A significant part of solvent effect on the excited-state energies is recovered by LR approaches using MO orbitals computed in solution (and, thus, including the polariza-

tion due to the solvent reaction field) [188, 189]. However, since the exact excited-state electron density is never computed, all the solvent contributions depending on the variation of the multipole moment upon excitation are missing in LR computations. An additional PCM correction is instead introduced which depends on the ground-excited state transition density [83]. As a consequence, the treatment of dynamical solvent effects is completely different with respect to that made by SS-PCM (in Eq. 2.8 explicit reference is made to the excited-state density). Furthermore, in the LR-PCM method the ground state is thus always fully equilibrated with the solvent degrees of freedom. As a consequence, in the nonequilibrium case the solvation contribution to the emission energy is computed as

$$\Delta G_{em}(LR) = \mathtt{G}^{(2)}_{neq}(LR) - \mathtt{G}^{(1)}_{eq} \tag{2.12}$$

whereas for long living excited state it is computed as

$$\Delta G_{em}(LR) = \mathtt{G}^{(2)}_{eq}(LR) - \mathtt{G}^{(1)}_{eq} \tag{2.13}$$

that is, in standard LR-PCM the equilibrium solvation energy for the ground state is used in both cases, making this method not suitable to an accurate treatment of the solvent effect on the emission process.

## 2.2 Models

### 2.2.1 Wild type neutral GFP

#### Neutral chromophore models

Neutral GFP chromophore consists of a p-hydroxybenzylideneimidazolinone (p-HBI) [194]. For our project we have constructed two models for the chromophore, namely HBI and HBDI (which differs from the first in the termination with two methyl groups instead of the hydrogens in the imidazolinone ring), represented in Figure 2.4.

#### GFP structure and QM/MM partitions

The *Aequorea victoria* GFP x-ray structure, obtained by Yang, Moss and Phillips in 1996 [61], has been chosen as starting reference (**1GFL** identification code from the Protein Data Bank).

The structure is a dimer in aqueous solution. However, previous theoretical and experimental studies have shown that the $\beta$-can is a stable and rigid structure and protects the chromophore from solvents and external agents [66]. For this reason the solvent can be, in a first approximation, neglected in a theoretical study of the GFP

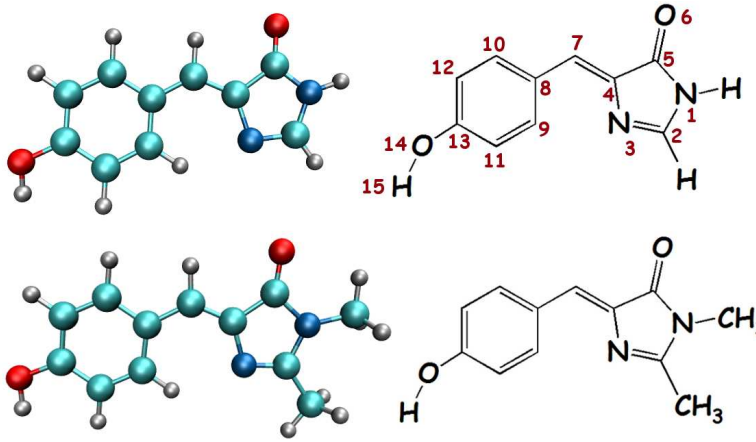



Figure 2.4: Minimal model (upper panel), HBI, and methyl-terminated model (lower panel), HBDI, for GFP neutral chromophore. On the left ball and stick representation and on the right chemical sketch.

photochemistry. Furthemore, the two units of GFP are identical, eachone has its own chromophore and the same spectral properties. On the basis of these considerations, we isolated from the 1GFL the monomer. Then by using the Dowser program [195], we included in the structure only the internal crystallographic waters.

GFP is composed of 238 amino acids, but the last eight C terminal residues are not present in the crystallographic structure. Being these residues far from the chromophore region, we could omit them from the structure we employed for our analysis.

One of the most discussed issues about GFP structure is the protonation state of residues, in particular of histidine, because the number of hydrogen atoms and their position on the ring are extremely influenced by the close environment and the solvent. Previous works adopted different solutions for this problem. For example, Patnaik et al. [77] in their MD simulation considered all histidine, arginine, lysine residues protonated and all glutamic and aspartic acid residues unprotonated, as it should be at pH 7, so that the total charge of GFP monomer was +3. A different structure was prepared by Helms et al. [66] who performed a dynamical simulation in which His77, 81, 139 and 169 are protonated at $N_\epsilon$, all the other at $N_\delta$. They assigned the protonation state by visual inspection and

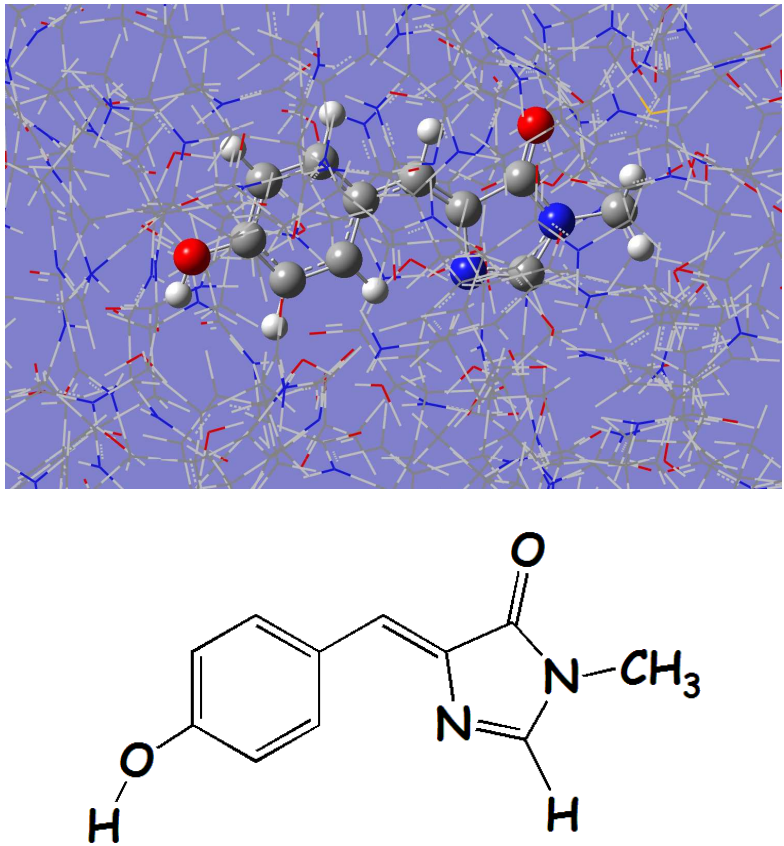
placed six $Na^+$ counterions in energetically favorable positions on the potential energy surface to obtain a neutral system.

In our structure the majority of residues are neutral, while Arg, Lis, Asp and Glu have been treated as charged, in accordance with Patnaik et al. [77]. In addition, the N-terminus of Ala 1 has been protonated and the C terminus of Thr 230 has been unprotonated in order to have the protein in the zwitterionic form. The histidine showed a fine equilibrium between the several protonation states: all histidine residues have been protonated, except for His 148, 169 and 181, closer to the chromophore, which have been considered neutral.

Following the ONIOM scheme [10–13], the GFP has been portioned into two layers, which have been treated at two different levels of theory: the "core", or the QM region, and the "real system", the entire protein, treated with molecular mechanics. In this work we considered six QM/MM partitions of GFP.

In order to preserve the chemical nature of peptide bonds for the residues belonging to the QM/MM interface, the cuts have been chosen along single bonds, and, in general, at nonpolar or slightly polar bonds, preferably C-C ones. The dandling valence have been capped by hydrogen atoms (link atoms).

Figure 2.5: The QM/MM **P1** for the GFP system: the QM region is drawn by ball and stick, while the MM region is drawn as a wireframe (upper panel). A sketch of the QM region is also reported (lower panel).



The first QM/MM partition considered (**P1**, see Figure 2.5) identifies the chromophore unit as the QM region. The QM/MM interface involves two covalent bonds, so the system requires two H-links. The remaining part of GFP monomer is treated at molecular mechanics level.

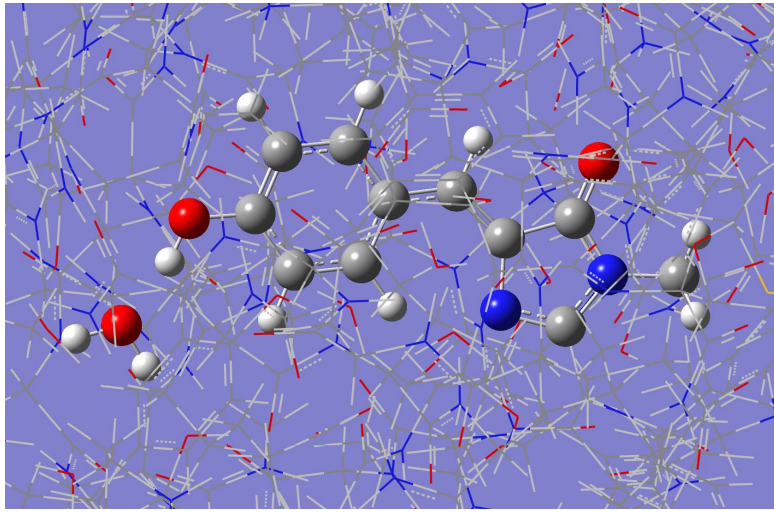
The second QM/MM partition (**P2**, see Figure 2.6) also includes the crystallographic water molecule (water 25 of the 1GFL PDB structure), which is hydrogen bonded to the OH of Tyr66 (chromophore) in the QM region.

In the third QM/MM partition (**P3**, see Figure 2.7) the QM region is further extended to the side chain of Ser205. This inclusion involves the $C_\alpha$ and $C_\beta$ bond across the QM/MM interface, for a total of three link atoms.

The fourth partition (**P4**, see Figure 2.8) includes the side chain of Glu222 in the QM part, beside the previous QM arrangement. A fourth link atom saturates the $C_\beta$-$C_\gamma$ valence of the Glu222.

The fifth partition (**P5**, see Figure 2.9) extends QM part to the side chain of Arg96. This choice has been motivated by the close proximity of this positively charged residue to the chromophore and the possible strong interaction with its carboxyl group. Previous studies suggested the importance of the Arg96 contribution to the spectroscopic behaviour [67, 91] This partition requires five link atoms.

Figure 2.6: The QM/MM **P2** for the GFP system (QM in ball and stick, MM region in wireframe- upper panel). A sketch of the QM region is also reported (lower panel).



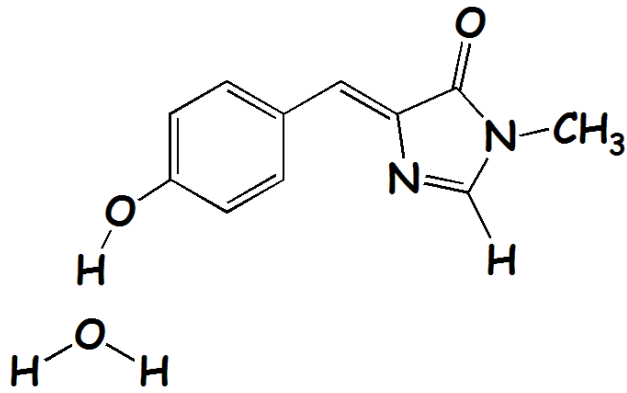


Figure 2.7: The QM/MM **P3** for the GFP system (QM in ball and stick, MM region in wireframe- upper panel). A sketch of the QM region is also reported (lower panel).

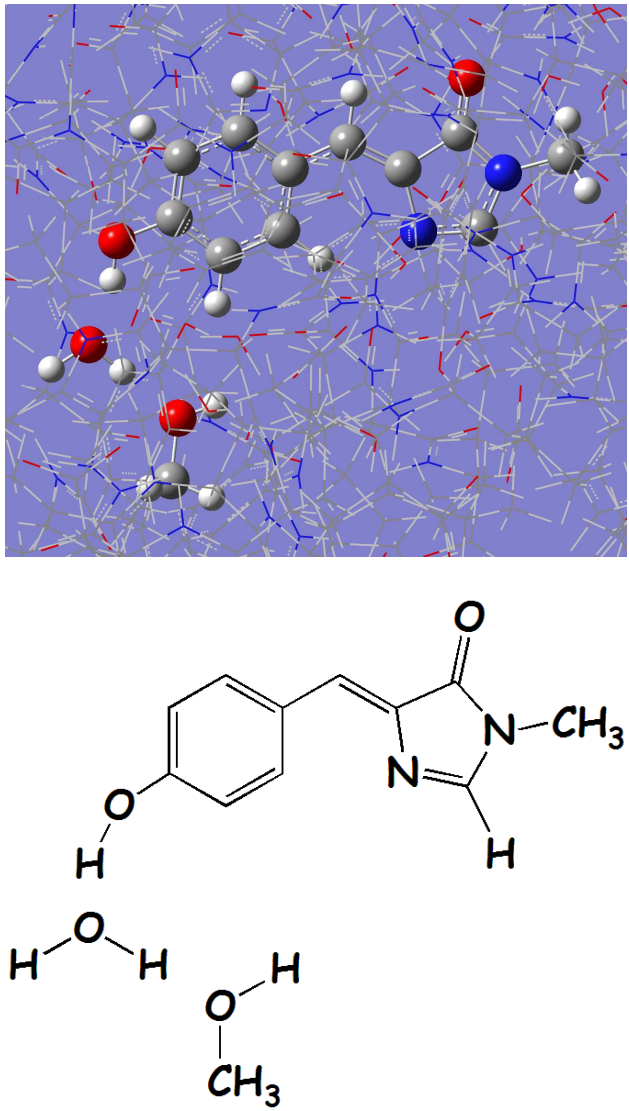


Figure 2.8: The QM/MM **P4** for the GFP system (QM in ball and stick, MM region in wireframe- upper panel ). A sketch of the QM region is also reported (lower panel).

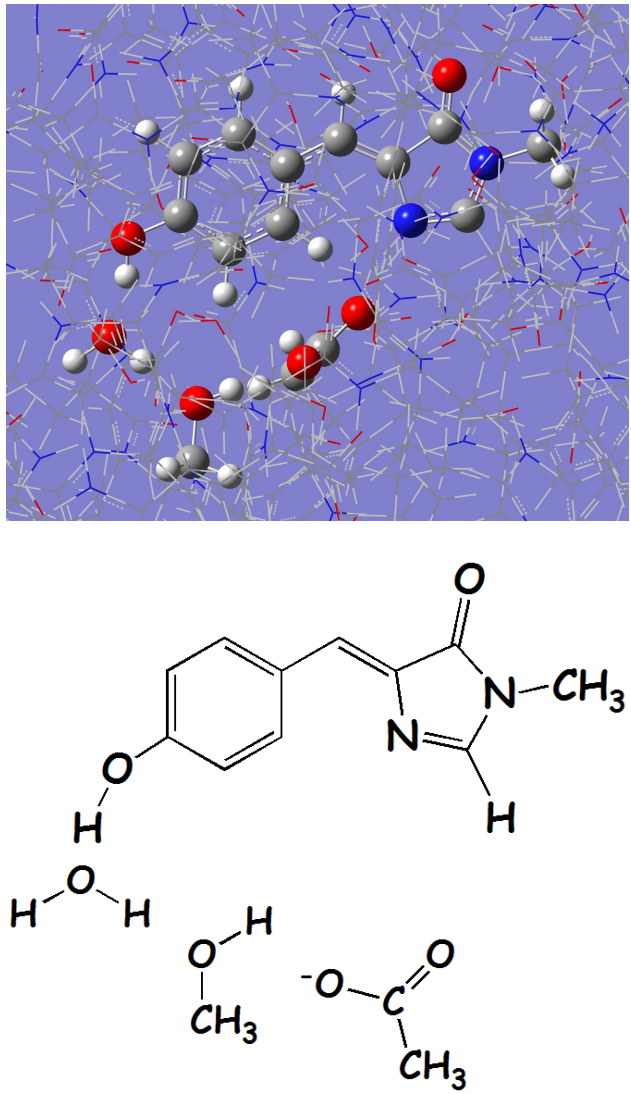


Figure 2.9: The QM/MM **P5** for the GFP system (QM in ball and stick, MM region in wireframe- upper panel). A sketch of the QM region is also reported (lower panel).

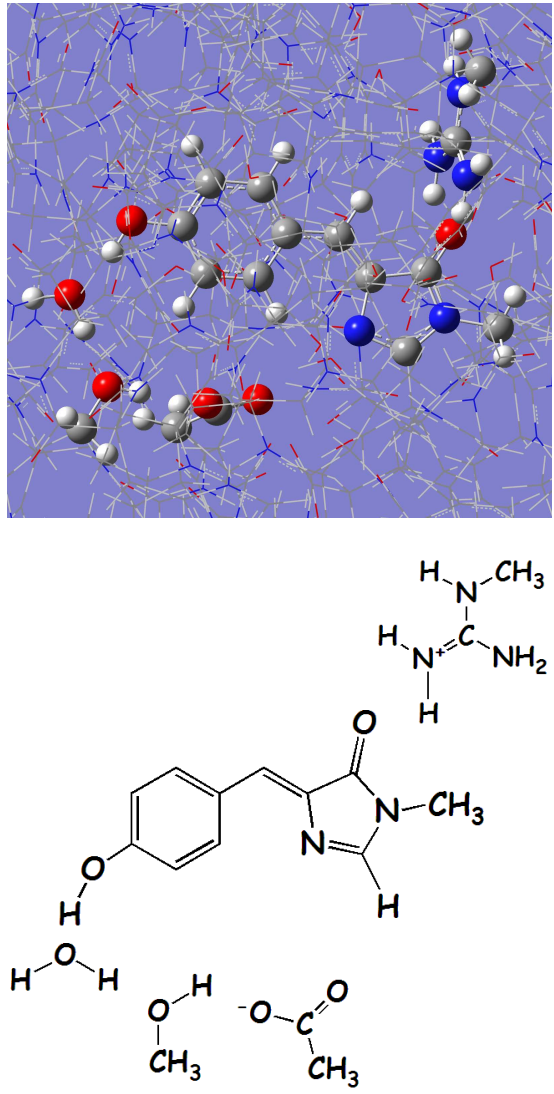
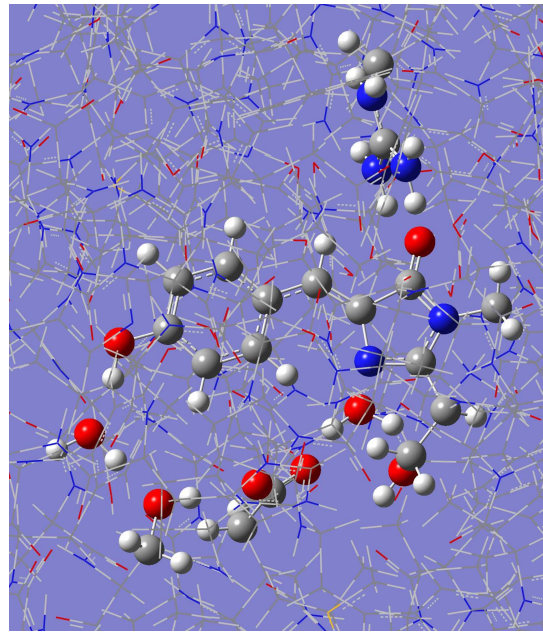


Figure 2.10: The QM/MM **P6** for the GFP system (QM in ball and stick, MM region in wireframe- upper panel). A sketch of the QM region is also reported (lower panel).



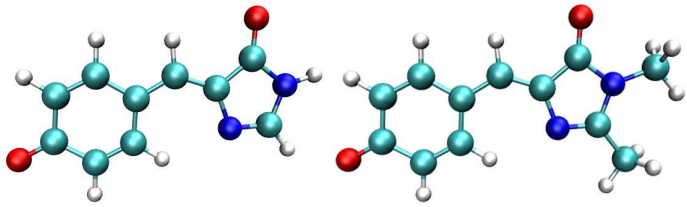


Figure 2.11: Minimal (left), $HBI^-$, and methyl-terminated (right), $HBDI^-$, models.

The last partition (**P6**, see Figure 2.10) further includes the side chain of Ser65 and a second crystallographic water molecule (water 60 of 1GFL PDB structure). In this way, the principal components of the hydrogen bond networking involving Glu222 are accounted for at the QM level.

### 2.2.2 Anionic GFP

#### Anionic chromophore models

We have used for the anionic chromophore two starting models, the $HBI^-$ and the methyl-terminated $HBDI^-$ (see Figure 2.11).
While there are not experimental characterizations for the minimal model in gas phase, for the $HBDI^-$ in vacuum there are two possible references, the photodestruction ex-

periments [94–96], which have placed the absorpion maximum at 2.59 eV [94, 95] and 2.75 eV [96], and the extrapolation from absorption peaks in different solvents [24].

### The anionic GFP models

As recently demostrated in a previous theoretical study[92], the structure of the GFP mutant S65T (1EMA as PDB-ID) [196] belonging to the Class 2 (EGFP - according to Tsien classification of the green fluorescent proteins [112]) is a good experimental benchmark structural GFP model in which the chromophore is in the anionic form. Starting from the 1EMA structure, the coordinates of heavy atoms were augmented by the hydrogen atoms, and the side chain of the residue at the 65 position was set to Ser as in the wild type.

For the residues protonation states, we worked considering the same aminoacids protonation form of the neutral form, except for the histidines: we protonated six histidines in both $\delta$ and $\epsilon$ positions, three in $\epsilon$ and only H148 in $\delta$.

Following the ONIOM scheme, the GFP has been then portioned into two layers (in Figure 2.12 the QM core of the QM/MM partition is shown). The *model* includes the anionic chromophore, the side chain of Ser65, Ser205, Glu222, His148, Asn146, Thr203, Arg96, Gln94, and the

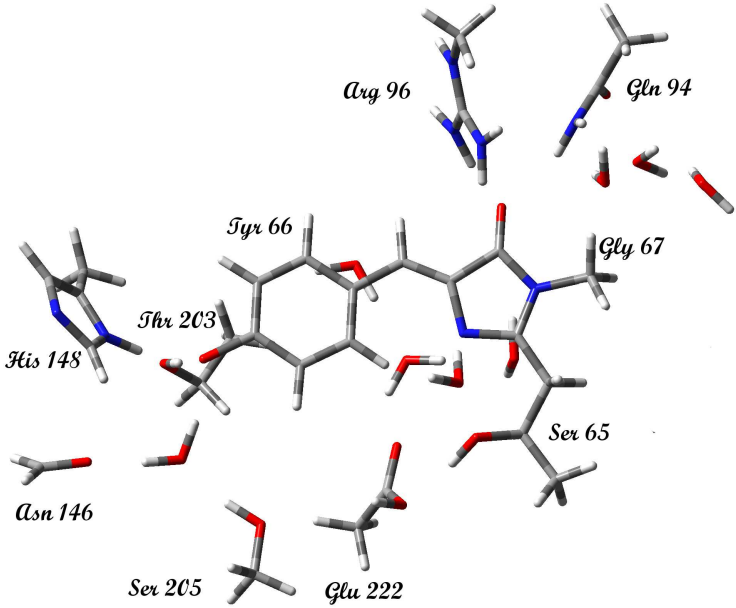



Figure 2.12: QM core of PI anionic GFP QM/MM partition.

crystallographic water molecules which were in proximity of the chromophore. In this way, the hydrogen bond networking involving the proton transfer has been fully accounted at the quantum mechanical level. The *real system*, on the other hand, is the entire protein.

**Anionic HBDI models in several solvents**.

The environment effects on the $HBDI^-$ transition energies have been monitored by analyzing different solvents of increasing polarity and proticity: 1,4-dioxane ($\varepsilon = 2.2099$), ethanol ($\varepsilon = 24.852$), methanol ($\varepsilon = 32.613$) and water ($\varepsilon = 78.3553$). The solvation effects have been monitored by both modeling the solvent with a self consistent reaction field and introducing explicit interactions in solute-solvent clusters. For this last aim, several explicit solvent-$HBDI^-$ clusters have been analyzed regarding all the solvents, except for the 1,4-dioxane, in order to take into account the explicit short-range and hydrogen bond contributions. We have monitored three solvation sites (see Figure 2.13a): the phenol-oxygen (O14 - Site 1), the imidazole-oxygen (O6 - Site 2) and the imidazole-nitrogen (N3 - Site 3).

The following nomenclature has been employed to identify the clusters: HBDI_*n1n2n3s*, where *s* is the solvent

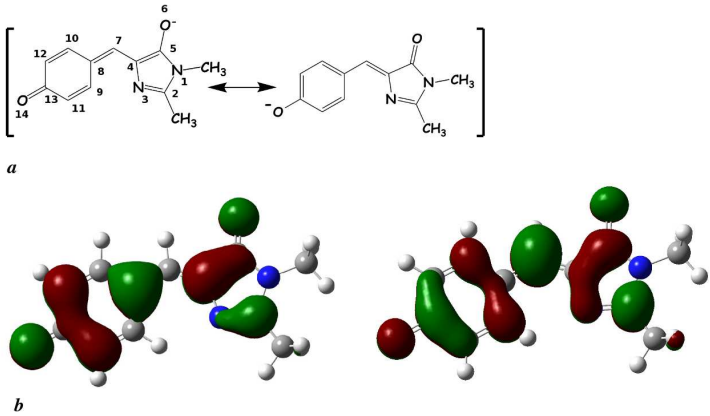



Figure 2.13: Resonance structures (a) and KS HOMO and LUMO countours (b) for $HBDI^-$.

type and *n1*, *n2* and *n3* are the number of explicit solvent molecules on sites 1, 2 and 3, respectively. $HBDI^-$ shows a complex topology of the solvation sites. The hydrogen bonds network retrieved from anionic GFP PDB structures [196–198] and QM/MM optimizations retraces the 321 arrangement, i.e. the phenol oxygen, the imidazole oxygen and nitrogen involved with residues and a crystallographic water in three, two and one hydrogen bonds, respectively. Three as coordination number is compatible with the negative charge partially located on the phenolic group, the $\pi$ electronic density on the oxygen acting as hydrogen bond acceptor. Mulliken charges calculated in the

gas phase for phenolic and imidalzolic oxygens show similar values (-0.627693 and -0.644698 a.u. calculated at the B3LYP/6-31+G(d,p) level), indicating a similar weight of the two represantive resonance structures of the molecule. In the isotropic solvent we can therefore expect also the imidazole oxygen possibly involved in up to three specific interactions. Moreover, the imidazole nitrogen (-0.282569 mulliken charge in the isolated molecule) can be engaged in specific interactions directed both above and below the molecular plane. Radial distribution functions calculated by MonteCarlo simulations of the trans $HBDI^-$ in water show indeed a coordination number of about three for the phenolic oxygen and of about two for the nitrogen (see Figure 2.14).

In summary, we can hypothesize a 332 arrangement to fully saturate the solvation sites of the molecule. We considered several 3n2n3 clusters with n2=2,3 and n3=1,2 for ethanol, methanol and water (the main are reported in rows 5,6 and 7 of table 2.1). In all cases, we found minima and calculated the chromophore solvation free energy, $\Delta G_{solv}$, and the solute/solvent interaction energy, $E_{int}$. Energetical parameters show increasing values of both $\Delta G_{solv}$ and $E_{int}$ when going from the 321 to the 332 arrangements for all the solvents. By nature of the continuum model, these values do not account for the entropic

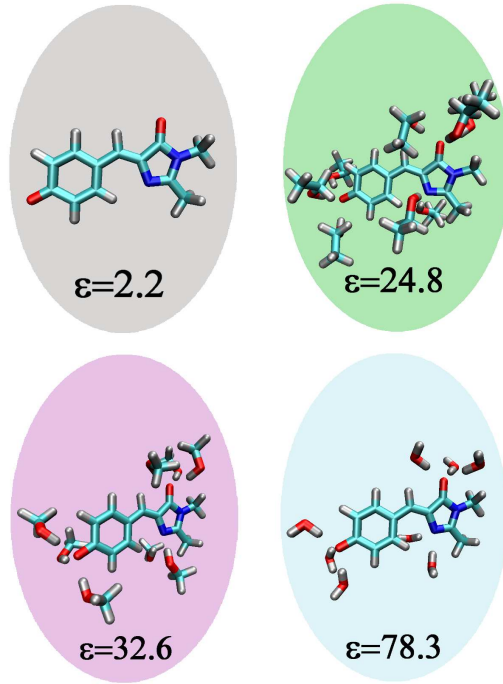



Figure 2.14: $HBDI^-$ in several solvents: 1,4-dioxane ($\varepsilon = 2.2$), ethanol ($\varepsilon = 24.8$), methanol ($\varepsilon = 32.6$) and water ($\varepsilon = 78.3$)

| System | O14(P) | O6(I) | N3 | solvent | reference name |
|---|---|---|---|---|---|
| $HBDI^-$ | - | - | - | dioxane | HBDI_d |
| $HBDI^-$ | - | - | - | ethanol | HBDI_e |
| $HBDI^-$ | - | - | - | methanol | HBDI_m |
| $HBDI^-$ | - | - | - | water | HBDI_w |
| $HBDI(CH_3CH_2OH)_8^-$ | 3 | 3 | 2 | ethanol | HBDI_332e |
| $HBDI(CH_3OH)_8^-$ | 3 | 3 | 2 | methanol | HBDI_332m |
| $HBDI(H_2O)_8^-$ | 3 | 3 | 2 | water | HBDI_332w |

Table 2.1: Summary of the main $HBDI^-$/solvent models (column 1). The systems in rows 1-4 are characterized by implicit solvation, while systems in rows 5-7 used explicit models. For each system the number of explicit solvent molecules on different solvation sites (columns 2,3 and 4), the solvent (column 5) and abbreviation name (column 6) are reported.

contribution in the cluster arrangement and may sensibly approximate the energetic balance of the solute/solvent and the solvent/solvent specific interactions. As matter of fact, an accurate prediction of the specific interactions distribution would require proper simulations with an explicit representation of the solvent. On the other hand, it is important to recall here that the modeling of the liquid state by a representative cluster should take into account, on average, the effect that the cybotactic region has on both the chromophore structure and the absorption spectrum. In the following, we discuss total solvation drifts of the electronic transitions considering the 331 and the full solvated 332 minima as reference. Individual effects on each solvation site are then separately analyzed.

The total solvent shift on the $S_1 \leftarrow S_0$ transition energy, $\Delta G^{01}_{tot}$, is obtained as the difference between the excitation energy in solvent, $G^{01}_{solv}$, and in the gas phase, $E^{01}_{vac}$.

$$\Delta G^{01}_{tot} = G^{01}_{solv}(R_{solv}) - E^{01}_{vac}(R_{vac}) \tag{2.14}$$

where $R_{vac/solv}$ collects the structural coordinates of the chromophore optimized in gas phase and the solvent, respectively. The suffix *vac* in the energy symbol indicates an excitation energies calculated from the electronic density of the unperturbed chromophore; the suffix *solv*, on the other hand, refers to excitation energies obtained from the chromphore electronic density which is polarized by both the explicit solvent molecules and the bulk solvent.

The total solvent shift can be decomposed in several contributes:

$$\Delta G^{01}_{tot} = \Delta E^{01}_{ind} + \Delta E^{01}_{pol/int} + \Delta G^{01} \tag{2.15}$$

where the first term, $\Delta\ E^{01}_{ind}$, accounts for the indirect effect due to chromophore structural change induced by the environment, and can be evaluated as the difference of the gas phase excitation energies evaluated at $R_{solv}$ and $R_{vac}$, respectively

$$\Delta E^{01}_{ind} = E^{01}_{vac}(R_{solv}) - E^{01}_{vac}(R_{vac}) \tag{2.16}$$

The second term in Eq. 2.15, $\Delta E^{01}_{pol/int}$, is the transition energy shift due to the electronic density polarization in presence of both the explicit solvent molecules and the bulk, and to the interaction between the solute and the explicit solvent molecules. It can be evaluated as

$$\Delta E^{01}_{pol/int} = E^{01}_{solv}(R_{solv}) - E^{01}_{vac}(R_{solv}) \tag{2.17}$$

When solute-solvent clusters are considered, it is difficult to separate contributions from the polarization and the excplicit interaction energy. We can instead evaluate the contribution due to the explicit solvent molecules by considering the excitation energies calculated from the electronic density of the chromophore perturbed by the explicit solvent molecules, $E^{01}_{cluster}$, given by

$$\Delta E^{01}_{pol/int,cluster} = E^{01}_{cluster}(R_{solv}) - E^{01}_{vac}(R_{solv}) \tag{2.18}$$

Finally, $\Delta$G$^{01}$ is $S_0 \leftarrow S_1$ variation in the solvation free energy and can be evaluated as

$$\Delta G^{01} = G^{01}_{solv}(R_{solv}) - E^{01}_{solv}(R_{solv}) \tag{2.19}$$

This latter contribution is intended in the so-called non equilibrium regime, according to which fast coordinates of the solvent instantaneously adjust to the electronic excita-

tion, while slow coordinates remain equilibrated to solute electrostatic potential in the ground state.

### 2.2.3 Proton Transfer

#### Approximated simple models

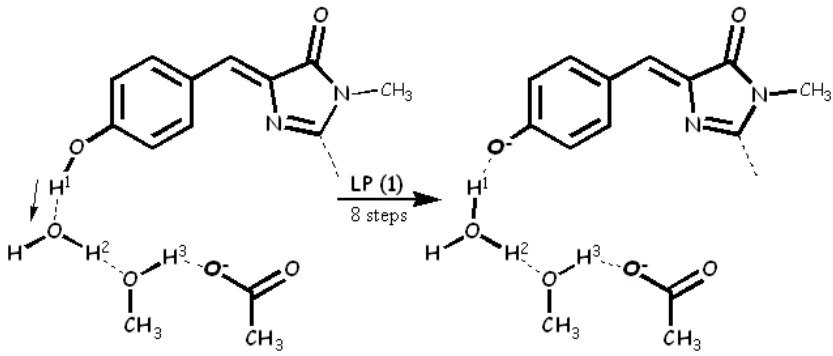
Regarding our theoretical study of the proton transfer (PT), our attention has been mainly captured by two mechanisms, which can possibly drive the shuttle of the proton from the chromophore to Glu222. The first is a $step-wise$ Grotthus-like *mechanism*, by which the proton is transfered from the chromophore to Glu222 through three successive donor/acceptor steps, involving a water molecule and Ser205. The second is a *concerted* mechanism, in which a simultaneous shift involves three hydrogen atoms.
We considered the GFP neutral structure represented by QM/MM P6 (see Figure 2.10). In this way not only the residues directly involved in the PT, but all the contributors to the hydrogen bond network close to the chromophore are treated at QM level. We optimized the whole structure in the neutral form at the B3LYP/6-31G/Amber level. This optimization mainly regarded the QM part, while the MM surrounding was loosely relaxed on the QM core. In this way, the scaffold of the $\beta$-can has not been

artificially modified.

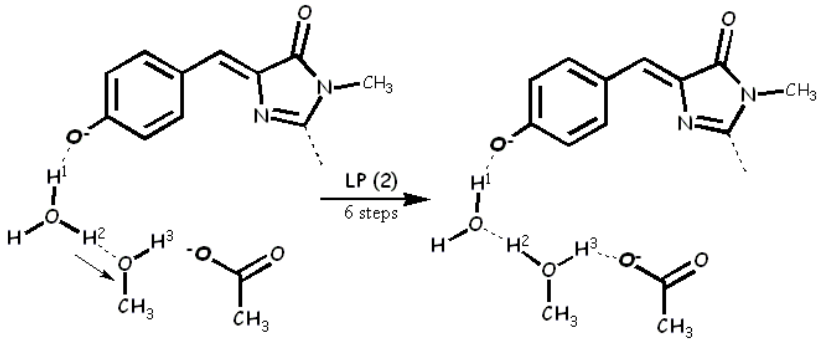
When compared to the x-ray results, the optimized structure shows a slight modification of the chromophore and of the hydrogen bond network. More in detail, the system is reoriented to enhance the whole hydrogen bond network around the chromophore.

Starting from this stucture, we built up two models for the reaction coordinate of the two mechanisms (step-wise and concerted). Regarding the step-wise proton shuttle, we considered the three oxygen-hydrogen bonds involved. Along such distance, the reaction coordinate refers to the structure of GFP with a progressive increasing of the donor-hydrogen distance and, as a consequence, a progressive decreasing of the acceptor-hydrogen distance. From the optimized neutral structure to the final anionic structure we individuated 20 values for the reaction coordinate, equally spatiated. A simple scheme for the linear step-wise (SWLP) reaction coordinate is sketched in Figures 2.15-2.15(c).

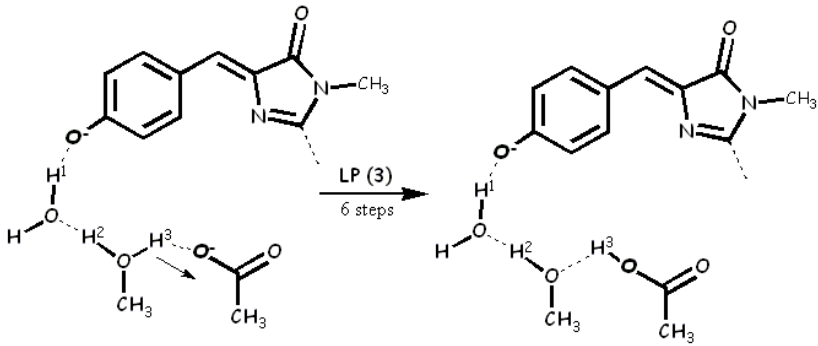
Regarding the concerted mechanism, we considered a linear synchronous path (LSP) along the free oxygen-hydrogen distances. In this case, the three protons move simultaneously along the reaction coordinate, represented by eight values. A scheme for the concerted LSP coordinate is reported in Figure 2.16.




(a) ***step one: $H^1$ from Tyr66 to water.***




(b) ***step two: $H^2$ from water to Ser205.***




(c) ***step three: $H^3$ from Ser205 to Glu222.***

Figure 2.15: Scheme of the SWLP.

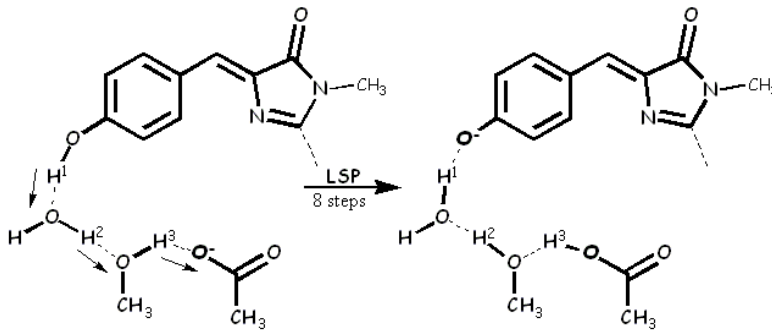



Figure 2.16: Scheme of the LSP.

### GFP structures models

To make a critical analysis of the PT mechanism the first step was to obtain accurate models for the entire GFP. We have worked on four structures: the neutral form A and the anionic intermediate I in both ground and the excited state. The reference PDB structure for the construction of both the neutral and the anionic models has been the 1GFL [61] (x-ray structure for the neutral GFP). 1GFL structure has been modified to simulate the anionic GFP I form: all the protons involved into the PT have been moved from their donors to their acceptors, without altering anything else of the remaining environment.

Starting from these reference structures, we have carried out optimizations for both the ground and the excited

state. For all the optimizations the solvent chosen has been water. We have first considered GFP QM/MM partitions with a bigger QM core, which includes the major residues interacting with the chromophore (the QM core of the partition is represented in Figure 2.17) and optimized them through B3LYP/6-31G(d,p)/Amber/CPCM and TD-B3LYP/6-31G(d,p)/Amber/CPCM potential for the ground and the excited state respectively [131, 132].

Then we have built for both A and I forms new GFP QM/MM partitions with a smaller core, which includes only the residues directly involved into the PT reaction and optimized them through B3LYP/6-31+G(d,p)/Amber/PCM TD-B3LYP/6-31+G(d,p)/Amber/PCM potential for the ground and the excited state respectively. (see the QM core of the partition in Figure 2.18).

We have employed these two approaches to find a compromise between accuracy and computational cost. In fact, the optimization of the bigger QM/MM partition (Figure 2.17) through TD-B3LYP/6-31+G(d,p)/CPCM potential would have been too expensive, so we allowed a first relaxation of all the residues in close contact with the chromophore and then a more specific accommodation of the residues directly involved into the transfer, which have a major influence on our study.

Figure 2.17: QM core of the QM/MM biggest GFP partition for both A (up panel) and I (down panel ).

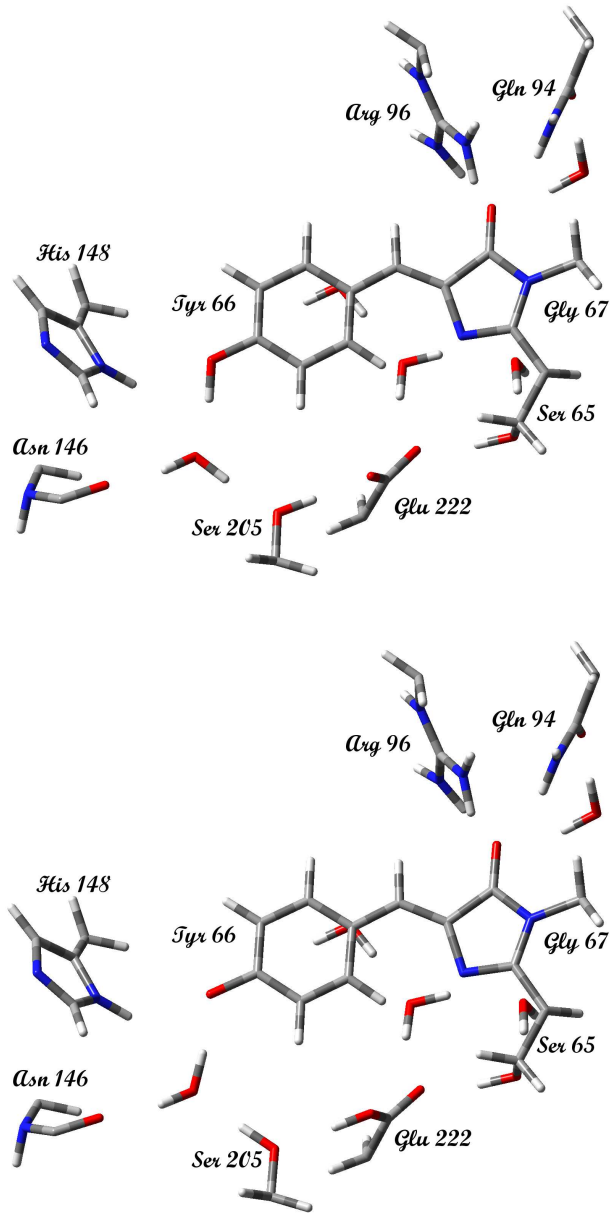
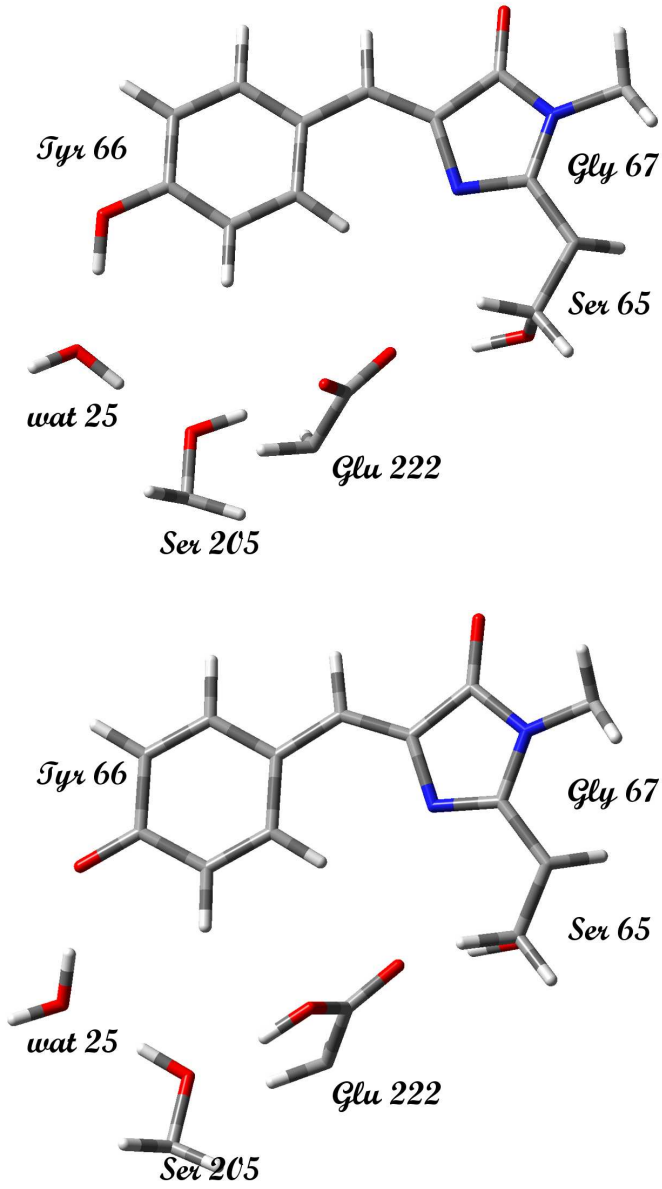


Figure 2.18: QM core of the QM/MM minimal GFP partition for both A (up panel) and I (down panel ).



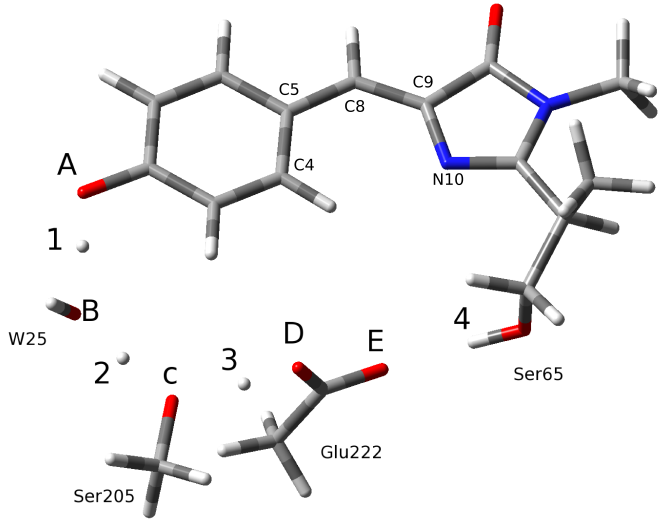



Figure 2.19: PT model compound A:chromophore, water25, Ser205, Glu222 and Ser65.

### IRC models

The search of IRC requests an optimization procedure to locate the transition state (TS). Starting configurations for TS have been extracted from GFP structures optimized at the (TD)DFT/ONIOM level [131, 132] as explained before.

Regarding the TS of the PES in the ground state, $S_0$, a minimal molecular cluster model, hereafter referred to

as A (See Fig. 2.19), has been extracted. This includes the chromophore, water25, and the residues Ser65, Ser205 (modelled by a methanol molecule) and Glu222 (modelled by an acetate ion). In this reference cluster model has been also included the effect of the Ser65, that is not directly involved in the proton transfer. The majority of the results obtained in this work are referred to model A. We considered two others models (Figure 2.20) for single point energy calculations. The first, hereafter referred to as B, can be obtained from A by adding the residue His148 in the protein like arrangement. The second one, hereafter referred to as C, can be obtained from A by eliminating the Ser65.

The TS optimization was performed with no constraints. In order to check the nature of the stationary point and the effective reaction vector involving the proton transfer, all force constants and the resulting vibrational frequencies have been evaluated. In particular an imaginary frequency of -276.6 $cm^{-1}$ was found, corresponding to a collective stretching of all protons of the wire, as shown in Fig. 2.21. No chromophore twisting or ring modes are involved in this coordinate.

Regarding the PES of the first excited state, $S_1$, we considered the same model A presented above. This time a free optimization procedure could not be possible, be-

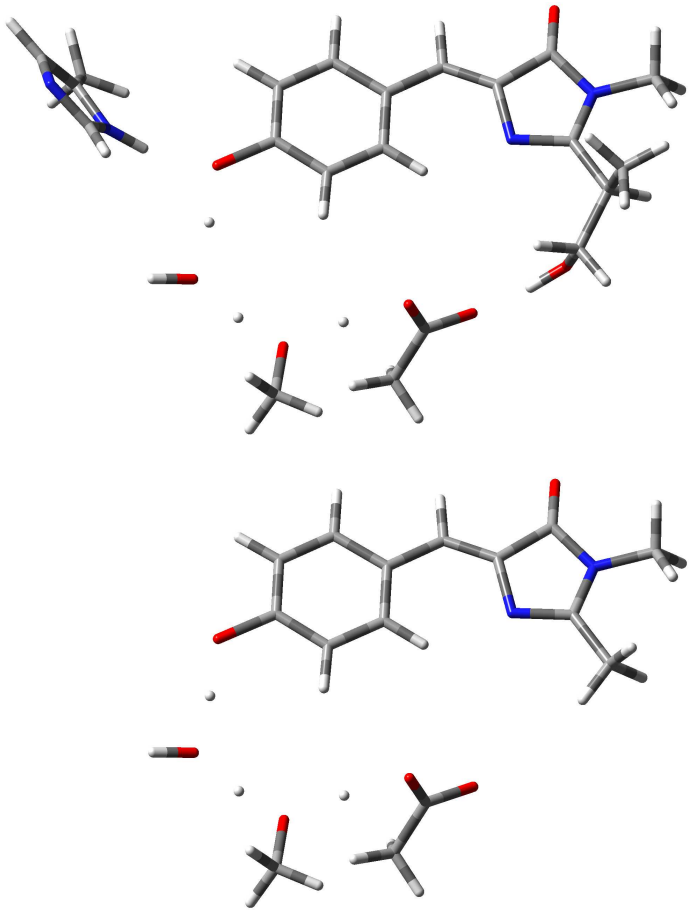
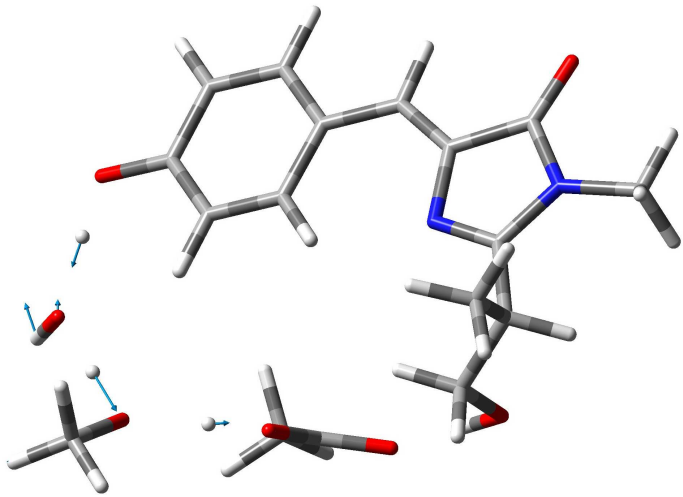


Figure 2.20: PT model compounds B (upper panel) and C (lower panel). B model consists of the chromophore, water25, Ser205, Glu222, Ser65 and His148; C model includes the chromophore, water25, Ser205 and Glu222.

Figure 2.21: TS for the proton transfer reaction in the electronic ground state ($S_0$). The light-blue vectors show the collective stretching of the protons of the wire. Relative frequency are obtained at B3LYP/6-31+G(d,p) level.



cause a full relaxing process of the TS leads to a twisting HBDI motion and the interconversion to a dark state. Therefore, a partial TS free optimisation was run and then re-started with the following constraints: all model atom were frozen, except the oxygen and hydrogen atoms directly involved in the proton transfer and the Ser65 side chain atoms. In Fig. 2.22 the fixed atoms are red coloured, while the unconstrained ones are in green. On this final structure all force constants and the resulting vibrational frequencies have been evaluated. In particular an imaginary frequency of -619.04 $cm^{-1}$ was found, corresponding to a collective stretching of all protons of the wire, as shown in Fig. 2.23. Starting from the TS geometries obtained, the intrinsic reaction coordinate have been integrated in both directions for both the electronic states. The two different structures corresponding to the reactant state, namely R in which the HBDI chromophore is neutral, and the product state, namely P in which the chromophore is present as a phenolate anion, have been after optimised in order to locate the corresponding local PES minima. It is important to underline that both the $S_0$ and $S_1$ IRC calculations have been performed without any constraints, following only the respective reaction vectors presented above.

Figure 2.22: $S_1$ schematic proton transfer constrained scheme for the model compound A. The atom in red are completely frozen during the optimization, while the green one are full free.

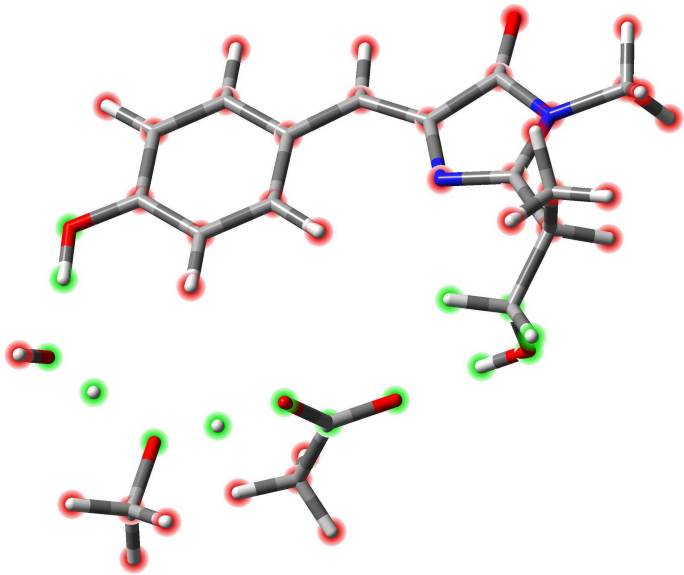
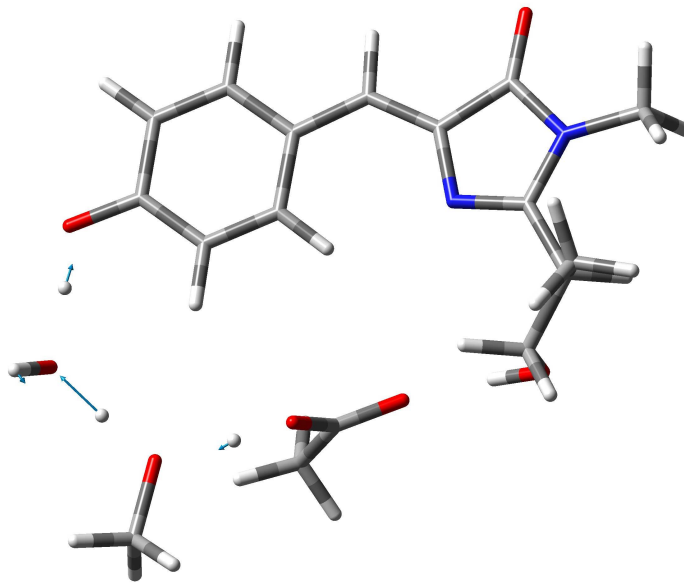


Figure 2.23: TS for the proton transfer reaction in the electronic excited state ($S_1$). The light-blue vectors show the collective stretching of the protons of the wire. Relative frequency are obtained at B3LYP/6-31+G(d,p) level.



### 2.2.4 Electronic level of theory

Validation tests and energy calculations have been performed on the neutral form of GFP using the following density functionals: M052X [151], CAM-B3LYP [148], B3LYP [199], LC-wPBE [149, 200, 201], PBE0 [153].

- M052X [151]: this functional is a long range corrected version of M06, the hybrid functional developed by Truhlar and Zhao [150].

- B3LYP: this is the hybrid functional developed by Becke. It employs the Becke exchange term and the non-local correlation provided by the LYP expression [199].

- CAM-B3LYP: This is the Handy and coworkers long range corrected version of B3LYP using the Coulomb-attenuating method [148].

- LC-wPBE is the long range-corrected version of wPBE [149, 200, 201].

- PBE0: PBE the 1996 pure functional of Perdew, Burke and Ernzerhof [202, 203], has been developed into a hybrid form by Adamo [153], resulting in PBE1PBE. This functional uses 25% exchange

and 75% correlation weighting and is known in the literature as PBE0.

These density functionals have been employed for both ONIOM/DFT and ONIOM/TD-DFT [131, 132] energy calculations, along with the 6-31+G(d,p) and the 6-31++G (2d,2p) basis sets [204–208]. Structure optimization by ONIOM [209] has also been considered using the B3LYP/ 6-31+G (d,p)/Amber level of theory. The structures of the neutral chromophores were relaxed in the ground state using DFT and a B3LYP/6-31G(d,p) potential. For the optimizations 6-31G(d,p) basis set was chosen as the use of a diffuse function in the 6-31+G(d,p) basis set does not change the models geometry significantly and, as a consequence, the values of the absorption peaks (see Table 2.2).

Following the ONIOM scheme [10–13], both neutral and anionic GFP have been portioned into two layers: the QM region, has been analyzed at DFT [1–3] and TD-DFT [4–7] level and the entire protein has been treated with molecular mechanics. In particular, the protein has been described through the Amber force field of Kollman et al. [210] and the water model used is TIP3P [211]. As the chromophore does not correspond to the standard structure and chemical properties performed by serine, tyrosine and glycine, there is a lack of specific Amber force

| vacuum | | water | |
|---|---|---|---|
| opt 6-31G(d,p) | opt 6-31+G(d,p) | opt 6-31G(d,p) | opt 6-31+G(d,p) |
| 3.45 | 3.45 | 3.30 | 3.29 |
| f= 0.6893 | f=0.6881 | f= 0.8088 | f= 0.8112 |

Table 2.2: Excitation energies eV and oscillator strengths for neutral chromophore in vacuum and in water obtained through TD-DFT B3LYP/6-31+G(d,p) potential on geometries optimized with 6-31G(d,p) or 6-31+G(d,p) basis set, respectively.

field parameters for the molecular-mechanics calculations. We included them with reference to the empirical force field developed for the neutral and the anionic GFP chromophores by Reuter et al. [127]. The electrostatic interaction between the two layers were accounted for by an electronic embedding scheme [11, 209]. In this case the *model* also includes molecular mechanics partial charges of the embedding protein. Therefore, the Kohn Sham operator exploited in the DFT calculation also includes interactions between the electronic density and surrounding charges, allowing for the electrostatic interactions to be considered at the *high level* of theory. The ONIOM/PCM calculations were performed according to different schemes. Structure optimizations were performed at the B3LYP/6-

31G(d,p)/Amber/PCM level for the ground state and at the TD-B3LYP/6-31G(d,p)/Amber/PCM level for the excited state, adopting the so-called ONIOM/PCM-C scheme. In this case the reaction field is included only at the low computational level for the real system, leading to the free energy functional $G^{MM/PCM,real}$, while model system calculations are carried out in the gas phase, giving as result internal energies $E^{DFT,model}$ and $E^{MM,model}$. These quantities are composed in the ONIOM free energy $G^{ONIOM/PCM-C}$ according to

$$G^{ONIOM/PCM-C} = E^{DFT,model} + G^{MM/PCM,real} - E^{MM,model} \tag{2.20}$$

VEE values were calculated exploting the so-called ONIOM/PCM-A formalism. In this case the solvent reaction field is computed separately in *high level* sub-calculation, leading to in the ONIOM free energy

$$G^{ONIOM/PCM-A} = G^{DFT/PCM,model} + E^{MM,real} - E^{MM/PCM,model} \tag{2.21}$$

For the anionic chromophore all the calculations were carried out by using the B3LYP [152] functional and the 6-31+G(d,p) basis set, in both DFT and TD-DFT formalism, this latter to model the vertical $S_1 \leftarrow S_0$ transi-

tion energy. Implicit solvent effects were taken into account by the polarizable contiuum model (CPCM) [128–130], by adopting the State Specific (SS) formalism [188–191, 212] when combined with TD-DFT calculations. Optimizations for solute-solvent clusters also included the Grimme's dispersion contributions in the electronic potential [213].

The calculations involved into the optimization procedure for the location of the TS in both the ground and the excited states of the PT reaction and the IRC calculations were carried out by using the B3LYP [152] density functional and the 6-31+G(d,p) basis set, in both the time independent and TD-DFT formalism [4–7], this latter to model the vertical $S_1 \leftarrow S_0$ transition energy.

The Gaussian [183] suite of programs has been used for all calculations.

# 3 Results and discussion

*The real act of discovery consists*
*not in seeking new landscapes,*
*but in having new eyes.*
*(M. Proust)*

In this chapter we report the results of our theoretical study on GFP in neutral and anionic forms. In order to elucidate some of the features affecting GFP photochemistry, we analysed in detail the role of the protein environment on both GFP forms absorption. Furthermore, we investigated the excited-state proton shuttle mechanism, which is responsible for the switching between the neutral and the anionic forms [120, 123–126].

## 3.1 Neutral GFP

As a preliminary study, we performed a robust methodological work in order to assess and validate a computa-

tional protocol for an accurate and quantitative description of the absorption phenomenon. This analysis mainly lead to the best choice of the QM/MM potential in terms of density functional and basis set: this potential has been there adopted for the remaining part of the study. However, for the sake of simplicity, in this discussion we postponed the description of the validation work, essentially of technical and methodological nature, to the end of the chapter (Section 3.1.3 ).

### 3.1.1 The isolated chromophore

Since we have defined the potential to thoroughly describe GFP system, we found useful to verify our method on the isolated chromophore. We chose as model (see Figure 2.4, lower panel) the one presented by Filippi et al.[108], because we had the use of the experimental data collected by Dong et al.[24] on the same chromophore in different solvents and of the extrapolation to the vacuum dielectric constant through Kamlet-Taft fit. In fact, for the neutral chromophore in vacuum no experimental data are reported, except the photodestruction one [26], which is in disagreement with both previous theoretical results [108] and does not offer a realistic estimation.

As it is possible to see in Table 3.1 our theoretical absorption value is really close to the one obtained by Kamlet-

| TD-DFT/B3LYP 6-31+G(d,p) | TD-DFT/B3LYP [108] cc-pVTZ | TD-DFT/SAOP [108] | CASPT2 [108] | *Kamlet-Taft [24]* |
|---|---|---|---|---|
| **3.45** | 3.46 | 3.20 | 3.58 | 3.51 |

Table 3.1: Excitation energies (eV) in vacuum for neutral chromophore. The value that we have obtained is compared with the results gained in vacuum by Filippi et al.[108] using different potentials. The extrapolation to the vacuum dielectric constant through Kamlet-Taft fit [24], taken as reference for the neutral form in vacuum, is also reported. In bold is expressed the more similar value to Kamlet-Taft fit.

Taft fit by Dong et al.[24]. In fact, they have predicted an absorption peak at 3.51 eV and we have an absorption of 3.45 eV, so $\Delta$E=0.06 eV. In addition, our result is in complete agreement with Filippi et al.[108] (TD-DFT/B3LYP result for the neutral chromophore) and the comparison with other methods, usually adopted to describe little systems, like TD/SAOP and CASPT2, points out that TD-DFT/B3LYP gives equal or better results, as is shown in Table 3.1 ($\Delta E_{SAOP}$=0.31 eV, $\Delta E_{CASPT2}$=0.07 eV). As we have taken as reference both theoretical and experimental works, it is possible to conclude that our treatment of the GFP chromophore is accurate and realistic.

### 3.1.2 The environment contribution

In order to elucidate some of the features affecting GFP photochemistry, we have analysed in detail the role of the protein on neutral chromophore absorption.

An important tool for our approach to the study of the GFP absorption has been the comparison among the performance of several QM/MM partitions of the GFP entire structure.

In the first partition (see Figure 2.5) the chromophore has been treated at the quantum mechanical (QM) level, while the remaining part of the GFP has been considered at molecular mechanical (MM) level. In the other partitions (from P2 to P6, see Figures 2.6- 2.10) the QM region has been further extended to the chromophore close environment.

In this way we have been able to switch on and off the explicit treatment of the electronic structure of some residues and two cristallographic water molecules in the calculation of the spectroscopic parameters (absorption energy and intensity). Therefore, we have been able to distinguish among the several features affecting the absorption and, in particular, the influence of the different residues on the chromophore electronic states in terms of pure electrostatic interactions and quantum mechanical effects (for

example, the charge transfer to the orbitals involved in the electronic transition).

The other fundamental characteristic of our methodological approach, which makes it a substantial tool of investigation, is the accuracy pursued in the results: the quantitative agreement with the experiment allows us to obtain a reliable analysis, when dissecting the different features influencing the GFP photochemistry.

We have divided the contribution of the environment into three effects: structural, electronic and quantistic.

**Structural effect**

As we tried to split up the contribution of the environment, we analyzed the absorption of the chromophore (P1 core) in both the optimized and the x-ray form.

Surprisingly, we obtained two absorption peaks which are rather different as it is shown in table 3.2. Main geometrical differences consist in the deviation from the planarity of $O_6C_5C_4C_7$, $C_2N_3C_4C_5$, $N_3C_4C_5N_1$ and $C_4C_5N_1C_2$ dihedrals, so the principal distortion is on the imidazolinone ring (see Table 3.3).

For the B3LYP/6-31G(d,p) optimized structure we reached an absorption of 3.44 eV, while for the x-ray structure the peak is at 3.33 eV, so the difference is of 0.10 eV.

| | OPTIMIZED | X-RAY |
|---|---|---|
| $\Delta$E | 3.45 | 3.38 |
| f | 0.6893 | 0.6709 |
| $\%_{H-L}$ | 0.70083 | 0.70342 |

Table 3.2: Excitation energies ($\Delta$E) expressed in eV, oscillator strengths (f) and percentage of homo-lumo transition ($\%_{H-L}$) are reported for the chromophore in vacuum. The first structure is optimized through B3LYP/6-31G(d,p) potential. The x-ray structure is obtained by 1GFL [61], isolating only the chromophore. The absorption peaks have been obtained using TD-DFT with B3LYP/6-31+G(d,p) potential.

| GEOMETRIC PARAMETERS | OPTIMIZED | X-RAY |
|---|---|---|
| $O_6C_5C_4C_7$ | **0.001** | **-5.299** |
| $C_5N_1C_2N_3$ | -0.014 | -1.587 |
| $N_1C_2N_3C_4$ | 0.007 | -1.222 |
| $C_2N_3C_4C_5$ | **0.002** | **3.232** |
| $N_3C_4C_5N_1$ | **-0.011** | **-4.293** |
| $C_4C_5N_1C_2$ | **0.014** | **3.501** |
| $C_5C_4C_7C_8$ | -179.998 | -178.146 |
| $N_3C_4C_7C_8$ | -0.002 | 0.532 |
| $C_4C_7C_8C_9$ | 0.002 | 0.045 |
| $C_4C_7C_8C_{10}$ | -179.999 | 179.764 |
| $C_9C_{11}C_{13}O_{14}$ | -179.993 | 178.136 |
| $C_4C_7C_8$ | 130.054 | 129.324 |
| $C_5C_4C_7$ | 122.145 | 121.205 |
| $C_{11}C_{13}O_{14}$ | 122.604 | 121.574 |
| $N_1C_2N_3$ | 115.446 | 116.152 |
| $C_5O_6$ | 1.221 | 1.238 |
| $C_4C_7$ | 1.358 | 1.380 |
| $N_3C_2$ | 1.297 | 1.365 |
| $C_{13}O_{14}$ | 1.361 | 1.371 |
| $O_{14}H_{15}$ | 0.967 | 0.969 |
| $C_5C_4$ | 1.489 | 1.503 |
| $C_7C_8$ | 1.448 | 1.412 |

Table 3.3: Main geometrical parameters referred to the chromophore optimized with B3LYP/6-31G(d,p) potential and x-ray structure. The atom labeling is referred to Figure 2.4

.

Starting from these data we assumed that the presence of the $\beta$-can has a first influence on the chromophore that is a structural effect. In fact, the geometry (eg. bond distances, angles, dihedrals) imposed by the environment determines by itself a batochromic shift in the absorption of about 0.10 eV, nearly a third of the entire contribution of the protein.

**Electrostatic and quantistic effect**

The neutral wt-GFP shows a neat band of absorption peaked at about 3.11 eV (398 nm) [41, 70]. In order to present a definitive and detailed analysis of the individual role of different residues on neutral GFP absorption, we considered our six QM/MM partitions, starting from P1 in which the core is the chromophore and increasing the QM part from one partition to another, including the residues which are closer to the chromophore. We performed TD-DFT calculations on all the partitions in order to characterize the ground and the excited electronic states, involved in the $\pi$-$\pi^*$ transition responsible for the absorption. In particular, we focused on the excitation energy, the corresponding oscillator strength and the nature of the electronic orbitals mainly involved in the transition. Regarding this latter aspect, we also analyzed the nature

of the transition, which can be pure $\pi$-$\pi^*$ or mixed, i.e. can also involve excitations different from the $\pi$ orbital of the chromophore.

In order to better evaluate the environmental influence, we also performed analogous calculations on isolated systems, corresponding to the QM region of GFP P1-P6. These *core* systems required pure QM calculations at the B3LYP/6-31+G(d,p) level.

In table 3.4 are reported the absorption peaks (eV), the oscillator strengths and the homo-lumo contribution to each transition calculated for the GFP six QM/MM partitions and the corresponding *core* systems.

The P1 *core* system corresponds to the isolated chromophore unit (see Figure 2.5). Our TD-DFT calculations can be considered accurate enough to estimate the total environmental effect: this latter can be calculated as the difference between the experimental value for GFP (3.11 eV) and the value calculated for the P1 *core* optimized (3.44 eV), i.e. $\Delta$ E $\approx$ -0.32 eV. We have already attributed the first -0.10 eV to a structural effect of the environment. The additional red-shift (about -0.32 eV) enforced by the protein is of the order of magnitude of previously reported results regarding the effect of a moderated polar medium on the $\pi - \pi^*$ absorption for several systems [24].

Comparing the value for the isolated chromophore (3.24

| **GFP** | *Band position* | *Osc. str.* | *homo-lumo %* |
|---|---|---|---|
| P1 | 3.25 | 0.4740 | 68 |
| P2 | 3.24 | 0.5089 | 68 |
| P3 | 3.24 | 0.5693 | 68 |
| P4 | 3.23 | 0.5337 | 69 |
| P5 | 3.23 | 0.5308 | 69 |
| P6 | 3.16 | 0.3624 | 55 |
| $P6_{opt}$ | **3.14** | 0.5909 | 65 |
| exp | 3.11 [41, 70] | | |
| **QM core** | *Band position* | *Osc. str.* | *homo-lumo %* |
| P1opt | 3.44 | 0.6654 | 70 |
| P1 | 3.24 | 0.6350 | 70 |
| P2 | 3.29 | 0.6883 | 70 |
| P3 | 3.27 | 0.7034 | 70 |
| P4 | 3.25 | 0.6909 | 70 |
| P5 | 3.16 | 0.5277 | 59 |
| P6 | 3.08 | 0.5627 | 53 |

Table 3.4: Band positions (eV), oscillator strengths and homo-lumo transition percentage calculated for GFP QM/MM partitions and corresponding QM core systems, at the B3LYP/6-31+G(d,p)/Amber and B3LYP/6-31+G(d,p) levels of theory, respectively.

eV) to the corresponding one obtained for GFP in the P1 representation (3.25 eV) we can assess the shift induced by the protein when entirely treated by molecular mechanics. The value of the isolated chromophore is red-shifted in energy by about 0.10 eV, therefore covering about another third of the total effect. Therefore, this result suggests that a pure electrostatic influence of the protein covers up to a third of the total effect.

The values obtained for the partitions P2-P4 of GFP (see from second to the fourth row of Table 3.4) are very similar; considering that such partitions differ each other by the inclusion in the QM region of the water molecule (P2), of Ser205 (P3) and Glu222 (P4) side chains, this result suggests that the influence of these residues is of pure electrostatic nature, being well represented already for the P1 choice of the QM and MM regions. Furthermore, it is worth of noting that the absorption value of P4 for both the GFP and the core systems are at convergence and, again, very similar to GFP P1 representation. This result also indicates the pure electrostatic nature of the contribution given by the set of water, Ser205 and Glu222.

P5 includes Arg96 in the QM region. This is a positive charged residue and it is expected to have a certain influence on the electronic states involving the nearby chromophore. As matter of fact, the P5 *core* system shows a

neat red-shift (3.16 eV) with respect to the P4 *core* (3.25 eV), where Arg96 is not present ($\Delta$E = -0.09 eV, corresponding to $\Delta\lambda$ = 11 nm). The value that we obtained for P4-core is in qualitative agreement with the data reported in the quantum mechanical study by Tozzini [91], in which is underlined that the presence of Arg96 and Gln94 produces a red shift in the absorption peak. However, this shift is estimated of about 30-40nm, so with respect to our evaluation is overestimated because of the lack of the remaining environmental effects, included considering the QM/MM partition. However, the corresponding value for GFP in the P5 QM/MM arrangement, also including Arg96 in the QM region, is very close to the previous P1-P6 values (3.23 eV, fifth row of Table 3.4). This suggests that the possible effect of Arg96 is compensated by the interaction with the remaining part of the system, which, for this choice of the QM and MM regions, leads to a blue-shift of the transition. This trend is confirmed in the last calculation for GFP (P6) and the corresponding QM *core*: the inclusion of the hydrogen bond network involving Ser65 and a second water molecule brings a strong contribution to the red-shift: the value calculated for the *core* is of 3.08 eV. In the corresponding QM/MM GFP partition, again, the remaining MM part partially cancel out this effect, leading to a final value of 3.16 eV. This

value is very close to the experimental one, and is our best result for the X-ray GFP structure. A loose relaxation of the P6 GFP structure (see later in the text) leads to a value which is in excellent agreement with the experiment (3.14 eV - 395 nm - seventh row of Table 3.4).

In summary, our analysis shows that the major contribution to the chromophore absorption is due to the hydrogen bond network and the interaction with water 25, Ser205, Glu222, Ser65 side chain and water 60. About an half of this contribution (-0.1 eV) has a prevalently pure electrostatic nature, and involves water 25, Ser 205, Glu 222. Most of the remaining shift is due to Ser65 and an hydrogen bonded crystallographic water, and this contribution is quantum mechanical in nature. These residues involve a modification of the $\pi$ orbital, which is mainly localized on the chromophore. We note that Arg96 scarcely influences the GFP absorption for a compensation effect due to the interaction with the remaining of the protein. This latter also contributes to the total effect (0.08 eV), and partially cancel out the red-shift due to closest environment.

The last part of our analysis on the GFP absorption regards some comments on the oscillator strengths (second column of Table 3.4) and on the percentage of the homo-lumo in the transition taken into analysis (third column of Table 3.4). For the oscillator strenght the general trend is

that in GFP the absorption is quenched when compared to the corresponding pure QM *core* system. As matter of fact, an inspection of the electronic states involved in the absorption revealed that the environment induces a mixed character to the transition. This latter maintains a main $\pi$-$\pi^*$ character, but is mixed to other dark transitions, which decrease the final intensity. The analysis of homo-lumo contribution to the transition is interesting because it is indicative of DFT ability to describe GFP orbitals and the potential occurrence of multiple transitions. The percentage of homo-lumo in the transition for each partition is of about 68% for GFP P1, P2, P3, P4 and P5. Then it becomes lower for P6, where the value is 55% and turns higher for the optimized structure(65%). This trend let us think that when the orbitals are extended to the side chain of the Ser65 there is a mixture of transition that determines a decrease of homo-lumo contribution. When the structure is optimized there is a rearrangement of the orbitals and the homo-lumo contribution newly increases. The same trend is repeated for the chromophore, with an exception. The percentage of homo-lumo for P5-core is 59%, so there is a decrese due to the involvement of Arg96 in the orbitals. This data is in accordance with our hypothesis that including Arg96 with its positive charge in the core determines a modification of

the orbitals. However, comparing this data with the corresponding for GFP is easily noted that considering the additional contribution of the protein environment there is a balance so that the effect of Arg96 on the transition becames not quantitative, but just electrostatic.

### 3.1.3 Validation tests

In this section we report the validation tests performed to obtain the best description of GFP electronic energy for both the ground and the excited states. These tests have been performed using the following density functionals: M052X [151], CAM-B3LYP [148], B3LYP [199], LC-wPBE [149, 200, 201], PBE0 [153].
These density functionals have been employed for both ONIOM/DFT and ONIOM/TD-DFT [131, 132] energy calculations, along with the 6-31+G(d,p) and the 6-31++G (2d,2p) basis sets [204–208]. Structure optimization by ONIOM [209] has also been considered using the B3LYP/ 6-31+G(d,p)/ Amber level of theory.
Our reference is the experimental absorption peak for the GFP in neutral form, corresponding to about 3.12 eV [41, 70].
B3LYP and PBE0 are very popular hybrid density functionals, well established in describing structure and energies for a wide range of molecular systems. However,

it is recognized that such functionals may not be the best choice when, comparing the ground and the excited states energy, a strong charge transfer between the orbitals is involved [214]. Therefore, we also considered the long range corrected functionals CAM-B3LYP, M052X and LC-wPBE, which can in principle provide better results for the calculation of the excitation energy.

We calculated the excitation energies at the TD-DFT/6-31+G(d,p)/Amber level employing the five density functionals chosen for the GFP. The results for P3 and P4 partitions, chosen as specimen, are collected in Table 3.5.

We note that the long range corrected functionals systematically overestimate the transition energy with respect to the more traditional B3LYP and PBE0.

In Table 3.6 we report the results obtained for partitions P1, P2 and P3 confronting the 6-31+G(d,p) and the more extended 6-31++G(2d,2p) basis sets along with the B3LYP and the PBE0 functionals.

We note that, in general, the extension provided by the inclusion of a second diffuse function and of further polarization functions does not lead to large variations of the excitation energy, except for the P4 partition.

In conclusion, we considered the B3LYP/6-31+G(d,p) potential the best cost/accuracy compromise to proceed further in our study of GFP absorption. We employed both

| | P3 | P4 |
|---|---|---|
| M052X | 3.51 | 3.43 |
| CAM-B3LYP | 3.47 | 3.38 |
| LC-wPBE | 3.66 | 3.59 |
| PBE0 | 3.32 | 3.37<br>3.19<br>2.96 |
| B3LYP | **3.24** | **3.20**<br>**3.00** |

Table 3.5: Excitation energies (eV) for GFP P3 and P4 partitions using five different density functionals and the 6-31+G(d,p) basis set. Multiple values refer to the spurious splitting of the HOMO orbital.

| | B3LYP | | PBE0 | | exp |
|---|---|---|---|---|---|
| | *6-31+G(d,p)* | *6-31++G(2d,2p)* | *6-31+G(d,p)* | *6-31++G(2d,2p)* | |
| P1 | 3.25 | 383.76 | 3.32 | 3.30 | 3.11 [41, 70] |
| P2 | 3.24 | 3.27 | 3.37 | 3.34 | |
| P3 | 3.24 | 3.22 | 3.32 | 3.30 | |

Table 3.6: Excitation energies (eV) for GFP obtained using 6-31+G(d,p) and 6-31++G(2d,2p) split valence basis sets, respectively, with both B3LYP and PBE0 functionals. In the last column is reported the experimental value.

the electronic and the mechanical embedding for GFP system (P3 and P4 partitions are reported), employing the B3LYP/6-31+G(d,p)/Amber level of theory. The results are collected in Table 3.7, along with the values obtained for the corresponding core systems. It is evident that the molecular embedding lead to results very similar to those obtained for the isolated QM system, clearly suggesting that an electronic embedding is mandatory to treat accurately the environmental effects.

| | GFP | | core |
|---|---|---|---|
| | *elec. embed.* | *mech. embed.* | |
| P3 | 3.24 | 3.27 | 3.27 |
| P4 | 3.20 | 3.23 | 3.23 |

Table 3.7: Comparison between the electronic embedding and the mechanical embedding treatment of ONIOM for GFP P3 and P4 partitions. The results obtained for P3 and P4 cores are reported, too. The absorption peaks (eV) were obtained using B3LYP/6-31+G(d,p) potential.

## 3.2 Anionic GFP

### 3.2.1 TD-DFT accuracy check

The uncertainty about the experimental absorption of the gas phase anionic chromophore and the real effect of the protein on the chromophore absorption peak combined with the assumptions of previous works about the capacity of TD-DFT method to describe GFP or GFP-like systems [108, 215] suggested us to verify if TD-DFT could accurately describe the anionic form of GFP. For this reason we tried to reproduce the experimental absorption

trend for the anionic chromophore in vacuum, in different solvents and in protein. We chose a non polar solvent, dioxane, polar protic solvents, ethanol, methanol and water and the protein. The experimental trend is

protein [41, 70] < ethanol $\simeq$ vacuum < dioxane < methanol < water [24]

At first we calculated the error, $\text{Error}_{TD-DFT}$, that our method determines in predicting the absorption maximum for all these environments:

$$\text{Error}_{TD-DFT} = \Delta\text{E}_{TD-DFT} - \Delta\text{E}_{exp}$$

where $\Delta\text{E}_{TD-DFT}$ are the absorption values obtained at TD-DFT level and $\Delta\text{E}_{exp}$ is the experimental absorption peak.
For the anionic chromophore in vacuum we have obtained an absorption peak at 3.06 eV through TD-B3LYP/6-31+G(d,p) level of theory. For the chromophore in the different solvents and in the protein environment we commit an average error of 0.23 eV with a standard deviation of 0.2 eV. Considering that the error committed is constant we take as reference the data fitted by Dong et al. for the absorption of the anionic chromophore [24] of 2.84 eV. Our method determines an error of about 0.22 eV (3.06 vs

2.84 eV) which can be easily explained with the intrinsic theoretical limitations of predicting a metastable excitation in the continuum and is perfectly comparable to the error obtained for the calculated absorptions in the other environments. In literature TD-DFT error for the vertical excitation of the $S_0$-$S_1$ transition goes from 0.44[76] to 0.1 eV [216]. We make a constant error (represented by the turquoise bar in the graph in Figure 3.1) in our estimation of the absorption maxima for all the environments and this let us say that we accurately introduce the environment and so we could analyze the effect that the environment has on the isolated chromophore absorption. Having proved the accuracy of our level of theory we have proceeded with our study, analyzing the effect of the protein environment and conducting the solvation analysis.

### 3.2.2 Protein effect

Although the numerous studies conducted it is not clearly defined the influence of the protein environment on $HBDI^-$ chromophore.

As it is possible to see in table 3.8, the protein has an indirect effect on the chromophore, that is a structural effect and determines a red shift of -0.34 eV ($HDBI_{x-ray//vac}$ vs $HDBI_{vac//vac}$) and a bulk effect that imply a blue shift of 0.10 eV (GFP $PI_{opt//wat}$ vs $HDBI_{x-ray//vac}$). The total

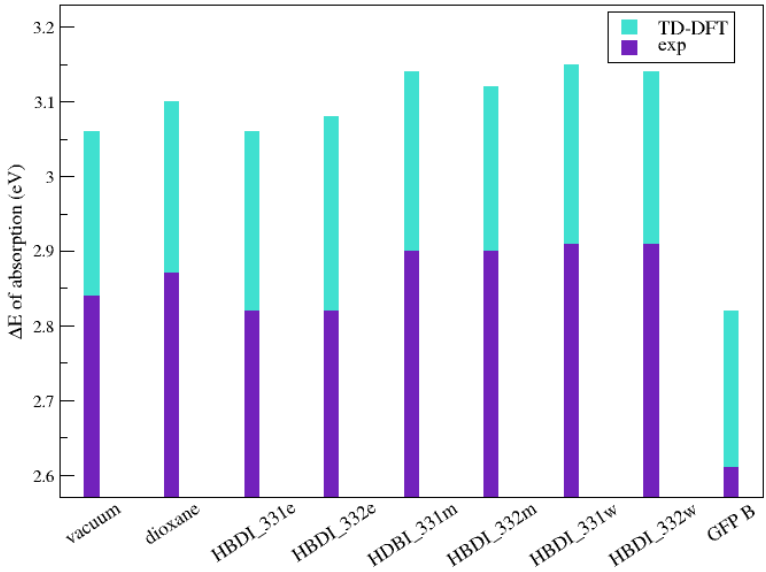



Figure 3.1: Experimental and TD-DFT absorption values (eV) for the $HBDI^-$ chromophore in vacuum, in different solvents (dioxane, ethanol, methanol, water - for reference names see table 2.1) and for the anionic GFP (GFP B). The theoretical absorption values are obtained through TD-B3LYP/6-31+G(d,p) and TD-B3LYP/6-31+G(d,p) SS CPCM level of theory for the gas phase and for the solvents, respectively. The experimental absorption value taken as references for the solvents are the ones obtained by Dong et al.[24], while for the anionic GFP is 2.61 eV [41, 70]

| $HBDI_{vac//vac}$ | $HBDI_{x-ray//vac}$ | GFP $PI_{opt//wat}$ | exp |
|---|---|---|---|
| 3.06 | 2.72 | 2.82 | 2.61 |

Table 3.8: TD-DFT absorption values (eV) are reported. $HBDI_{vac//vac}$ is the $\Delta$E in gas phase for the chromophore optimized in vacuum; $HBDI_{x-ray//vac}$ is the $\Delta$E in gas phase for the chromophore in the x-ray geometry; GFP $PI_{opt//wat}$ is the$\Delta$E of the entire QM/MM protein in water. The experimental value for the anionic GFP [41, 70] is also reported.

effect of the protein results into a $-0.24$ eV red shift, to be compared to the experimental value of -0.23 eV (2.61 [41, 70] vs 2.84 eV [24]).

### 3.2.3 HBDI anion in solution

The properties of the GFP chromophore in solutions have been characterized both experimentally [72] and theoretically [107, 109–111, 217, 218], but at the best of our knowledgements there is not a systematic solvation analysis on GFP anionic chromophore absorption. One of the major issues in literature is that the chromophore does not show regular trends of absorption bands with polarity or proticity of the solvent, as it is possible to verify looking at the experimental trend proposed before. In addition, the trend in solution seems to be of different nature when compared to the protein. In fact, if we take as reference the gas phase absorption value extrapolated by Dong, the

| | CO(P) | CO(I) | $\alpha_{478}$ | $\Gamma_{N3C4C6C7}$ |
|---|---|---|---|---|
| vacuum | 1.259 | 1.243 | 132.22 | 0.01 |
| HBDI_d | 1.276 | 1.245 | 131.44 | 0.00 |
| HBDI_e | 1.276 | 1.245 | 131.44 | 0.00 |
| HBDI_m | 1.276 | 1.245 | 131.42 | 0.00 |
| HBDI_w | 1.276 | 1.245 | 131.32 | 0.01 |
| HBDI_332e | 1.300 | 1.265 | 130.59 | 1.84 |
| HBDI_332m | 1.302 | 1.265 | 131.66 | 1.66 |
| HBDI_332w | 1.299 | 1.265 | 131.54 | 1.53 |

Table 3.9: Structural parameters (Å and degrees) for the anionic chromophore in vacuum and different solvents.

protein environment determines a red shift of about 0.2 eV (2.84 eV -absorption of gas phase chromophore- vs 2.61 eV -GFP anionic absorption in vacuum), while all the solvents determine a blue shift in the absorption.

In Table 3.9 we list some important geometrical parameters of the anionic HBDI chromophore optimized at the B3LYP/6-31G+(d,p) level of theory, corresponding to the gas phase, and at the B3LYP/6-31G+(d,p)/CPCM level of theory, corresponding to dioxane, ethanol, methanol and water solutions. For these structures, we calculated corresponding VEE values at TD-B3LYP/6-31G+(d,p) and TD-B3LYP/6-31G+(d,p)/CPCM levels of theory, respectively.

In Table 3.10 we list the energetical contributions to the VEE shift induced by the solvents, compared to experi-

| | HBDI_d | HBDI_e | HBDI_m | HBDI_w |
|---|---|---|---|---|
| $\Delta E^{01}_{ind}$ | -0.04 | -0.04 | -0.04 | -0.04 |
| $\Delta E^{01}_{pol}$ | 0.03 | 0.07 | 0.07 | 0.07 |
| $\Delta G^{01}$ | 0.04 | 0.09 | 0.09 | 0.09 |
| $\Delta G^{01}_{tot}$ | 0.03 | 0.12 | 0.12 | 0.12 |
| Exp.[a] $\Delta G^{01}_{tot}$ | 0.03 | -0.02 | 0.06 | 0.07 |

Table 3.10: Contributions to the induced solvent shift $\Delta G^{01}_{tot}$ (eV) on the $HBDI^-$ VEE. $\Delta E^{01}_{ind}$, $\Delta E^{01}_{pol}$, $\Delta G^{01}$, $\Delta G^{01}_{tot}$ are defined in eq. 2.16, 2.17, 2.19 and 2.15 respectively. 2.84 eV is the vertical excitation energy in the gas phase.

mental counterparts. The VEE calculated value for the gas phase is 3.06 eV, in agreement to previously reported results obtained with comparable potentials, while experimental shifts are reported considering 2.84 eV as the gas phase experimental value.

Dioxane leads to an hypsochromic shift of 0.04 eV with respect to the gas phase. The solvent effect can be dissected in a bathochromic shift of -0.03 eV due to an structural indirect effect, which slightly stabilizes the $S_1$ state. This effect is compensated by a hypsochromic shift of 0.03 eV due to the solvent polarization, which sligtly destabilizes the $S_1$ state. As matter of fact, $S_1$ results to be more polarizable, leading to a larger polarization work against the solvation process. Therefore, the contribution from the internal energy of the system when distorted and polarized

by the dioxane environment corresponds to the VEE in vacuum $E^{01}_{solv}(R_{solv}) = 3.05$ eV versus $E^{01}_{vac}(R_{solv}) = 3.06$ eV. The total solvent influence can be then indentified with the hypsochromic shift of 0.04 eV due to free energy shift, caused by interactions with the bulk solvent which are more favorable when the molecule is in the $S_0$ state.

The analysis of the $HBDI^-$ VEE in ethanol, methanol and water shows equivalent results, namely a total hypsochromic shift of 0.13 eV. This can be dissected in a structural indirect effect analogue to the dioxane solution (bathochromic shift of -0.03 eV). On the other hand, the hypsochromic shift to the solvent polarization effect is larger than the apolar dioxane (0.07 eV versus 0.03 eV). The free energy contributions is also more consistent (hypsochromic shift of 0.09 eV), in agreement with the polar nature of the solvents.

### 3.2.4 HBDI anion/solvent clusters in solution

In the second part of Table 3.9 we list geometrical parameters of the HBDI_332 clusters optimized at the B3LYP/6-31G+(d,p)/CPCM level of theory, corresponding to ethanol, methanol and water solutions. For these structures, we calculated corresponding VEE values at the TD-B3LYP/6-

| | HBDI_332e | HBDI_d | HBDI_332m | HBDI_332w |
|---|---|---|---|---|
| $\Delta E^{01}_{ind}$ | -0.08 | -0.04 | -0.06 | -0.06 |
| $\Delta E^{01}_{pol/int}$ | 0.01 | 0.03 | 0.03 | 0.07 |
| $\Delta G^{01}$ | 0.08 | 0.04 | 0.09 | 0.07 |
| $\Delta G^{01}_{tot}$ | 0.01 | 0.03 | 0.06 | 0.08 |
| Exp.[a] $\Delta G^{01}_{tot}$ | -0.02 | 0.03 | 0.06 | 0.07 |

Table 3.11: Contributions to the induced solvent shift $\Delta G^{01}_{tot}$ (eV) on the HBDI VEE. $\Delta E^{01}_{ind}$, $\Delta E^{01}_{pol/int}$, $\Delta G^{01}$, $\Delta G^{01}_{tot}$ are defined in eq. 2.16, 2.17, 2.19 and 2.15 respectively. 2.84 eV is the vertical excitatione energy in the gas phase.

31G+(d,p)/CPCM level of theory and reported results in Table 3.11.

All the clusters show a similar structural effect on the VEE, namely a bathochromic shift of -0.08 eV (ethanol) and of -0.06 eV (methanol and water). Therefore, the structural effect is enhanced by the presence of explicit solvent molecules, but, as it happens for the $HBDI^-$ in the bulk solvent, is not decisive to discriminate the solvent nature in the VEE.

On the other hand, contributions due to the explicit solvation and to the polarization effect are different, going from a nearly nihil effect in the ethanol (hypsochromic shift of 0.01 eV), to a more sensible variation in the methanol (0.03 eV) and water (0.07 eV) solutions, respectively. When this contribution is evaluated in the isolated clusters, the trend is similar, although the absolute values are red-

shifted to -0.04 eV, -0.02 eV and 0.03 eV for ethanol, methanol and water,respectively. This result evidences an important effect due to the bulk solvent polarization. Finally, the free energy contribution is similar (about 0.08 eV) in all cases.

It is noteworthy to compare in more detail the analysis of the VEE values in dioxane and ethanol. Due to the compensation of the polarizazion and explicit interactions effect, ethanol shows a behaviour similar to dioxane as regards the $\Delta E^{01}_{pol/int}$ term. Because of the larger polarity, absolute values of $\Delta E^{01}_{ind}$ and $\Delta G^{01}$ are instead greater in ethanol than in dioxane, but in the two solvents they compensate in a similar way with a slight red-shifted result in the ethanol case. As a consequence, $HBDI^-$ shows a VEE value in ethanol with a bathochromic shift of -0.02 eV with respect to dioxane. This is in excellent agreement with the experiment, which shows the absorption maximum in ethanol less of that in dioxane of -0.04 eV.

Regarding methanol and water, the larger polarity leads to greater $\Delta E^{01}_{pol/int}$ terms. These solvent show indeed a similar hypsochromic total shift $\Delta G^{01}_{tot}$ (0.06 eV and 0.08 eV for methanol and water, respectively), again in excellent agreement with the experiment (0.06 and 0.07 eV).

We tried to recognize the parameters which mainly influence the variation of transition energy for the chro-

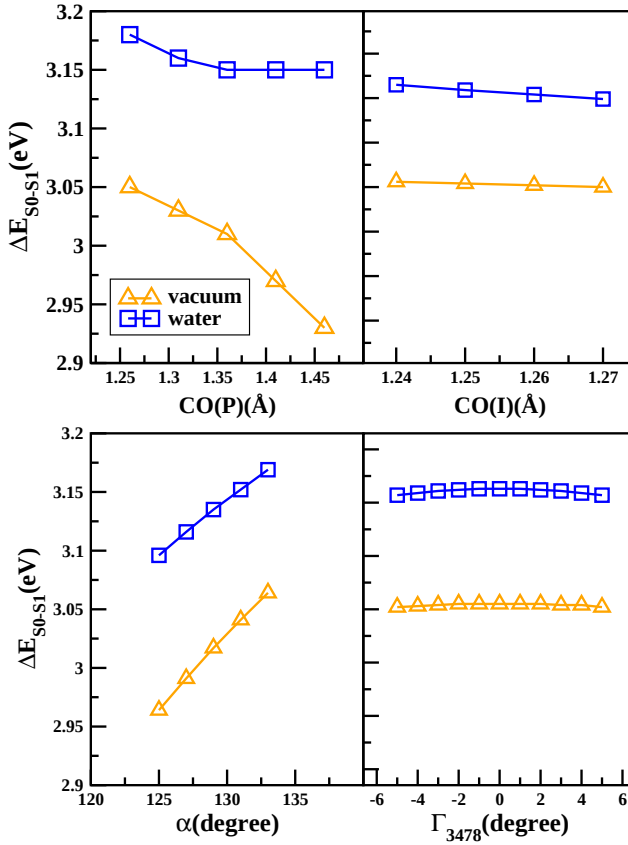



Figure 3.2: Trends of vertical excitation energies (eV) with respect to $HBDI^-$ structural parameters (Åand degree) calculated at the TD-B3LYP/6-31+G(d,p) level of theory.

mophore and reported them in Table 3.9. In order to verify our hypothesis we have taken as reference the structure of the anionic chromophore in vacuum, made some scan calculations in which we have frozen the entire structure of the chromophore, changing the value of one parameter each time into a range of values included into the ones the parameter assumes in the structures we have analyzed and have monitored the changes in the absorption band. These scan are collected in Figure 3.2

We have identified both the enhancement of the CO bonds and the reduction of the 478 angle as the parameters which determine a red shift in the absorption energy. This can be explained looking at the chromophore HOMO and LUMO orbitals (see Figure 2.13b). In fact, the reduction of 478 angle determines a stabilization of the LUMO orbital in which the electronic density is more localized on the central C atom than on the 4-7 and 7-8 bonds and the enhancement of the carboxyl bonds imply a destabilization in the HOMO orbital, in which there is a huge density on the oxygens of the CO groups with respect to the LUMO.

In order to understand the contributions which influence this interaction we analyze the interaction energy for a single molecule of each solvent on the anionic chromophore. We constructed several models in which a molecule (M)

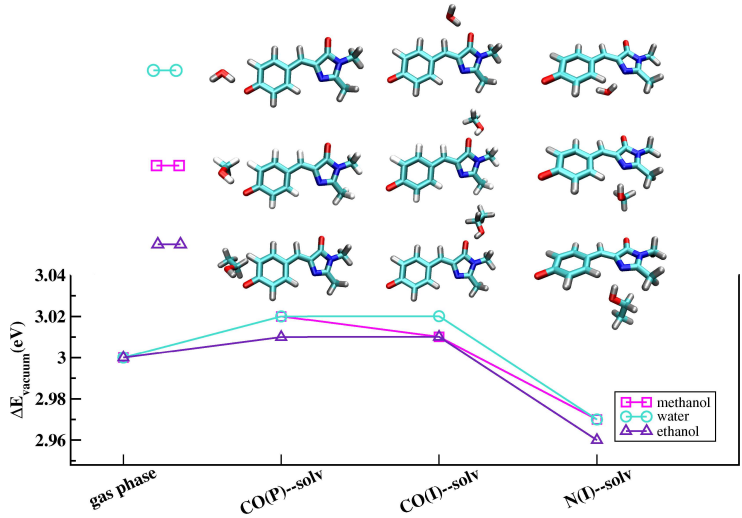



Figure 3.3: Vertical excitation energies (eV) for HBDI101, HBDI010 and HBDI001 clusters calculated at the TD-B3LYP/6-31+G(d,p) level of theory.

of each solvent is hydrogen bonded to a site of solvation of the chromophore. This for all the solvation sites. Then we calculated the interaction energy for the process:

$$\mathrm{HBDI}_{gas} + 1\mathrm{M}_{gas} = \mathrm{HBDI(M)}_{gas}$$

The interaction energy is given by $E_{int}$_1M S0 = $E_{HDBI+M}$ S0 − $E_{HBDI}$ S0 - $E_M$ S0 for the ground state and $E_{int}$_1M S1 = $E_{HDBI+M}$ S1 − $E_{HBDI}$ S1 - $E_M$ S0 for the excited state. In Figure 3.3 are reported the results obtained.

$\Delta E_{int}$_1M $S_0-S_1$ is the difference between $E_{int}$_1M S1 and $E_{int}$_1M S0 for each solvation site and $\Delta E_{int}S_0-S_1$tot is the sum of the $\Delta E_{int}$_1M $S_0-S_1$ for the different solvation sites.

We verify that with already an only molecule the blue shift is enhanced by the polarity of the solvent. In fact, the total effect on the absorption for the ethanol is null, while for methanol and water it is a little blue. In addition, the solvation on the carboxil groups determine a blue shift, while the solvation of the imidazole nitrogen determines a red shift. This could be explained looking at the HOMO and LUMO orbitals: the solvation of CO sites stabilizes the HOMO more than the LUMO (blue shift), because on its carbonil groups there is a major density localization. The N solvation stabilizes the LUMO more than the HOMO (red shift), because on LUMO there is more density on the N site. The predominant effect is determined by both the strength of the interaction involving the CO and the nitrogen site for each solvent. The dipole is higher in the ground state and so it is more stabilized by the solvent than the excited one (blue shift).

Starting from the solvation analysis we have tried to rationalize the data obtained and compared them with the effect obtained for the protein environment on the absorption of the isolated anionic chromophore $HDBI^-$. It

is confirmed for both the protein and the solvent environment a red shift due to the structural effect. For the protein a major shift is reasonable, because it forces the chromophore in a defined geometry more than the solvents do. The explicit interactions with the residues or the waters are compensated as it happens in the protein. The bulk effect is a blue shift for both the environments. Usually, the protein is considered like a non polar solvent, such as dioxane; we have obtained for dioxane a bulk effect of 0.08 eV, in perfect accordance with the results obtained for the protein.

### 3.2.5 Linear Response versus State Specific solvation

For our systems (see tables 3.12, 3.13) the use of SS and LR methods give different results in the description of the absorption transition energies.

As it is possible to see from table 3.12, for $HBDI^-$ systems LR could not distinguish among the different solvents. The excitation energies obtained for ethanol, methanol and water are really similar, whether when considering the solvent as a continuum or when clusters have been treated. In addition, LR does not determine a constant error referred to the experimental values [24] (see third

| system | LR | Δ(LR-exp) |
|---|---|---|
| HBDI_d | 2.86 | -0.01 |
| HBDI_e | 2.92 | 0.10 |
| HBDI_m | 2.93 | 0.03 |
| HBDI_w | 2.93 | 0.02 |
| HBDI_332e | 2.91 | 0.09 |
| HBDI_221e | 2.93 | 0.11 |
| HBDI_432m | 2.94 | 0.04 |
| HBDI_431m | 2.96 | 0.06 |
| HBDI_432w | 2.96 | 0.05 |
| HBDI_431w | 2.96 | 0.05 |

Table 3.12: Excitation energies (eV) for the systems referred to in column 1 (references names are defined follow the scheme in Table 2.1) computed using LR-PCM through a B3LYP/6-31+G(d,p) level of theory. In column 3 the discrepancies between the energies obtained through LR and the experimental values are indicated. Experimental reference values are 2.87 eV, 2.82 eV, 2.90 eV and 2.91 eV for dioxane, ethanol, methanol and water, respectively [24].

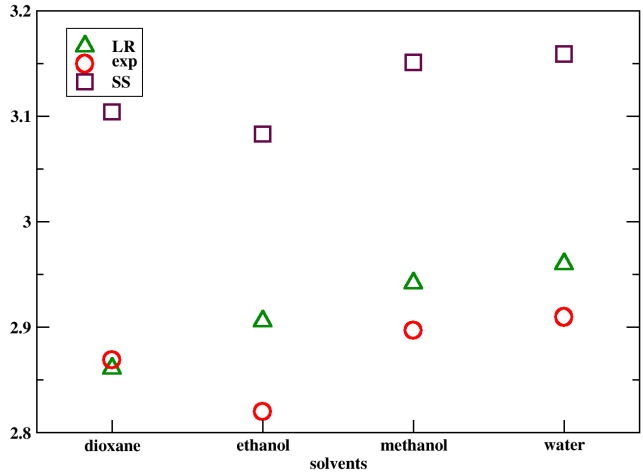



Figure 3.4: Excitation energies (eV), $\Delta E_{S_0-S_1}$, calculated through LR-CPCM and SS-CPCM methods for dioxane, ethanol, methanol and water are reported compared with the experimental values [24]

column of table 3.12) and so does not allow us to reproduce the experimental trend (see Figure 3.4) and to have a fine control of the systems treated.

Through the SS-CPCM method it is possible to operate a better distinction among the different solvents (see the second column of table 3.13), in particular between ethanol and methanol/water (which have similar excitation energies experimentally too). In addition SS performs

a quite similar error with respect to the experimental absorption values [24] and this allows us to qualitatively reproduce the experimental trend (see Figure 3.4) and to have a better control on our models. The excitation energies obtained through SS-CPCM shown an average error comparable with the error we made through our methodology for the $HBDI^-$ in vacuum, if 2.84 eV [24] is considered as experimental reference value (3.06 vs 2.84 eV, error of 0.22 eV). This supports our hypothesis that the reference value for the absorption of the anionic chromophore in vacuum is the one obtained by Dong et al. [24].

## 3.3 Proton transfer mechanism analysis

The PT from the chromophore to Glu222 leads to an anionic form of GFP. According to several studies [120], this latter is only an intermediate structure, which through a conformational change of the chromophore and of its close environment, gives the anionic form responsible of the fluorescence.

The characterization of both the intermediate and the anionic species lifetime is a crucial step to understand the GFP photochemestry. In this respect, the elucidation of

| system | SS | Δ(SS-exp) | Δ(SS-LR) |
|---|---|---|---|
| HBDI_d | 3.10 | 0.23 | 0.24 |
| HBDI_e | 3.18 | 0.36 | 0.26 |
| HBDI_m | 3.18 | 0.28 | 0.25 |
| HBDI_w | 3.18 | 0.27 | 0.25 |
| HBDI_332e | 3.08 | 0.26 | 0.18 |
| HBDI_221e | 3.14 | 0.32 | 0.21 |
| HBDI_432m | 3.15 | 0.25 | 0.21 |
| HBDI_431m | 3.16 | 0.27 | 0.20 |
| HBDI_432w | 3.16 | 0.25 | 0.20 |
| HBDI_431w | 3.17 | 0.26 | 0.20 |

Table 3.13: Excitation energies (eV) for the systems referred to in column 1 (references names are defined follow the scheme in Table 2.1) computed using SS-PCM method through a B3LYP/6-31+G(d,p) level of theory. In columns 3 the discrepancies between the energies obtained through SS and the experimental values are indicated. In column 4 the differences between the SS and the LR excitation energies are reported. Experimental reference values are 2.87 eV, 2.82 eV, 2.90 eV and 2.91 eV for dioxane, ethanol, methanol and water, respectively [24].

the PT mechanism and kinetics is a prerequisite, and the debate about the occurrence of a concerted or a step-wise mechanism is still open.

In this work we attempted to give a contribution to the characterization of both the possible paths. Our work on the PT could be divided into three steps:

1) we have conducted a qualitative analysis in which we have made some approximations. In fact, we have considered a static model in which the protein cavity is motionless and quantum effects (eg. proton tunnelling) are not taken into account.

Starting from this preliminary work, we have described the proton transfer process in both the ground and first singlet excited state using a convenient modelization that includes the main conformational protein effects on the reaction.

2)We have obtained GFP structures for the neutral and anionic forms involved into the transfer.

3) We have proposed a molecular description of the proton transfer mechanism by integrating the intrinsic reaction coordinate (IRC) [219, 220].

### 3.3.1 Approximated simple models analysis

We built up a simple model for the reaction coordinate: we chose a linear synchronus path (LSP) for the concerted mechanism and three linear paths along the three oxygen-hydrogen distances for step-wise mechanisms (SWLP).

For all the structures representing both the LSP and the SWLP reaction mechanism we performed TD-DFT calculations at the B3LYP/ 6-31+G(d,p)/Amber level. For comparison, we also performed analog calculations on the corresponding isolated *core* systems at the B3LYP/ 6-31+G(d,p) level of theory. In this way we obtained an energetic profile along the paths for both the SWLP and LSP mechanisms, and for both the ground and the excited state involved in the $\pi$-$\pi^*$ transition. As matter of fact, it is known that the shuttle is triggered by the electronic excitation and we judged interesting to evaluate the energetic differences provided by the absorption.

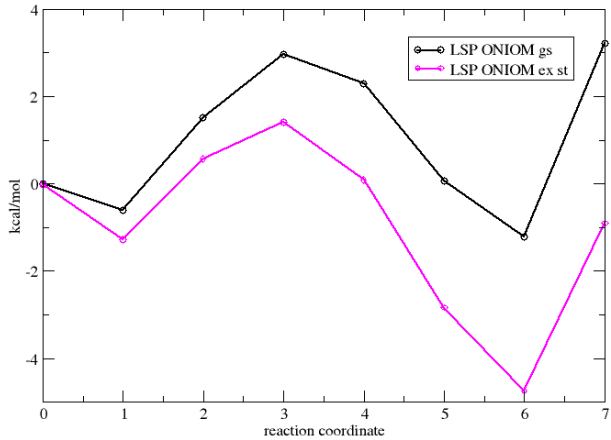
In Figure 3.5(a) we report the energetic profiles for the concerted mechanism in both the ground and the excited states of GFP. We note (Table 3.14) that the activation energy is lowered in the excited state by about 0.9 kcal/mol. In addition, while in the ground state the final form of GFP is slightly stabilized ($\Delta$E=-0.62 kcal/mol), in the

excited state it is neatly favoured ($\Delta$E=-3.5 kcal/mol). According to this description, the shuttle is favoured in the excited state.

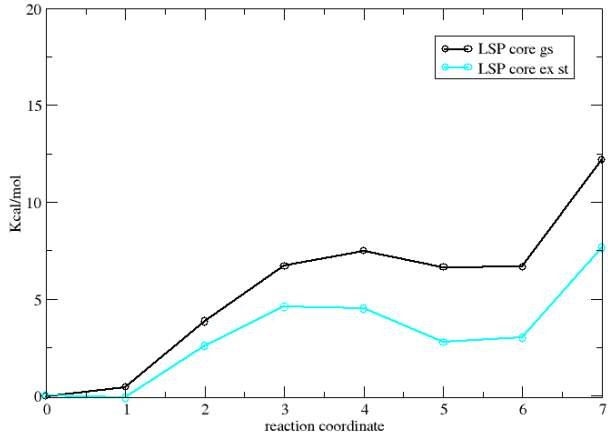
In Figure 3.5(b) we report the energetic profiles for the concerted mechanism in both the ground and the excited for the isolated QM *core* corresponding to the P6 (see Figure 2.10). Also in this case the activation energy is higher in the ground state (about 2.4 kcal/mol) than in the excited one. On the contrary, the final form of GFP is thermodinamically unfavoured in both the electronic states considered. This result (see Table 3.14 for a summary) clearly indicates an important role of the protein environment in stabilizing the anionic form and, as a consequence, the mandatory treatment of the whole system in order to get reliable conclusions.

Regarding the wise-step mechanism the whole path is composed by twenty values of the reaction coordinate: the first eight refer to the transfer of the proton from the chromophore to water; the second step (six values) regards the shuttle from water to Ser205; the latter step (6 values) regards the transfer from Ser205 to Glu222.

In Figure 3.6 are reported the energetic profiles of the wise-step path for GFP in both the ground and the excited state. Following the proton in its transfer from the chromophore to Glu222 along the hydrogen bond network,

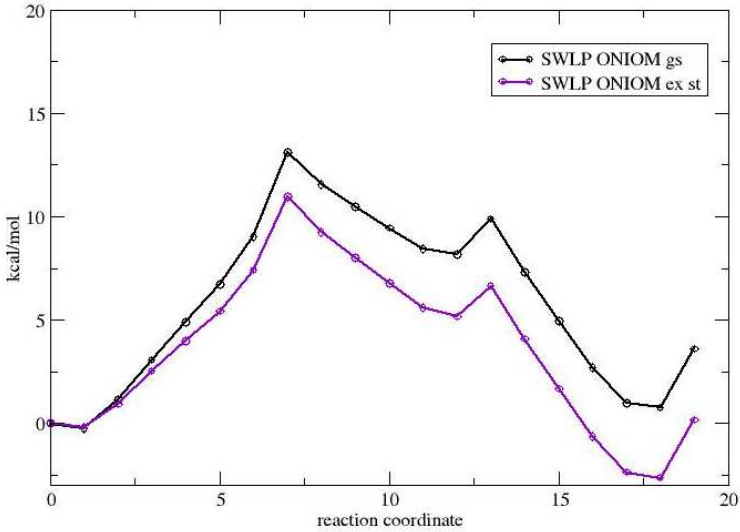



**(a) *GFP***

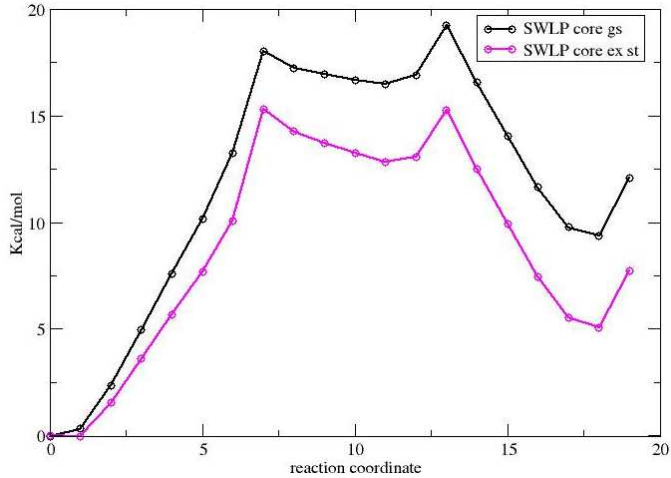



**(b) *core***

Figure 3.5: Energetic profiles (kcal/mol) of the LSP, in both ground and excited states, for GFP and *core* systems, calculated at the B3LYP/6-31+G(d,p)/Amber and at the B3LYP/6-31+G(d,p) levels of theory, respectively.




**(a) *GFP***




**(b) *core***

Figure 3.6: Energetic profiles (kcal/mol) of the wise-step PT in both the ground and the excited states for GFP and correspondin *core*, calculated at the B3LYP/6-31+G(d,p)/Amber and at the B3LYP/6-31+G(d,p) levels of theory, respectively.

| | GFP | | *core* | |
|---|---|---|---|---|
| LSP | excited | ground | excited | ground |
| $\Delta E^{\dagger}$ | 2.68 | 3.55 | 4.68 | 7.11 |
| $\Delta E$ | -3.51 | -0.62 | 3.12 | 6.30 |

Table 3.14: Activation energy ($\Delta E^{\dagger}$, kcal/mol) and energy variance ($\Delta E$, kcal/mol) of the linear synchronus PT in the ground and the excited states, calculated for GFP and *core* at the B3LYP/6-31+G(d,p)/Amber and B3LYP/6-31+G(d,p) levels of theory, respectively.

it is possible to identify two activation energies for each path. If we consider the first peak, it is possible to note that the activation energy for the excited state is lower (of about 2 kcal/mol) than the one for the ground state (see also Table 3.15 for a summary). In addition, the intermediates in the ground state are less stable than in the excited state. On the contrary, if we look at the second peak, it is easy to note that the activation energy for the ground and the excited states are very similar and the final product is favoured in both cases. So, our results bring to the conclusion that the passage from the intermediate to the final form of GFP is both thermodynamically and kinetically favoured. The first proton transfer from the chromophore to water, on the contrary

is highly activated (about 11.1 kcal/mol) and represents the rate limiting step. If we consider the whole process the wise-step proton shuttle mechanism is an esothermal reaction for the excited state, while it is endothermal for the ground state (see row 5 in Table 3.15).

In Figure 3.6(b) we also report our analysis on wise-step linear path of QM *core* system. The analysis of the SWLP for the core supports the hypothesis that the transfer is favoured in the excited state. However, the activation energies for both the excited and the ground state are larger when compared with the profiles obtained for the GFP.

It is important to underline that we made important approximations in performing these calculations: we considered the protein cavity to be motionless, while it would have been more appropriate to relax the structure at each step of the paths and to consider other dynamical aspects. Furthermore, quantum effects (proton tunnelling) have not been taken into account. These approximations, unfortunately, make this first analysis accurate only from a qualitative point of view.

In summary, the information gathered from this first approximated model of the proton shuttle is that the account of the entire protein is mandatory to study GFP photochemistry, because it decreases the activation energy of

| | GFP | | *core* | |
|---|---|---|---|---|
| SWLP | excited | ground | excited | ground |
| $\Delta E^{\dagger}_{(1)}$ | 11.11 | 13.23 | 15.22 | 17.71 |
| $\Delta E_{(1)}$ | 5.35 | 8.32 | 12.76 | 16.16 |
| $\Delta E^{\dagger}_{(2)}$ | 1.34 | 1.75 | 2.53 | 2.85 |
| $\Delta E_{(2)}$ | - 7.81 | - 7.36 | - 7.67 | - 7.03 |
| $\Delta E_{tot}$ | - 2.45 | 0.97 | 5.08 | 9.13 |

Table 3.15: Activation energy ($\Delta E^{\dagger}$, kcal/mol) and energy variance ($\Delta E$, kcal/mol) of the step-wise proton shuttle in the ground and the excited statesi, calculated for GFP and *core* at the B3LYP/6-31+G(d,p)/Amber and B3LYP/6-31+G(d,p) levels of theory, respectively.

| | LSP | | SWLP | |
|---|---|---|---|---|
| | GFP | core | GFP | core |
| $\Delta E^{\dagger}$ | 2.68 | 4.68 | 11.11 | 15.22 |
| $\Delta E$ | -3.51 | 3.12 | 5.35 | 12.76 |

Table 3.16: Summary of the energetic parameters (kcal/mol) for the LSP and the SWLP proton shuttle mechanism, calculated for the GFP and the QM *core* in the excited state.

the proton transfer for both the LSP and SWLP. Moreover, a kinetical and thermodynamical analysis of our results leads to the conclusion that the concerted mechanism is favoured with respect to the wise-step one (see Table 3.16).

### 3.3.2 GFP structure analysis

To give an answer to the requirement of using a model which can accurately represent the protein influence on the reaction, we obtained QM/MM structures which could accurately describe GFP system in the neutral and anionic form involved into the PT in both the ground state (A and I) and in the excited state involved into the absorption ($A^*$ and $I^*$).

We have conducted a structural analysis on the geometries obtained and we have identified the main geometrical parameters (reported in table 3.17, atom labels are shown in Figure 3.7) characterizing the four structures.
Our aim was to identify the protein geometrical arrangement which could favour the PT in the excited state with respect to the ground. It is worth making some remarks:

- Comparing A and $A^*$:

  1) the dihedral $N_3C_4C_7C_8$ (second row of table 3.17), referred to the bridge region of the chromophore, is bigger in the excited state with respect to the value it has in the ground state. This means that the chromophore is more distorted in $A^*$ than in A, where it is nearly planar (see upper and middle panel of Figure 3.8 ).

  2) The dihedrals referred to the PT coordinates (last eight rows of table 3.17 ) are all reduced in the excited state compared to the ground. The values of these geometrical parameters are the expression of the minor distortion of the H-bond network among the residues involved into the PT. To show this arrangement we considered an avarage plane built taking into account all the oxygens involved into the transfer and also the Ser65 oxygen which forms an

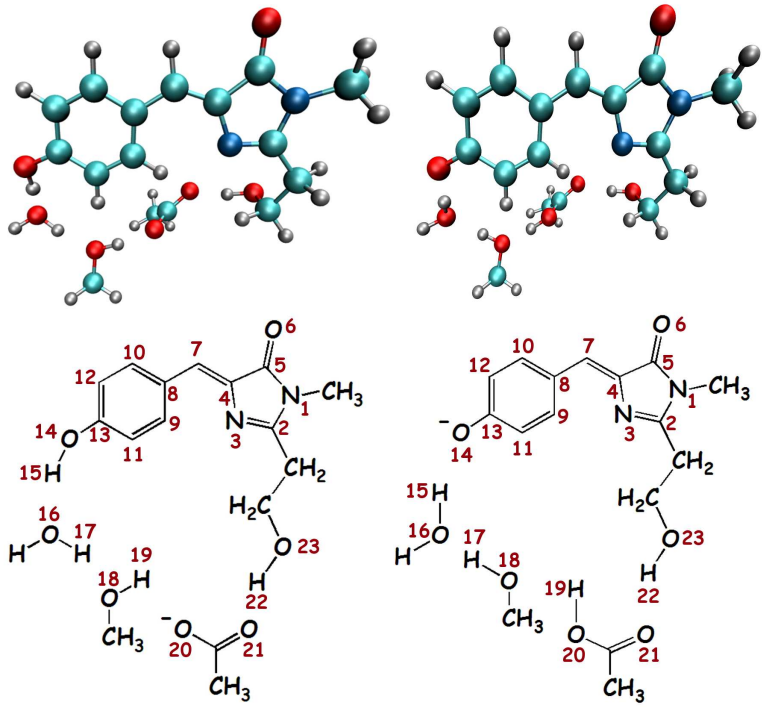


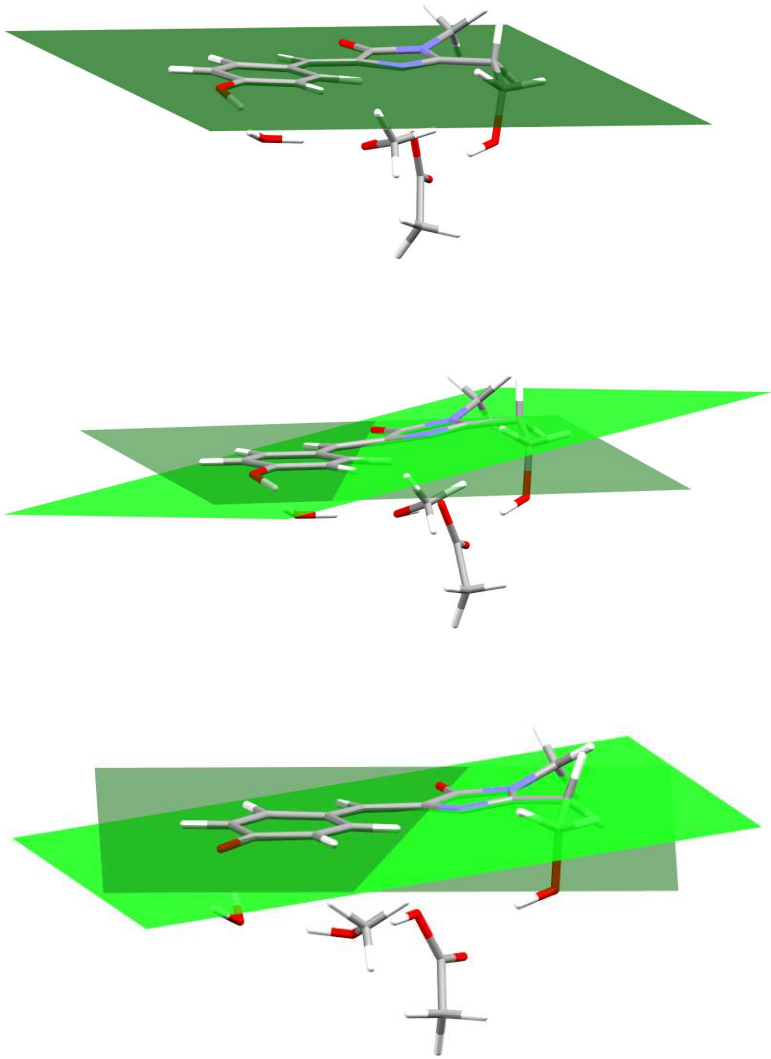

Figure 3.7: Sketches of the QM cores of A form (on the left in the panel) and I form (on the right) for the PT minimal QM/MM GFP partition. The atom labels for both the forms are reported in the lower panel.

| Geometrical parameter | A | A* | I | I* |
|---|---|---|---|---|
| $C_4C_7C_8$ | 130.71 | 129.34 | 130.74 | 127.73 |
| $\mathbf{N_3C_4C_7C_8}$ | 0.60 | -7.07 | 0.54 | -2.56 |
| $C_{13}O_{14}$ | 1.342 | 1.348 | 1.288 | 1.291 |
| $C_5O_6$ | 1.251 | 1.267 | 1.267 | 1.264 |
| $O_{14}H_{15}$ | 1.015 | 1.001 | 1.726 | 1.725 |
| $H_{15}O_{16}$ | 1.569 | 1.685 | 0.997 | 0.994 |
| $O_{14}O_{16}H_{15}$ | 4.13 | 5.66 | 1.59 | 1.79 |
| $O_{16}H_{17}$ | 0.990 | 0.996 | 1.736 | 1.675 |
| $H_{17}O_{18}$ | 1.699 | 1.684 | 0.991 | 0.997 |
| $O_{16}O_{18}H_{17}$ | 3.22 | 2.51 | 0.84 | 0.30 |
| $O_{18}H_{19}$ | 1.004 | 1.007 | 1.590 | 1.560 |
| $H_{19}O_{20}$ | 1.632 | 1.608 | 1.008 | 1.014 |
| $O_{18}O_{20}H_{19}$ | 3.10 | 2.21 | 4.39 | 4.39 |
| $\mathbf{O_{21}H_{22}}$ | 1.575 | 1.631 | 1.781 | 1.818 |
| $H_{22}O_{23}$ | 1.006 | 0.999 | 0.980 | 0.980 |
| $\mathbf{O_{21}O_{23}H_{22}}$ | 6.68 | 7.20 | 8.98 | 13.43 |
| $\mathbf{C_{11}C_{13}O_{14}O_{16}}$ | -23.29 | -19.88 | -29.73 | -33.65 |
| $\mathbf{C_{11}C_{13}O_{14}O_{18}}$ | -12.16 | -9.99 | -16.43 | -19.68 |
| $\mathbf{C_{11}C_{13}O_{14}O_{20}}$ | -12.74 | -11.22 | -17.61 | -20.61 |
| $\mathbf{C_{13}O_{14}O_{16}O_{18}}$ | 21.26 | 17.73 | 22.89 | 24.89 |
| $\mathbf{C_{13}O_{14}O_{16}O_{20}}$ | 10.40 | 8.38 | 11.68 | 13.05 |
| $\mathbf{C_{13}O_{14}O_{16}O_{23}}$ | 2.21 | 3.16 | 6.44 | 10.50 |
| $\mathbf{O_{14}O_{16}O_{18}O_{20}}$ | -22.64 | -19.48 | -23.68 | -22.92 |
| $\mathbf{O_{14}O_{16}O_{18}O_{23}}$ | -28.91 | -22.07 | -26.67 | -22.22 |

Table 3.17: Geometrical parameters for the neutral and anionic forms of GFP in the ground (A, I) and in the excited state (A*, I*) calculated using B3LYP/6-31+G(d,p)/Amber/CPCM and TD-B3LYP/6-31+G(d,p)/Amber/CPCM potentials.

.

Figure 3.8: The distortion of the chromophore is represented through planes. A form is the upper panel, A* is the middle panel and I* the down panel. The dark green plane is the one located by the atoms of the tyrosine phenolic ring, while the light green is referred to the atoms of the imidazole ring.



hydrogen bond with Glu222 (see upper and middle panel of Figure 3.9). It is possible to see that in the excited state $A^*$ the oxygen atoms involved into the transfer are less distorted respect the avarage plane than the ones in the ground.

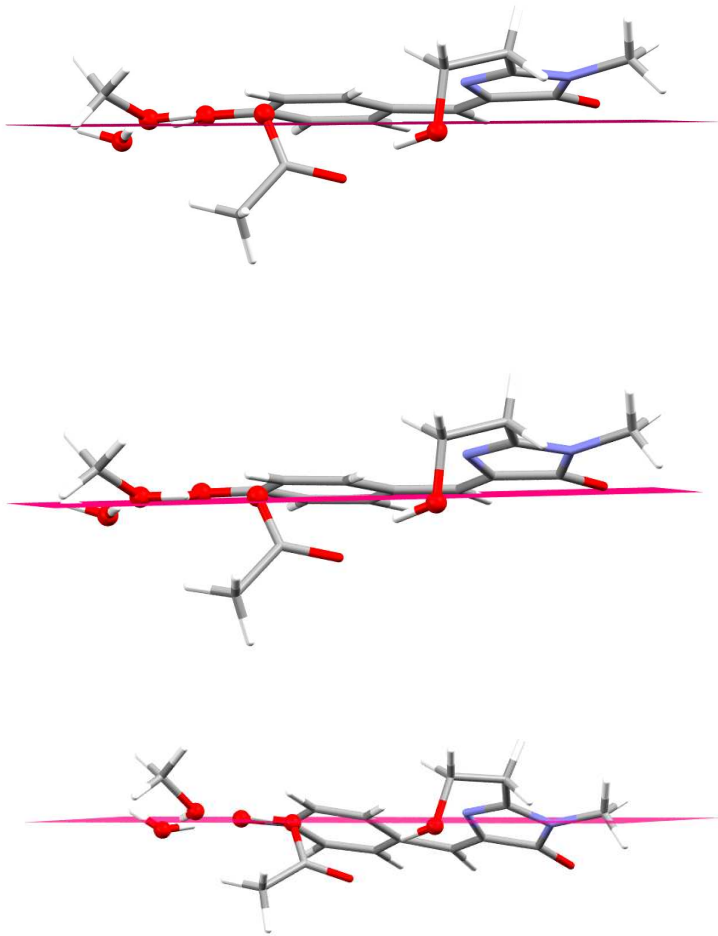
This geometric arrangement could reasonably promote the transfer in the excited state rather than in the ground. In fact, if the chromophore is more distorted the release of the phenolic proton could be favoured; in addition, the enhanced planarity of the PT H-bond network could plausibly bring to a reduction of the energy barrier of the transfer.

- Comparing $A^*$ and $I^*$:

  1) in $I^*$ the dihedral $N_3C_4C_7C_8$ (second row of table 3.17) becomes smaller again, indicating that the chromophore turns to be more planar than in $A^*$ (see down panel in Figure 3.8)

  2) In $I^*$ the H-bond network involved into the transfer (dihedrals reported in the last eight rows of table 3.17) becomes not only more distorted than the one in $A^*$, but also than the one in A form (see the down panel of Figure 3.9). This structure, in which the active site of the protein (chromophore and residues involved into the transfer) displays the

Figure 3.9: The distortion of the residues involved into the PT is considered evaluating the distances of the oxygen atoms involved into the transfer (shown in balls) from an avarage plane. The avarage plane was built taking into account the oxygens involved into the transfer (chromophore, wat25, ser205 and glu222) and the oxygen of Ser65 which forms an hydrogen bond with Glu222. A form is represented up in the panel, A* in the middle and I* in the down panel.



characteristics belonging to the arrangement of the neutral form A before the excitation, even accentuated, could be considered the result of the occurred transfer.

3) In $I^*$ (as it is already possible to see in $A^*$ if compared to A form) the H-bond distances between Glu222 and Ser65 ($O_{21}H_{22}$ in table 3.17) is lengthened and the angle $O_{21}O_{23}H_{22}$ is augmented with respect to the neutral forms, A and $A^*$. The elongation of this H bond during the proton transfer witness the less negative charge localised on Glu222 to be accomodated.

- The geometrical features of $I^*$ are mostly conserved in I.

  1) in I the dihedral $N_3C_4C_7C_8$ is nearly null, like in A form.

  2) The majority of the dihedrals describing the PT H-bond network are reduced if referred to the ones in $I^*$, but augmented with respect to A form.

### 3.3.3 IRC analysis

Having located the TS we have conducted a structural analysis of the stationary points. Then we followed the

proton transfer reaction by integrating the IRC on both the ground and the excited state potential energy surfaces, we have proposed a mechanism for the PT and evaluated the effect of the environment on the PT reaction considering two additive models for the protein.

**Structural analysis**.

In order to monitor the degree of asynchronicity of the transfer, we define the position asymmetry coordinate $\delta_i$ (as in Ref. [221]) for each event i of the whole reaction, more precisely for the HBDI-Wat25 (i=1), the Wat25-Ser205 (i=2), and the Ser205-Glu222 (i=3) proton transfer, respectively. The $\delta_i$ coordinate can be expressed as follows:

$$\delta_1 = R_{A1} - R_{1B}$$
$$\delta_2 = R_{B2} - R_{2C}$$
$$\delta_3 = R_{C3} - R_{3D}$$

where $R_x$ are hydrogen bond distances indicated according to labels of Fig. 2.19. Please note that when $\delta_i$ is equal to zero the $i$th proton is exactly in the middle of the hydrogen bond. When the $\delta_i$ value is negative the $i$th proton is closer to the hydrogen bond donor (proton transfer reactant-like). Finally, when $\delta_i$ is positive,the $i$th

| | $R_{A1}$ | $R_{1B}$ | $R_{B2}$ | $R_{2C}$ | $R_{C3}$ | $R_{3D}$ | $\delta_1$ | $\delta_2$ | $\delta_3$ |
|---|---|---|---|---|---|---|---|---|---|
| $S_0$-R | 0.992 | 1.729 | 1.005 | 1.630 | 1.035 | 1.502 | -0.737 | -0.625 | -0.467 |
| $S_0$-TS | 1.074 | 1.406 | 1.295 | 1.135 | 1.470 | 1.047 | -0.332 | 0.160 | 0.423 |
| $S_0$-P | 1.483 | 1.045 | 1.627 | 1.005 | 1.651 | 1.003 | 0.438 | 0.622 | 0.648 |
| $S_1$-R | 1.004 | 1.602 | 1.017 | 1.522 | 1.068 | 1.399 | -0.598 | -0.505 | -0.331 |
| $S_1$-TS | 1.078 | 1.391 | 1.214 | 1.200 | 1.407 | 1.066 | -0.313 | 0.014 | 0.341 |
| $S_1$-P | 1.462 | 1.048 | 1.552 | 1.014 | 1.592 | 1.005 | 0.414 | 0.538 | 0.587 |

Table 3.18: Structural parameters (Å) for the stationary points of the GFP model A

proton is closer to the hydrogen bond acceptor (proton transfer product-like).

In Table 3.18 all the the $\delta_i$ values relative to all stationary structures have been reported. By the inspection of these $\delta_i$ values is clear how the TS structure represents an intermediate structure between the R and the P states for both the electronic states. It also clear that the mechanism appears concerted and there is a neat shortening of whole H-bond distances in the TS structures. In the Table 3.19 the distances between the three oxygen-Donor and the oxygen-Acceptor atoms have been reported for all three proton chain involved in the proton transfer reaction.

The main structural changes brought up by photoexcitation are an overall shortening of hydrogen bond network

Table 3.19: $HBDI^-$ Structural parameters of the GFP model A stationary points (Å and Degree)

| | $R_{AB}$ | $R_{BC}$ | $R_{CD}$ | $\omega_{ABCD}$ | $\alpha_{C5C8C9}$ |
|---|---|---|---|---|---|
| $S_0$-R | 2.720 | 2.634 | 2.535 | -47.71 | 128.91 |
| $S_0$-TS | 2.480 | 2.429 | 2.515 | -27.65 | 129.84 |
| $S_0$-P | 2.525 | 2.628 | 2.643 | -12.05 | 130.66 |
| $S_1$-R | 2.600 | 2.530 | 2.466 | 8.14 | 128.79 |
| $S_1$-TS | 2.466 | 2.413 | 2.472 | 12.86 | 128.95 |
| $S_1$-P | 2.501 | 2.566 | 2.597 | 6.81 | 128.11 |

Table 3.20: $HBDI^-$ Structural parameters of the stationary points (Å and Degree)

| | $R_{E4}$ | $R_{COph}$ | $R_{COim}$ | $\tau_{N10C9C8C5}$ | $\phi_{C9C8C5C4}$ |
|---|---|---|---|---|---|
| $S_0$-R | 1.708 | 1.352 | 1.232 | -3.39 | -7.66 |
| $S_0$-TS | 1.835 | 1.325 | 1.233 | -2.22 | -1.10 |
| $S_0$-P | 1.873 | 1.289 | 1.237 | -0.97 | 2.18 |
| $S_1$-R | 1.695 | 1.344 | 1.244 | -1.61 | -2.98 |
| $S_1$-TS | 1.785 | 1.326 | 1.245 | -0.65 | -3.62 |
| $S_1$-P | 1.820 | 1.294 | 1.244 | -0.01 | -3.45 |

upon the excitation for all the stationary points, as can be seen by the inspection of the values in Table 3.19. All the three structures in the $S_1$ state seem to be more shrunk in order to facilitate the proton transfer, as it can be observed by narrower values of $\alpha_{C5C8C9}$ angle assumed in all the $S_1$ structures. Furthermore, the oxygen atoms involved in the proton transfer are more planar in the excited state, as it can be observed by the $\omega_{ABCD}$ dihedral value.

In the Table 3.20 we also monitor other structural parameters not directly involved in the proton transfer. These

are the distance between the Glu222 oxygen atom (O-E) and the Ser65 hydrogen atom (4), $R_{E4}$, the C-O distance of the Tyr66 ($R_{COph}$) and of imidazole ($R_{COim}$), and the dihedral angles accounting for the planarity of the chromophore ($\tau_{N10C9C8C5}$ and $\phi_{C9C8C5C4}$, see Fig. 2.19 for labels). The explicit evaluation of the Ser65 effects on the structures are mandatory. By the inspection of the $R_{E4}$ distances in all structures emerged that the electrostatic interactions of this residue on the Glu222 are very important as witnessed by the elongation of this distance during the proton transfer in both $S_0$ and $S_1$ in order to easier accommodate the less negative charge localised on Glu222 in the products respect to the reactants (see the first column of Table 3.20). From Table 3.20 we observe that in both the electronic states the proton transfer affects the $R_{COph}$ distance in a similar manner. On the contrary, the imidazole group is more involved by the electronic excitation, as we can infer by the longer $R_{COim}$ distances in all the $S_1$ structures. Moreover, by the inspection of the $\tau_{N10C9C8C5}$ and $\phi_{C9C8C5C4}$ dihedrals we note that the electronic excitation affects the the so-called bridge region of the chromophore (C5-C8-C9 group) in a significant way. This result suggests an important role played by the protein as a structural constraint on these degrees of freedom in both the energetics and the mechanism of the proton

transfer.

**Transition states**

The TS of the PES in the ground state, $S_0$, is reported in Fig. 2.21. The TS structure seems to prepare the hydrogen shuttle. All the oxygen distances involved in the wire are shrunk in respect to the reactants and products (see Table 3.19), in order to make easier the hydrogen hopping. At same time the wire is not so symmetric. By the inspection of the $\delta_i$ values is clear that the TS looks like more similar to the products, even if the first hydrogen atom in the wire (Chro-Wat) is still attached in a reactants like arrangement. By the inspection of the $\delta_2$ (Wat25-Ser205), and the $\delta_3$ (Ser205-Glu222) the positive values witness the products like shape. The negative frequency of -276.6 $cm^{-1}$ is found corresponding to the $S_0$ IRC and the main contribution to reaction coordinate is due to a collective and concerted stretching of all protons of the wire, as can be seen by the vectors belonging to hydrogen atoms of the wire (Chro-Wat25-Ser205-Glu222).

The TS of the PES in the excited state, $S_1$, is reported in Fig. 2.23. As in the $S_0$, the TS structure seems to prepare the hydrogen shuttle. All the oxygen distances involved in the wire are shrunk in respect to the reac-

tants and products (see Table 3.19), but also respect to the corresponding $S_0$ TS structure. Moreover this time all the wire network is in a very planar arrangement as we can see by the inspection of the $\omega_{ABCD}$ dihedral value in Table 3.19. The wire seems also to be more symmetric by the inspection of absolute $\delta_1$ and $\delta_3$ values and the near zero $\delta_2$ value. On this final structure all force constants and the resulting vibrational frequencies have been evaluated, in particular a negative frequency of -619.04 $cm^{-1}$ was found, corresponding to a collective stretching of all protons of the wire, as shown in Fig. 2.23.

Even if the two IRC vectors seems to be very similar each other, the excited state shrunk oxygen wire distances and their planar arrangement, combined to a more symmetric proton distribution in $S_1$ TS, explain the low barrier activation energy in the $S_1$ rather than in $S_0$.

**The proton shuttle**

We followed the proton transfer reaction by integrating the intrinsic reaction coordinate (IRC) for the GFP model A, on both the $S_0$ and $S_1$ potential energy surfaces. The initial geometry has been that of the optimized transition state, and the path has been integrated in both directions from that point.

Figure 3.10: IRC Energy profile (full black line) representing proton transfer reaction in the electronic ground state. The reactants and products energies are also reported (x symbols). Relative energies in kcal/mol are obtained at B3LYP/6-31+G(d,p) level.

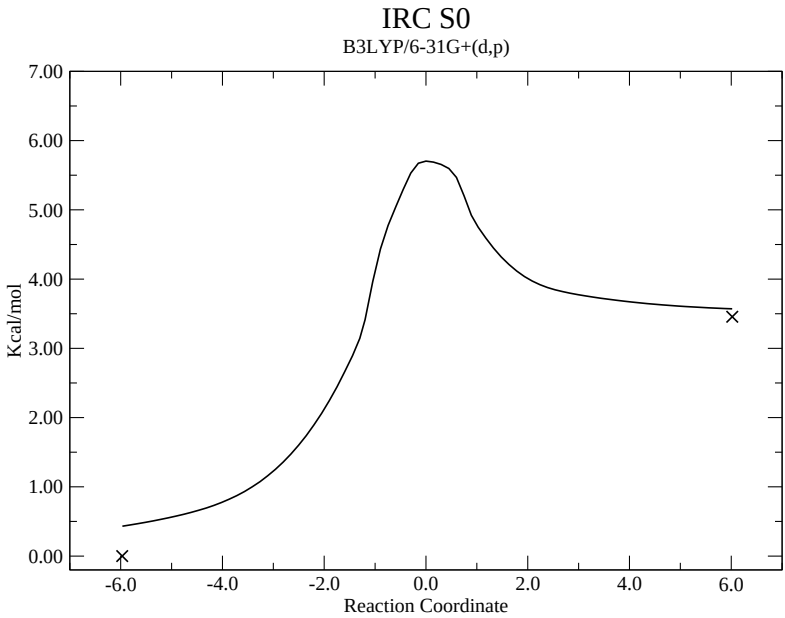


Table 3.21: $HBDI^-$ (TD)-B3LYP/6-31+G(d,p) electronic activation energy ($E_a$) and electronic energetic difference between the products and reactants ($\Delta E(P-R)$). The values are in kcal/mol.

| | $E_a$ | $\Delta E(P-R)$ |
|---|---|---|
| $E_{elect}(S_0)$ A | 5.704 | 3.456 |
| $E_{elect}(S_0)$ B | 2.990 | -1.308 |
| $E_{elect}(S_0)$ C | 0.878 | -2.529 |
| $E_{elect}(S_1)$ A | 4.081 | -0.051 |
| $E_{elect}(S_1)$ B | 2.968 | -3.208 |
| $E_{elect}(S_1)$ C | 0.355 | -4.504 |

In Fig. 3.10 we present the IRC energy profile relative to the reactant energy in the electronic ground state. It can be observed a smooth energy profile, characterized by one maximum located at TS (reaction coordinate= 0), with an activation energy of 5.70 kcal/mol. The product is 3.5 kcal/mol higher in respect to the reactant (see Table 3.21).

In Fig. 3.11 we plot the evolution of the three position asymmetry coordinates $\delta_i$ during the IRC. The coordinates are normalised, going from 0 (the reactant) to 1 (the product), and represent in percentage the progress

of the proton transfer. In spite of different oxygen-oxygen distances for each proton transfer event (see Table 3.19), the evolution of the three $\delta_i$ values suggests a synchronous and concerted mechanism. In fact, the three values of $\delta_i$ change simultaneously and become positive in proximity of the TS structure (Reaction coordinate= 0).

In Fig. 3.12 we present the IRC energy profile relative to the reactant in the first singlet electronic excited state.

In $S_1$ we observe an energy profile rougher than the corresponding in $S_0$, characterized by an higher maximum located at TS, and an activation energy of 4.08 kcal/mol ($\sim$ 1.5 kcal/mol lower than $E_a$ found in $S_0$). At variance with the ground state reaction, the product is about isoenergetic with the reactant in $S_1$. Therefore, the whole reaction is more favored in the excited state.

In Fig. 3.13 we plot the evolution of the three position asymmetry coordinates $\delta_i$ during the IRC. We observe that the reaction appears to be concerted, but the three proton movements are asynchronous, unlike the reaction in $S_0$. For example the value of $\delta_3$, associated with the proton transfer from Ser205 to Glu222, is the first to become positive, while the other two values remain negative until the proximity of the TS structure. This means that the ESPT mechanism is driven by the protonation of the Glu222 in spite of the idea that the driving force for the

Figure 3.11: IRC position asymmetry coordinate $\delta_i$ evolution representing proton transfer path in the electronic ground state (upper pannel) and normalised one (lower pannel). The $\delta_1$ (Chro-Wat), $\delta_2$ (Wat-Ser), and $\delta_3$ (Ser-Glu) are plotted in full , dashed, and dotted lines respectively. The reactants and products $\delta_1$ (Chro-Wat), $\delta_2$ (Wat-Ser), and $\delta_3$ (Ser-Glu) are also reported ($x$, circle, and star symbols respectively)

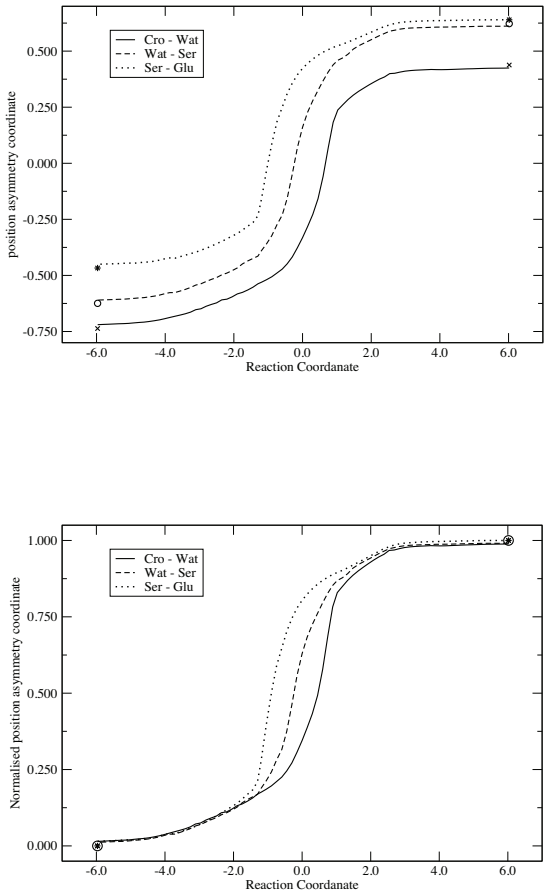


Figure 3.12: IRC Energy profile representing proton transfer reaction in the first singlet excited electronic state. The reactants and products energies are also reported (x symbols). Relative energies in kcal/mol are obtained at TD-B3LYP/6-31+G(d,p) level.

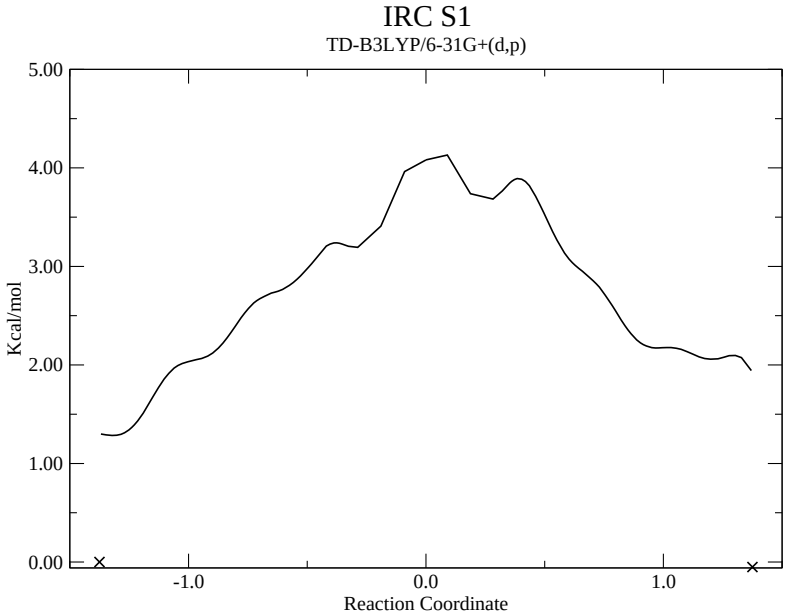
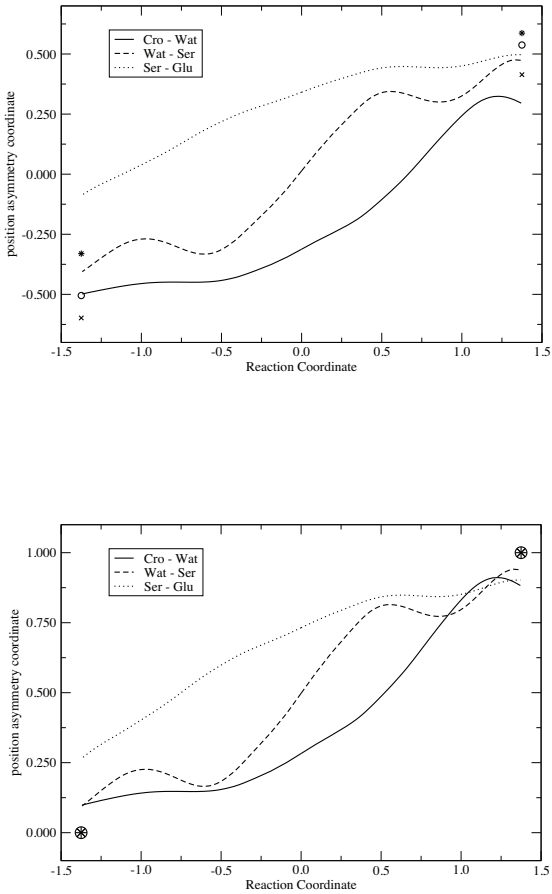


Figure 3.13: IRC position asymmetry coordinate $\delta_i$ evolution representing PT path in the first singlet excited electronic state (upper pannel) and normalised one (lower pannel). The $\delta_1$ (Chro-Wat), $\delta_2$ (Wat-Ser), and $\delta_3$ (Ser-Glu) are plotted in full , dashed, and dotted lines respectively. The reactants and products $\delta_1$ (Chro-Wat), $\delta_2$ (Wat-Ser), and $\delta_3$ (Ser-Glu) are also reported ($x$, circle, and star symbols respectively)



ESPT is the proton detachment from the chromophore.

**Energy analysis**

We considered two others models (Figure 2.20) for single point energy calculations. The first, hereafter referred to as B, can be obtained from A by adding the residue His148 in the protein like arrangement. The second one, hereafter referred to as C, can be obtained from A by eliminating the Ser65.

In Table 3.21 we reported the electronic activation energy ($E_a$) in kcal/mol for the proton transfer in the gas phase. The relative energies for all models in the ground and excited electronic states are also represented in Fig. 3.14.

By the inspection of Table 3.21 we note that the barrier is higher for model A (5.70 and 4.08 kcal/mol in the $S_0$ and $S_1$ state, respectively), while the process is favored by including the effects of His145 (2.99 and 2.97 kcal/mol), or by eliminating the electrostatic effects of Ser65 (0.88 and 0.35 kcal/mol). It is noteworthy that the B model shows the same activation energy in both the electronic states, while the barrier is lower in the $S_1$ electronic state for model A and model C systems. In the second column of Table 3.21 the difference between the products and reactants energy ($\Delta E(P-R)$) is reported in kcal/mol. In

Figure 3.14: Schematic energy profiles representing proton transfer reaction. Relative energies in kcal/mol are obtained at B3LYP/6-31+G(d,p) level.

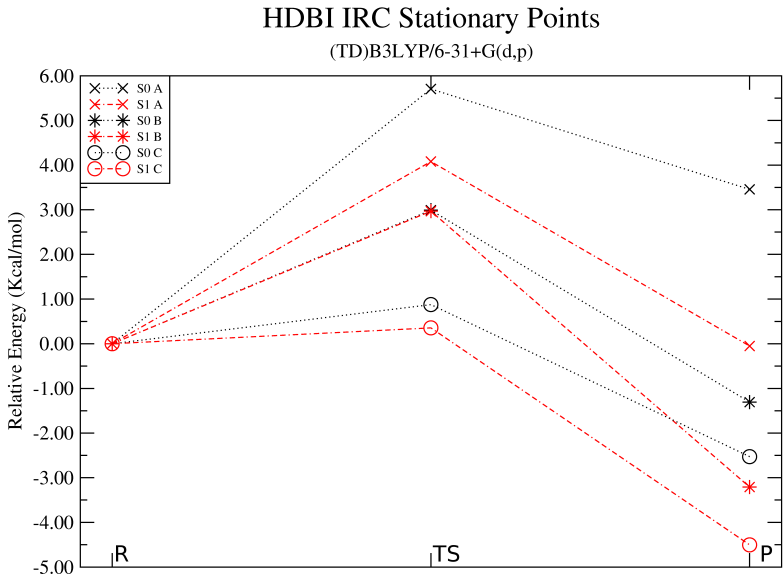
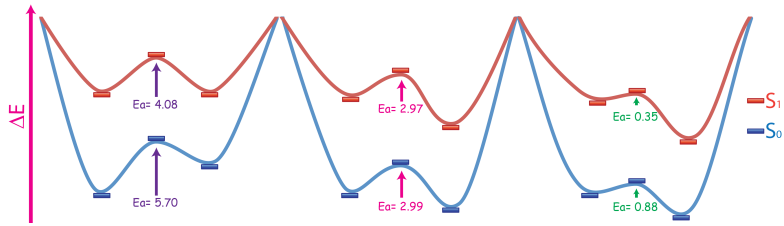


Figure 3.15: Schematic energy profiles of the PT reaction for model A, B and C, respectively. Activation energies ($E_a$, Kcal/mol) for each profile are reported.




particular for the model A system, in $S_0$ the product is less favored of 3.5 kcal/mol, while in $S_1$ the product and reactant become iso-energetic ($\Delta$E =-0.05 kcal/mol). In order to have an immediate view of the protein effect on the PT we report a sketch of the reaction energy profiles for A, B and C models (see Figure 3.15)
In Tab. 3.22 the stationary points energies (eV) relative to the ground state reactants one are also reported for each model system.
In summary:

- The protein environment has a determining role in the transfer reaction in both the ground and the excited states. This result supports the Raman spectroscopy experiments [222] attributing to the protein the function of inhibiting the radiationless de-

Table 3.22: HBDI$^-$ (TD)-B3LYP/6-31+G(d,p) stationary points relative energies (eV) for each model system.

| | A | B | C |
|---|---|---|---|
| $S_0$-R | 0.000 | 0.000 | 0.000 |
| $S_0$-TS | 0.247 | 0.130 | 0.038 |
| $S_0$-P | 0.150 | -0.057 | -0.110 |
| $S_1$-R | 3.493 | 3.494 | 3.334 |
| $S_1$-TS | 3.670 | 3.622 | 3.349 |
| $S_1$-P | 3.491 | 3.355 | 3.138 |

activation and inducing a preferential path for the transfer.

- The protons motion during the transfer could be attributed to a concerted mechanism, which is synchronous in the ground state, while more asynchronous in the excited state. The single barrier process in excited state has been already mentioned in high level post Hartree-Fock study of Vendrell et al. [221], but they found a different pathway in the ground state, probably because of their planar constrained HBDI geometry of their model compound.

- As main result this study seems to validate the VIS/ midIR pump probe hypothesis [223], suggesting that the excitation could favor the proton transfer mostly for changes in the position of protons in the H-bond network that lead to a partial initial protonation of Glu222, rather than for an increased HBDI photo-acidity.

Starting from these results and from the QM/MM GFP structures obtained for the 4 forms involved into the transfer (A,$A^*$,I and $I^*$), a future perspective could be to implement the analysis made on reduced models on the entire protein.

# 4 Conclusions

*Have no fear of perfection,*
*you' ll never reach it!*
*(S. Dalí)*

GFPs peculiar properties make them and their mutants ideal noninvasive markers in living cells, suited for numerous applications, like reporter of gene expression, cell lineage tracer, measure of protein-protein interactions, signaling and trafficking in cellular systems [44–51] and useful in many fields, such as molecular biology, cell biology [15] and biotechnology, drug discovery [16] or medicine. One of the most interesting aspects in the study of the GFPs is the analysis of their photophysics and photochemistry in order to improve the efficiency and efficacy of their applications.

In this work we have performed an in-depth study of the GFP photochemistry, focusing on the analysis of GFP absorption in order to investigate the relation among struc-

ture, optical properties and GFP functionality. In addition, evaluating GFP photoinduced behaviour, we have focused our attention on the proton shuttle triggered by excitation and we have tried to shed light on this complex issue proposing an hypothesis for its mechanism.

The first problem we have faced is a methodological question. In fact, we need to use a level of theory capable to describe all GFP forms, neutral and anionic ones. We have used DFT and TD-DFT to characterize the absorption properties of the isolated chromophores and the solvent clusters, while a DFT(TD-DFT)/MM treatment when considering the entire protein environment and the study of the relation between the structural properties and the absorption. The uncertainty about the experimental absorption of the gas phase anionic chromophore and the protein contribution to the chromophore absorption combined with the assumptions of previous works about the capacity of TD-DFT method to describe GFP or GFP-like systems [108, 215] suggested us to verify TD-DFT accuracy in describing GFP photochemistry, in particular for the anionic species. The $HBDI^-$ lowest bright excited state is metastable with respect to electron detachment, which has implications in the computation and interpretation of the $HBDI^-$ excitation energies [92, 93] and makes the system difficult to be described not only

experimentally, but also theoretically. We have been able to reproduce the experimental absorption trend for the anionic chromophore in vacuum, in different solvents and in protein, making a constant error on our predictions. This means that we accurately introduces the environment (solvent or protein) and that TD-DFT is capable to describe and control the different elements which influence GFP anionic chromophore absorption. In addition, holding a realistic tool for the evaluation of GFP absorption, we can apport our contribution to the debate about the gas-phase anionic chromophore absorption peak position. Our findins support the picture emerging from the extrapolation of the Kamlet-Taft fit of absorption experimental data in solution [24] and indicate that the protein gives rise to a considerable bathochromic shift with respect to vacuum. These results also open some questions on the interpretation of photodestruction spectroscopy experiments in the gas phase [94–96, 98] as well as on the accuracy of previous theoretical calculations in protein environment. The determination of the gas-phase anionic chromophore absorption peak is interesting by itself, but becomes really significant because it is a fundamental element to define the influence of the protein on the chromophore absorption, which is null if the photodestruction experiments are taken into account, while determines a consistent shift of

the absorption peak if considering the extrapolated value. For all GFP forms it is evident that, even if the majority of the contribution to the optical properties is due to the chromophore, it is really difficult to accurately define how the protein environment influences the intrinsic properties of the fluorescent protein, such as the position of absorption and emission peaks and the possible excited state reactions. This knowledge allow to improve the application of GFP and its mutants as biosensors and to project similar, but more efficient systems.

To treat the entire GFP structure we use a QM/MM approach. In this way we have been able to switch on and off the explicit treatment of the electronic structure of some residues and water molecules in the calculation of the spectroscopic parameters (absorption energy and intensity) and to distinguish among the several features affecting the absorption. This type of analysis can bring to results which are different from the ones obtained by quantum mechanical approach to large models. We have monitored the influence of different residues in the proximity of the chromophore and we have found that the major contribution to the chromophore absorption is due to the hydrogen bond network and the interaction with water 25, Ser205, Glu222 and Ser65 side chain. We have also dissected the protein contribution into three effects:

the first is structural and is due to the geometry imposed by the protein on the chromophore. The second is electrostatic and involves water 25, Ser 205, Glu 222. The third contribution is quantum mechanical in nature and is due to Ser65 and an hydrogen bonded crystallographic water. These residues involve a modification of the $\pi$ orbital, which is mainly localized on the chromophore.

To reach a deeper comprehension of the anionic chromophore behaviour we have conducted a solvation analysis. In fact, one of the major issues in literature is that the chromophore does not show regular trends of absorption bands with polarity or proticity of the solvent, as it is possible to verify looking at the experimental trend [24]. In addition, the trend in solution seems to be of different nature when compared to the protein. In fact, if we take as reference the gas phase absorption value extrapolated by Dong [24], the protein environment determines a red shift of about 0.2 eV (2.84 eV -absorption of gas phase chromophore- vs 2.61 eV -GFP anionic absorption in vacuum), while all the solvents determine a blue shift in the absorption. We have obtained the vertical excitation energy (VEE) of the GFP anionic chromophore in the environment of both the protein and several solvents, we have reproduced, for the first time, the following experimental trend of the absorption peak: protein (GFP)

< ethanol < dioxane < methanol < water and, as a consequence, have analyzed with confidence the relative weight of different effects affecting the solvation drift of the VEE.

Starting from the solvation analysis we have tried to rationalize the data obtained and compared them with the effect obtained for the protein environment on the absorption of the isolated anionic chromophore $HDBI^-$. It is confirmed for both the protein and the solvent environment a red shift due to the structural effect. For the protein a major shift is reasonable, because it forces the chromophore in a defined geometry more than the solvents do. The explicit interactions with the residues or the waters are compensated as it happens in the protein. The bulk effect is a blue shift for both the environments. Usually, the protein is considered like a non polar solvent, such as dioxane; we have obtained for dioxane a bulk effect of 0.08 eV, in perfect accordance with the results obtained for the protein.

We have also analyzed the proton transfer mechanism. The characterization of both the intermediate and the anionic species lifetime is a crucial step to understand the GFP photochemestry. In this respect, the elucidation of the PT mechanism and kinetics is a prerequisite, and the debate about the occurrence of a concerted (synchronous or asynchronous) or a step-wise mechanism is still open.

In this work we attempted to give a contribution to the characterization of both the possible paths. To give an answer to the requirement of using a model which can accurately represent the protein influence on the reaction, we have obtained QM/MM structures which could accurately describe GFP system in the forms involved into the PT (neutral species A and anionic intermediate I) in both the ground and the excited state involved into the absorption. From these structures we have extracted reduced models for which we have proposed a molecular description of the proton transfer mechanism by integrating the intrinsic reaction coordinate (IRC) [219, 220]. In $S_1$ we observe an energy profile rougher than in $S_0$, characterized by a maximum corresponding to the TS, and an activation energy of 4.08 kcal/mol ($\sim$ 1.5 kcal/mol lower than that found in $S_0$). At variance with the ground state reaction, where the product is less stable versus the reactant, in $S_1$ the product is about iso-energetic with the reactant. Therefore, the whole reaction is more favoured in the excited state. From inspection of the geometrical coordinates the reaction appears to be concerted, but the three proton movements are asynchronous, unlike the reaction in $S_0$. Our analysis on the PT mechanism supports the idea that the protein environment has a determining role in the transfer reaction in both the ground and the ex-

cited states. This result supports the Raman spectroscopy experiments [222] attributing to the protein the function of inhibiting the radiationless deactivation and inducing a preferential path for the transfer. The protons motion during the transfer could be attributed to a concerted mechanism, which is synchronous in the ground state, while more asynchronous in the excited state. As main result this PT study seems to validate the VIS/midIR pump probe hypothesis [223], suggesting that the excitation could favor the proton transfer mostly for changes in the position of protons in the H-bond network that lead to a partial initial protonation of Glu222, rather than for an increased HBDI photo-acidity.

We have reached a fine control on GFPs absorption, a clear understanding of the photophysics and photochemistry changes induced by the molecular environment and a deep comprehension of the relation between GFP structure and properties. These results allow to achieve highly modulated design of fluorescence and to maximize the potentiality of these proteins and use them in a more effective way to improve bioimaging techniques and nanotechnology.

# Bibliography

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[3] R. G. Parr and W. Yang, *Density- functional theory of atoms and molecules* (Oxford Univ. Press, Oxford, 1989).

[4] M. E. Casida, C. Jamorski, K. C. Casida, and D. R. Salahub, J. Chem. Phys. **108**, 4439 (1998).

[5] R. E. Stratmann, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **109**, 8218 (1998).

[6] C. V. Caillie and R. D. Amos, Chem. Phys. Lett. **308**, 249 (1999).

[7] C. V. Caillie and R. D. Amos, Chem. Phys. Lett. **317**, 159 (2000).

[8] A. Dreuw, J. L. Weisman, and M. Head-Gordon, J. Chem. Phys. **119**, 2943 (2003).

[9] D. Tozer, R. D. Amos, N. C. Handy, B. O. Roos, and L. Serrano-Andrs, Mol. Phys. **97**, 859 (1999).

[10] S. Dapprich, I. Komaromi, K. S. Byun, K. Morokuma, and M. J. Frisch, J. Mol. Struct.(Theochem) **462**, 1 (1999).

[11] T. Vreven, K. S. Byun, I. Komaromi, S. Dapprich, J. A. M. Jr., K. Morokuma, and M. J. Frisch, J. Chem. Theory and Comput. **2**, 815 (2006).

[12] T. Vreven and K. Morokuma, *Annual Reports in Computational Chemistry* (Vol.2, pp 35-51, Ed. D. C. Spellmeyer, 2006).

[13] F. Clemente, T. Vreven, and M. J. Frish, *Quantum Biochemistry* (Ed. C. Matta (Wiley VCH), 2008).

[14] M. Zimmer, Chem. Rev. **102**, 759 (2002).

[15] T. Misteli and D. L. Spector, Nat. Biotechnol. **15**, 961 (1997).

[16] D. L. Taylor, E. S. Woo, and K. A. Giuliano, Curr. Opin. Biotechnol. **12**, 75 (2001).

[17] Chem. Soc. Rev. **38** (2009).

[18] W. W. Ward, H. J. Prentice, A. F. Roth, C. W. Cody, and S. C. Reeves, Photochem. Photobiol. **35**, 803 (1982).

[19] H. Lossau, A. Kummer, R. Heinecke, F. Poellinger-Dammer, C. Kompa, G. Bieser, T. Jonsson, C. M. Silva, M. M. Yang, D. C. Youvan, and M. E. Michel-Beyerle, Chem. Phys. **213**, 1 (1996).

[20] A. Kummer, J. Wiehler, H. Rehaber, C. Kompa, B. Steipe, and M. E. Michel-Beyerle, J. Phys. Chem. **104**, 4791 (2000).

[21] C. Scharnagl and R. A. Raupp-Kossmann, J. Phys. Chem. B **108**, 477 (2004).

[22] M. Kondo, I. A. Heisler, D. Stoner-Ma, P. J. Tonge, and S. R. Meech, J. Am. Chem. Soc. **132**, 1452 (2010).

[23] S. Olsen, K. Lamothe, and T. J. Martinez, J. Am. Chem. Soc. **132**, 1192 (2010).

[24] J. Dong, K. M. Solntsev, and L. M. Tolbert, J. Am. Chem. Soc. **128**, 12038 (2006).

[25] A. Sinicropi, T. Andruniow, N. Ferr, R. Basosi, and M. Olivucci, J. Am. Chem. Soc. **127**, 11534 (2005).

[26] L. Lammich, M. A. Petersen, M. B. Nielsen, and L. H. Andersen, Biophysical Journal **92**, 201 (2007).

[27] S. B. Nielsen, A. Lapierre, J. U. Andersen, U. V. Pedersen, S. Tomita, and L. H. Andersen, Phys. rev. Lett. **87**, 228102 (2001).

[28] A. Matsuura, T. Hayashi, H. Sato, A. Takahashi, and M. Sakurai, Chem. Phys. Lett. **484**, 324 (2010).

[29] I. Topol, J. Collins, I. Polyakov, B. Grigorenko, and A. Nemukhin, Biophys. Chem. **145**, 1 (2009).

[30] C. Filippi, F. Buda, L. Guidoni, and A. Sinicropi, J. Chem. Theory Comput. **8**, 112 (2012).

[31] J. Morin and J. Hastings, J. Cell Physiol. **77**, 313 (1971).

[32] D. C. Prasher, V. Eckenrode, W. Ward, F. Pendergast, and M. J. Cornier, Gene **111**, 229 (1992).

[33] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Pendergast, Science **263**, 802 (1994).

[34] O. Shimomura, *In Green Fluorescent Protein* (M. Chalfie and S. Kain eds., Ed. Wiley-Liss:New York, 1998).

[35] O. Shimomura, B. Musicki, and Y. Kishi, Biochem. J. **251**, 405 (1988).

[36] O. Shimomura, Biol. Bull. **189**, 1 (1995).

[37] J. M. Kendall and M. N. Badminton, TIBTECH **16**, 216 (1998).

[38] W. W. Ward, Photochem. Photobiol. Rev. **1**, 57 (1979).

[39] D. Davenport and J. A. C. Nicol, Proc. R. Soc. London, Ser. B **144**, 399 (1955).

[40] W. W. Ward, *Bioluminescence and Chemiluminescence* (M. A. De Luca and W. D. McElroy eds., pp. 235-242, Ed. Academic Press, 1981).

[41] W. W. Ward and S. H. Bokman, Biochemistry **21**, 4535 (1982).

[42] M. W. Cutler and W. W. Ward, Photochem. Photobiol. **57**, 63S (1993).

[43] M. A. Ehrmann, C. H. Scheyling, and R. F. Vogel, Lett. Appl. Microbiol. **32**, 230 (2001).

[44] R. Y. Tsien, Annu. Rev. Biochem. **67**, 509 (1998).

[45] A. B. Cubitt, R. Heim, S. R. Adams, A.E.Boyd, L. Gross, and R. Y. Tsien, Trends Biochem. Sci. **20**, 448 (1995).

[46] R. Heim and R. Y. Tsien, Curr. Biol. **6**, 178 (1996).

[47] A. Miyawaki, J. Llopis, R. Heim, J. McCaffery, J. A. Adams, M. Ikura, and R. Y. Tsien, Nature **388**, 882 (1997).

[48] J. F. Presley, N. A. Cole, T. A. Schroer, K. Hirschborn, J. M. Zaal, and J. Lippincott-Schwartz, Nature **389**, 81 (1997).

[49] R. Pepperkok, A. Squire, S. Geley, and P. I. H. Bastiaens, Curr. Biol. **9**, 269 (1999).

[50] A. Miyawaki, O. Griesbeck, R. Heim, and R. Y. Tsien, Proc. Natl. Acad. Sci. **96**, 2135 (1999).

[51] R. Mitra, C. Silva, and D. Youvan, Gene **173**, 13 (1996).

[52] T. Chishima, Y. Miyagi, X. Wang, E. Baranov, Y.Tan, H. Shimada, A. R. Moossa, and R. M. Hoffman, Clin. Exp. Metastasis **15**, 547 (1997).

[53] T. Chishima, Y. Miyagi, X. Wang, H. Yamaoka, H. Shimada, A. R. Moossa, and R. M. Hoffman, Cancer Res. **57**, 2042 (1997).

[54] R. M. Hoffman, Biotechniques **30**, 1024 (2001).

[55] J. L. et al., Nature **450**, 56 (2007).

[56] R. N. Day, Mol. Endocrinol. **12**, 1410 (1998).

[57] N. P. Mahajanm, K. Linder, G. Berry, G. W. Gordon, R. Heim, and B. Herman, Nat. Biotechnol. **16**, 547 (1998).

[58] X. Xu, A. L. Gerad, B. C. Huang, D. C. Anderson, D. G. Payan, and Y. Luo, Nucleic Acids Res. **26**, 2034 (1998).

[59] Y. Suzuki, T. Yasunaga, R. Ohkura, T. Wakabayaschi, and K. Sutoh, Nature **396**, 380 (1998).

[60] L. L. Pearce, R. E. Gandley, W. Han, K. Wasserloos, M. Stitt, A. J. Kanai, M. K. McLaughlin, B. R. Pitt, and E. S. Levitan, Proc. Natl. Acad. Sci. **97**, 477 (2000).

[61] F. Yang, L. G. Moss, and G. N. Phillips, Nature Biotechnology **14**, 1246 (1996).

[62] B. Rao, M. Kemple, and F. Pendergast, Biophys. J. **32**, 630 (1980).

[63] W. W. Ward and S. H. Bokman, Biochemistry **21**, 4535 (1982).

[64] W. W. Ward, H. J. Prentice, A. F. Roth, C. W. Cody, and S. C. Reeves, Photochem. Photobiol. **35**, 803 (1982).

[65] R. Heim, D. C. Prasher, and R. Y. Tsien, Proc. Natl. Acad. Sci. **91**, 12501 (1994).

[66] V. Helms, T. P. Straatsma, and J. A. McCammon, J. Phys. Chem. B **103**, 3263 (1999).

[67] M. A. L. Marques, X. Lopez, D. Varsano, A. Castro, and A. Rubio, Phy. Rev. Lett. **90**, 258101 (2003).

[68] M. Chattoraj, B. A. King, G. U. Bublitz, and S. G. Boxer, Proc. Natl. Acad. Sci. **93**, 8362 (1996).

[69] H. Lossau, A. Kummer, R. Heinecke, F. Poellinger-Dammer, C. Kompa, G. Bieser, T. Jonsson, C. M. Silva, M. M. Yang, D. C. Youvan, and M. E. Michel-Beyerle, Chem. Phys. **213**, 1 (1996).

[70] T. M. H. Creemers, A. J. Lock, V. Subramaniam, T. M. Jovin, and S. Volker, Nat. Struct. Biol. **6**, 557 (1999).

[71] G. Striker, V. Subramaniam, C. A. M. Seidel, and A. Volkmer, J. Phys. Chem. B **103**, 8612 (1999).

[72] M. Vengris, I. H. M. van Stokkum, X. He, A. F. Bell, P. J. Tonge, R. van Grondelle, and D. S. Larsen, J. Phys. Chem. A **108**, 4587 (2004).

[73] A. A. Voityuk, M.-E. Michel-Beyerle, and N. Rosch, Chem. Phys **231**, 13 (1998).

[74] W. Weber, V. Helms, J. McCammon, and P. Langhoff, Proc. Natl. Acad. Sci. **96**, 6177 (1999).

[75] M. E. Martin, F. Negri, and M. Olivucci, J. Am. Chem. Soc. **126**, 5452 (2004).

[76] O. Vendrell, R. Gelabert, M. Moreno, and J. M. Lluch, Chem. Phys. Lett. **396**, 202 (2004).

[77] S. S. Patnaik, S. Trohalaki, and R. Pachter, Biopolymers **75**, 441 (2004).

[78] S. S. Patnaik, S. Trohalaki, R. R. Naik, M. O. Stone, and R. Pachter, Biopolymers **85**, 253 (2007).

[79] H. Niwa, S. Inouye, T. Hirano, T. Matsuno, M. Kojima, M. Kubota, M. Ohashi, and F. I. Tsuji, Proc. Natl. Acad. Sci. **93**, 13617 (1996).

[80] A. D. Kummer, C. Kompa, H. Lossau, F. Poellinger-Dammer, M. E. Michel-Beyerle, C. M. Silva, E. J. Bylina, W. J. Coleman, M. M. Yang, and D. C. Youvan, Chem. Phys. **237**, 183 (1998).

[81] R. M. Dickson, A. B. Cubitt, R. Y. Tsien, and W. E. Moerner, Nature **388**, 355 (1997).

[82] E. J. G. Peterman, S. Brasselet, and W. E. Moerner, J. Phys. Chem. **103**, 10553 (1999).

[83] M. F. Garcia-Parajo, G. M. J. Segers-Nolten, J.-A. Veerman, J. Greeve, and N. F. Hulst, Proc. Natl. Acad. Sci. **97**, 7237 (2000).

[84] D. W. Pierce, N. HomBooher, and R. D. Vale, Nature **388**, 338 (1997).

[85] D. W. Pierce and R. D. Vale, Methods Cell. Biol. **58**, 49 (1998).

[86] A. A. Voityuk, M. E. Michel-Beyerle, and N. Rosch, Chem. Phys. Lett. **272**, 162 (1997).

[87] A. F. Bell, X. He, R. M. Wachter, and P. J. Tonge, Biochemistry **39**, 4423 (2000).

[88] M. Elsliger, R. M. Wachter, G. T. Hanson, K. Kallio, and S. J. Remington, Biochemistry **38**, 5296 (1999).

[89] V. Tozzini and R. Nifosi, J. Phys. Chem. B **105**, 5797 (2001).

[90] R. Carr and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[91] T. Laino, R. Nifosi, and V. Tozzini, Chem. Phys. **298**, 17 (2004).

[92] K. B. Bravaya, M. G. Khrenova, B. L. Grigorenko, A. V. Nemukhin, and A. I. Krylov, J. Phys. Chem. B **115**, 8296 (2011).

[93] K. B. Bravaya, B. L. Grigorenko, A. V. Nemukhin, and A. I. Krylov, Acc. Chem. Res. **45**, 265 (2012).

[94] S. B. Nielsen, A. Lapierre, J. U. Andersen, U. V. Pedersen, S. Tomita, and L. H. Andersen, Phys. Rev. Lett. **87** (2001).

[95] L. H. Andersen, H. Bluhme, S. Boye, T. J. D. Jorgensen, H. Krogh, I. B. Nielsen, S. B. Nielsen, and A. Svendsen, Phys. Chem. Chem. Phys. **6**, 2617 (2004).

[96] M. W. Forbes and R. A. Jockusch, J. Am. Chem. Soc. **131**, 17038 (2009).

[97] N. M. Webber and S. R. Meech, Photochem. Photobiol. Sci. **6**, 976 (2007).

[98] K. Chingin, R. M. Balaboin, V. Frankevich, K. Barylyuk, R. Nieckarz, P. Sagulenko, and R. Zenobi, Int. J. Mass Spectrom. **306**, 241 (2011).

[99] A. A. Voityuki, M. E. Michel-Beyerle, and N. Rosch, Chem. Phys. Lett. **227**, 162 (1997).

[100] V. Helms, C. Winstead, and P. W. Langhoff, J. Molec. Struct. (Theochem) **506**, 179 (2000).

[101] A. K. Das, J.-Y. Hasegawa, T. Miyahara, M. Ehara, and H. Nakatsuji, J. Comput. Chem. **24**, 1421 (2003).

[102] A. Toniolo, S. Olsen, L. Manohara, and T. J. Martinez, Faraday Discuss. **127**, 149 (2004).

[103] I. Polyakov, B. G. E. M. Epifanovsky, A. I. Krylov, and A. V. Nemukhin, J. Chem. Theory Comput. **6**, 2377 (2010).

[104] S. Olsen and S. C. Smith, J. Am. Chem. Soc. **130**, 8677 (2008).

[105] E. Epifanovsky, I. Polyakov, B. Grigorenko, A. Nemukhin, and A. I. Krylov, J. Chem. Theory Comput. **5**, 1895 (2009).

[106] K. Burke and E. K. U. Gross, *Density functionals: theory and applications* (pp. 116-146, Springer Berlin, 1998).

[107] A. V. Nemukhin, I. A. Topol, and S. K. Burt, J. Chem. Theory Comput. **2**, 292 (2006).

[108] C. Filippi, M. Zaccheddu, and F. Buda, J. Chem. Theory Comput. **5**, 2074 (2009).

[109] W. Yan, L. Zhang, D. Xie, and J. Zeng, J. Comput. Chem **26**, 1487 (2005).

[110] I. Polyakov, E. Epifanovsky, B. Grigorenko, A. I. Krylov, and A. V. Nemukhin, J. Chem. Theory Comput. **5**, 1907 (2009).

[111] D. Zuev, K. B. Bravaya, M. V. Makarova, and A. I. Krylov, J. Chem. Phys. **135**, 194304 (2011).

[112] R. Y. Tsien, Annu. Rev. Biochem. **67**, 509 (1998).

[113] G. H. Patterson, S. M. Knobel, W. D. Sharif, S. R. Kain, and D. W. Piston, Biophys. J. **73**, 2782 (1997).

[114] K. Andersson, P. A. Malmqvist, and B. Roos, J. Chem. Phys. **96**, 1218 (1992).

[115] A. A. Voityuk, M.-E. Michel-Beyerle, and N. Rosch, Chem. Phys. Lett. **296**, 269 (1998).

[116] D. Mandal and T. T. andS. R. Meech, J. Phys. Chem. B **108**, 1102 (2004).

[117] A. K. Das, J.-Y. Hasegawa, T. Miyahara, M. Ehara, and H. Nakatsuji, J. Comp. Chem. **24**, 1421 (2003).

[118] M. C. Chen, C. R. Lambert, J. D. Urgitis, and M. Zimmer, Chemical Physics **270**, 157 (2001).

[119] S. L. Maddalo and M. Zimmer, Photochemistry and Photobiology **82**, 367 (2006).

[120] P. J. Tonge and S. R. Meech, J. Photochem. Photobiol. A **205**, 1 (2009).

[121] H.-Y. Yoo, J. A. Boatz, V. Helms, J. A. McCammon, and P. W. Langhoff, J. Phys. Chem. B **105**, 2850 (2001).

[122] M. A. Lill and V. Helms, Proc. Natl. Acad. Sci. **99**, 2778 (2002).

[123] H. Zhang, S. Wang, Q. Sun, and S. C. Smith, Phys. Chem. Chem. Phys **11**, 8422 (2009).

[124] A. Shinobu and N. Agmon, J. Phys. Chem. A **113**, 7253 (2009).

[125] S. Wang and S. C. Smith, Chemical Physics **326**, 204 (2006).

[126] X. Shu, P. Leiderman, R. Gepshtein, N. R. Smith, K. Kallio, D. Huppert, and S. J. Remington, Protein Science **16**, 2703 (2007).

[127] N. Reuter, H. Lin, and W. Thiel, J. Phys. Chem. B **106**, 6310 (2002).

[128] M. Cossi, N. Rega, G. Scalmani, and V. Barone, J. Comput. Chem. **24**, 669 (2003).

[129] S. Miertus, E. Scrocco, and J. Tomasi, Chem. Phys. **55**, 117 (1981).

[130] V. Barone and M. Cossi, J. Phys. Chem. A **102**, 1995 (1998).

[131] T. Vreven and K. Morokuma, Theor Chem Acc **109**, 125 (2003).

[132] T. Vreven and K. Morokuma, J. Chem. Phys. **113**, 2969 (2000).

[133] T. Vreven, B. Mennucci, C. O. da Silva, K. Morokuma, and J. Tomasi, J. Chem. Phys. **115**, 62 (2001).

[134] S. J. Mo, T. Vreven, B. Mennucci, K. Morokuma, and J. Tomasi, Theor. Chem. Acc. **111**, 154 (2003).

[135] M. A. L. Marques and E. K. U. Gross, *A Primer in Density Functional Theory* (Springer,Berlin, 2003).

[136] R. J. Buenker, S. D. Peyerimhoff, and P. J. Bruna, *Computational Theoretical Organic Chemistry* (I. G. Csizmadia, R. Daudel, Eds., Reidel, Dordrecht, The Netherlands, p. 91, 1981).

[137] J. Gauss, *Encyclopedia of Computational Chemistry* (P. v. R. Schleyer, N. L. Clark, J. Gasteiger, H. F. Schaefer III, P. R. Schreiner, Eds., Wiley, Chichester, p.615., 1998).

[138] *Modern Ideas in Coupled-Cluster Methods* (R. J. Barlett, Ed., World Scientific, Singapore, 1997).

[139] M. A. L. Marques and E. K. Gross, Annu. Rev. Phys. Chem. **55**, 427 (2004).

[140] M. E. Casida, In Recent Advances in Density Functional Methods, Part I; Chong D. P. Ed.; World Scientific Singapore pp. 155–188 (1995).

[141] E. R. Bitnner and D. S. Kosov, J. Chem. Phys. **110**, 6645 (1999).

[142] F. Furche and R. Ahlrichs, J. Chem. Phys. **117**, 7433 (2004).

[143] M. Cossi and V. Barone, J. Chem. Phys. **115**, 4708 (2001).

[144] *Structure and Dynamics of Electronic Excited States* (J. Laane, H. Takahashi, A. D. Bandrauk, Eds., Springer, Berlin, 1998).

[145] M. Wanko, M. Garavelli, F. Bernardi, T. A. Niehaus, T. Frauenheim, and M. Elstner, J. Chem. Phys. **120**, 1674 (2004).

[146] A. Dreuw and M. Head-Gordon, J. Am. Chem. Soc. **126**, 4007 (2004).

[147] A. Dreuw and M. Head-Gordon, Chem. Rev. **105**, 4009 (2005).

[148] T. Yanai, D. Tew, and N. Handy, Chem. Phys. Lett. **393**, 51 (2004).

[149] O. A. Vydrov and G. E. Scuseria, J. Chem. Phys. **125**, 234109 (2006).

[150] Y. Zhao and D. G. Truhlar, Theor. Chem. Acc. **120**, 215 (2008).

[151] Y. Zhao, N. E. Schultz, and D. G. Truhlar, J. Chem. Theory and Comput. **2**, 364 (2006).

[152] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[153] C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).

[154] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frish, J. Phys. Chem. **98**, 11623 (1994).

[155] C. Adamo and V. Barone, J. Chem. Phys. **108**, 664 (1998).

[156] F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy, J. Chem. Phys. **109**, 6264 (1998).

[157] H. L. Schmider and A. D. Becke, J. Chem. Phys. **108**, 9624 (1998).

[158] B. Paizs and S. Suhai, J. Comput. Chem. **19**, 575 (1998).

[159] D. Jacquemin, J. Andr, and E. A. Perpete, J. Chem. Phys. **121**, 4389 (2004).

[160] D. Jacquemin, A. Femenias, H. Chermette, I. Ciofini, C. Adamo, J. Andr, and E. A. Perpete, J. Phys. Chem. A **110**, 5952 (2006).

[161] B. Champagne, E. A. Perpete, S. van Gisbergen, E. J. Baerends, J. G. Snijders, C. Soubra-Ghaoui, K. Robins, and B. Kirtman, J. Chem. Phys. **109**, 10489 (1998).

[162] B. Champagne, E. A. Perpete, D. Jacquemin, S. van Gisbergen, E. J. Baerends, C. Soubra-Ghaoui, K. Robins, and B. Kirtman, J. Phys. Chem. A **104**, 4755 (2000).

[163] D. Guillaumont and S. Nakamura, Dyes Pigm. **46**, 85 (2000).

[164] D. J. Tozer, J. Chem. Phys. **119**, 12697 (2003).

[165] R. J. Magyar and S. J. Tretiak, J. Chem. Theory Comput. **3**, 976 (2007).

[166] J. B. Krieger, Y. Li, and G. J. Iafrate, J. Phys. Rev. A **45**, 101 (1992).

[167] C. Legrand, E. Suraud, and P. G. Reinhard, J. Phys. B **35**, 1115 (2002).

[168] G. Vignale and W. Kohn, Phys. Rev. Lett.. **77**, 2037 (1996).

[169] J. Antony and S. Grimme, Phys. Chem. Chem. Phys. **8**, 5287 (2006).

[170] D. Y. Wang, Q. Ye, B. G. Li, and G. L. Zhang, Nat. Prod. Res. **17**, 365 (2003).

[171] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, J. Chem. Phys. **115**, 3540 (2001).

[172] D. jacquemin, E. A. Perpete, G. Scalmani, M. J. Frish, R. Kobayashi, and C. Adamo, J. Chem. Phys. **126**, 144105 (2007).

[173] T. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai, and K. Hirao, J. Chem. Phys. **120**, 8425 (2004).

[174] A. J. A. Aquino, M. Barbatti, and H. Lischka, Chem. Phys. Chem. **7**, 2089 (2006).

[175] R. Nifosi, P. Amat, and V. Tozzini, J. Comput. Chem. **28**, 2366 (2007).

[176] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).

[177] U. C. Singh and P. A. Kollman, J. Comput. Chem. **7**, 718 (1986).

[178] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).

[179] D. Bakowies and W. Thiel, J. Phys. Chem. **100**, 10580 (1996).

[180] P. Amara and M. J. Field, Theor. Chem. Acc. **109**, 43 (2003).

[181] D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. Chatfield, M. Hodoscek, and B. R. Brooks, J. Chem. Phys. **117**, 10534 (2002).

[182] H. Lin and D. G. Trulahr, J. Phys. Chem. A **109**, 3991 (2005).

[183] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, M. Ehara, K. Toyota, M. Hada, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, O. Kitao, H. Nakai, Y. Honda, H. Nakatsuji, X. Li, J. Knox, H. Hratchian, J. Cross, C. Adamo, J. Jaramillo, R. Cammi, C. Pomelli, R. Gomperts, R. E. Stratmann, J. Ochterski, P. Y. Ayala, K. Morokuma, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz,

I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, , and J. A. Pople, *Gaussian 01, Development Version (Revision B.01)*, Gaussian, Inc., Pittsburgh PA, 2001.

[184] J. Tomasi and M. Persico, Chem. Rev.(Washington, D. C.) **94**, 2027 (1994).

[185] C. J. Cramer and D. G. Truhlar, Chem. Rev. **99**, 2161 (1999).

[186] J. Tomasi, B. Mennucci, and R. Cammi, Chem. Rev. (Washington, D. C.) **105**, 2999 (2005).

[187] V. Barone and M. Cossi, J. Phys. Chem. A **102**, 1995 (1998).

[188] R. Cammi, S. Corni, B. Mennucci, and J. Tommasi, J. Chem. Phys. **122**, 104513 (2005).

[189] S. Corni, R. Cammin, B. Mennucci, and J. Tommasi, J. Chem. Phys. **123**, 134512 (2005).

[190] R. Improta, V. Barone, G. Scalmani, and M. J. Frisch, J. Chem. Phys. **125**, 54103 (2006).

[191] R. Improta, G. Scalmani, M. J. Frisch, and V. Barone, J. Chem. Phys. **127**, 74504 (2007).

[192] *Computational strategies for spectroscopy* (V. Barone, Ed, Wiley, 2012).

[193] G. Scalmani, M. J. Frisch, B. Mennucci, J. Tomasi, R. Cammi, and V. Barone, J. Chem. Phys. **124**, 094107 (2006).

[194] O. Shimomura, FEBS Lett. **104**, 220 (1979).

[195] L. Zang and J. Hermans, Proteins **24**, 433 (1996).

[196] M. Ormo, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien, and S. J. Remington, Science **273**, 1392 (1996).

[197] J. van Thor, T. Gensch, K. J. Hellingwerf, and L. N. Johnson, Nat. Struct. Biol. **9**, 37 (2002).

[198] A. Shinobu, J. P. Gottfried, A. J. Schierbeek, and N. Agmon, J. Am. Chem. Soc. **132**, 11093 (2010).

[199] A. D. Becke, Phys. Rev. **38**, 3098 (1988).

[200] Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai, and K. Hirao, J. Chem. Phys. **120**, 8425 (2004).

[201] O. A. Vydrov, J. Heyd, A. Krukau, and G. E. Scuseria, J. Chem. Phys. **125**, 074106 (2006).

[202] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[203] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **78**, 1396 (1997).

[204] R. Ditchfield, W. J. Hehre, and J. A. Pople, Can. J. Phys. **54**, 724 (1971).

[205] W. J. Hehre, R. Ditchfield, and J. A. Pople, Can. J. Phys. **56**, 2257 (1972).

[206] P. C. Hariharan and J. A. Pople, Theor. Chem. Acc. **28**, 213 (1973).

[207] T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. v. R. Schleyer, J. Comptu. Chem. **4**, 294 (1983).

[208] M. J. Frisch, J. A. Pople, and J. S. Binkley, J. Chem. Phys. **80**, 3265 (1984).

[209] T. Vreven, K. Morokuma, O. Farkas, H. B. Schlegel, and M. J. Frisch, J. Comput. Chem. **24**, 760 (2003).

[210] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer,

T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).

[211] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[212] B. Mennucci, C. Chiappelli, C. A. Guido, R. Cammi, and J. Tommasi, J. Phys. Chem. A **113**, 3009 (2009).

[213] S. Grimme, J. Comput. Chem. **27**, 1787 (2006).

[214] D. Jacquemin, V. Wathelet, E. A. Perpte, and C. Adamo, J. Chem. Theory Comput. **5**, 2420 (2009).

[215] X. Li, L. W. Chung, H. Mizuno, A. Miyawaki, and K. Morokuma, J. Phys. Chem. B **114**, 1114 (2010).

[216] X. Lopez, M. A. L. Marques, A. Castro, and A. Rubio, J. Am. Chem. Soc. **127**, 12329 (2005).

[217] D. Xie and J. Zeng, J. Comput. Chem. **26**, 1487 (2005).

[218] L. Zhang, D. Xie, and J. Zeng, J. Theor. Comput. Chem **1**, 375 (2006).

[219] K. Fukui, Acc. Chem. Res. **14**, 363 (1981).

[220] H. P. Hratchian and H. B. Schlegel, *Theory and Applications of Computational Chemistry: The First 40 Years* (C. E. Dykstra and G. Frenking and K. S. Kim and G. Scuseria , Ed., Elsevier Science, Amsterdam, pp.195-249, 2005).

[221] O. Vendrell, R. Gelabert, M. Moreno, and J. M. Lluch, J. Chem. Theory Comput. **4**, 1138 (2008).

[222] C. Fang, R. Frontiera, R. Tran, and R. A. Mathies, Nature **462**, 200 (2009).

[223] M. D. Donato, L. J. G. W. van Wilderen, I. H. M. V. Stokkum, T. C. Stuart, J. T. M. Kennis, K. J. Hellingwerf, R. van Grondelle, and M. L. Groota, Phys. Chem. Chem. Phys. **13**, 16295 (2011).

# ...l'esperienza è quello che l'uomo fa con ciò che gli accade

# Il mio Grazie...

*L'amore è paziente,*
*è benigno l'amore,*
*non è invidioso l'amore,*
*non si vanta, non si gonfia,*
*non manca di rispetto,*
*non cerca il suo interesse,*
*non si adira,*
*non tiene conto del male ricevuto,*
*non gode dell'ingiustizia,*
*ma si compiace della verità.*
*Tutto copre,*
*tutto crede,*
*tutto spera,*
*tutto sopporta.*
*L'amore non avrà mai fine.*
*(1 Cor 13, 4-8)*

**A tutti coloro che con il loro Amore danno gioia, sostegno, pienezza e senso alle mie giornate.**