One-, Two-, and Three-Photon Absorption Studies of Fluorescent Proteins and Their Chromophores Using Quantum Mechanical and Quantum Mechanical/Molecular Mechanical Approaches

by

Maria Rossano Tapia

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry
University of Alberta

Abstract

In the present thesis, the multi-photon absorption features of different Aequorea fluorescent proteins (FPs) are explored computationally using quantum mechanical and combined quantum mechanical/molecular mechanical approaches. In chapters two and three of the present thesis, we evaluate the performance of the semi-empirical time-dependent tight-binding density functional theory and its long-range corrected version, TD-DFTB and LC-TD-DFTB, respectively, in the computation of the two-photon absorption (2PA) cross-sections (σ^{2PA}) for a set of canonical and non-canonical FPs chromophores previously studied using time-dependent density functional theory (TD-DFT). From these investigations, we found that through the two-level model (2LM), TD-DFTB and LC-TD-DFTB lead to 2PA cross-sections that largely deviate from the TD-DFT values obtained in previous investigations. In comparison with TD-CAM-B3LYP, a common DFT method used in the computation of σ^{2PA}s of FPs and their chromophores, the σ^{2PA}s obtained using LC-TD-DFTB are up to 125 GM larger. Such a deviation is mainly due to an overestimation of the excited state permanent dipole moment, on which the 2PA cross-sections are strongly dependent. Chapter four explores the computation of the three-photon absorption of the intrinsic probe serotonin, as well as the dyes, fluorescein and rhodamine 6g, which are used as references in multi-photon absorption studies. The results obtained here, using a series of basis sets along with the CAM-B3LYP functional, are in reasonably good agreement with experiment. Furthermore, we found that the CAM-B3LYP/aug-cc-pVDZ method is the best option as it exhibits the best accuracy with respect to experimental results to computing cost ratio among the tested basis sets. In the case of rhodamine 6g, the three-photon absorption cross-section we obtained using the CAM-B3LYP/aug-ccpvDZ method in vacuum, 4.86×10^{-81} cm^6 s^2 photon^{-2}, is close to the average measurement of different

experimental sources, 6.0×10^{-81} cm^6 s^2 photon^{-2}. Experimental and computational investigations have concluded that the MPA of a given FP depends on the interaction between its chromophore and the environment, which is comprised of the protein and other moieties, such as water molecules, if present. Therefore, a given FP chromophore can exhibit a 2PA cross-section of either 10 GM or 80 GM, depending on the environment in which it is embedded. In this context, chapter five discusses the 2PA of a set of non-canonical chromophores while we consider the environmental effects through the DsRed protein barrel. The σ^{2PA}s obtained for these systems using the polarizable quantum mechanical/molecular mechanical (QM/MM) approach are about four times smaller than results reported without taking into account the protein effects (in vacuum). These results reinforce the idea of including the environment effects in the computation of multi-photon absorption properties of FPs to get a more "realistic" view of their photophysical properties. Finally, chapter six assesses some of the possible paths to engineer FPs with enhanced 2PA in comparison to the existent FPs. In this chapter, the σ^{2PA} of the chromophore surrounded by a selected set of neighbouring amino acids is computed using the polarizable QM/MM approach. This investigation is not conclusive, however, as from the results obtained, it was determined that some neighbouring amino acids can enhance the 2PA of the red fluorescent protein chromophore up to 130%. The examination of the roles of specific amino acids provides information about possible positions that can be mutated in order to engineer FPs with enhanced multi-photon absorption in comparison to the existing ones. Overall, the results presented here will be useful as benchmark to those wanting to study the MPA features of FPs.

Preface

The candidate confirms that the work submitted is her own, except where work that has formed part of jointly-authored publications has been included.

The candidate created this thesis under the scientific and editorial guidance of Prof. Alexander Brown. Dr. Anna Jordan, Thesis Editor in Residence at the Department of Chemistry of the University of Alberta, contributed with editorial corrections to all chapters. Other contributors are mentioned below.

Chapter 1 is an adaptation of the focus review published as Rossano-Tapia, M. and Brown, A. "Quantum Mechanical/Molecular Mechanical Studies of Photophysical Properties of Fluorescent Proteins." *WIREs Comput Mol Sci.* DOI: 10.1002/wcms.1557. The candidate wrote the first manuscript and refined the text based on the peer-review and Prof. Brown's feedback. Prof. A. Brown contributed with editorial corrections and scientific advice.

Chapter 2 is an adaptation of the paper first published as Rossano-Tapia, M. and Brown, A. "Determination of Two-Photon-Absorption Cross Sections Using Time-Dependent Density Functional Theory Tight Binding: Application to Fluorescent Protein Chromophores". *J. Chem. Theory Comput.* **2019**, 15, 3153–3161. Prof. A. Brown conceived the original idea of the project and provided scientific advice to the candidate for the performance of the computations. The candidate performed all the computations, wrote the first manuscript, and co-edited in collaboration with Prof. A. Brown a refined version of the manuscript.

Chapter 3 is based on the outcomes found in Chapter 2, and it evaluates the performance of a long-range corrected version of TD-DFTB, LC-TD-DFTB. The project was conceived originally by the candidate and Prof. A. Brown. The candidate performed all the computations, collected the data, and wrote the draft of the chapter. Prof. A Brown provided editorial feedback.

Chapter 4 is an adaptation of the manuscript submitted for publication "Three-photon Absorption Cross-sections of Serotonin, Fluorescein, Rhodamine 6G, and Some Fluorescent Protein Chromophores." The project was conceived by the candidate and Prof. A. Brown. The computations were carried out by Marina Mondrusova, Mitacs Summer Intern in the Prof. A. Brown's research group in 2019, under the supervision of the candidate, and by the candidate herself. The candidate collected the data, wrote the first draft of the chapter, and refined the manuscript. Prof. A. Brown provided editorial and scientific advice.

Chapter 5 is an adaptation of the work originally published as Rossano-Tapia, M.; Olsen, J. M. H.; and Brown, A. "Two-Photon Absorption Cross-Sections in Fluorescent

Proteins Containing Non-canonical Chromophores Using Polarizable QM/MM." *Front. Mol. Biosci.* **2020**,*7*, 111. The work of this chapter was performed in collaboration with Dr. Jógvan Magnus Haugaard Olsen from the Hylleraas Centre for Quantum Molecular Sciences and the Department of Chemistry at UiT The Arctic University of Norway, Tromsø, Norway. The candidate and Prof. A. Brown conceived the original idea of the project. Prof. A Brown provided editorial feedback to the manuscript. The candidate performed all the computations under the supervision of Dr. J.M.H. Olsen and wrote the first draft of the manuscript. Dr. J.M.H. Olsen contributed to the creation and refinement of the manuscript.

Chapter 6 is motivated by the outcomes discussed in Chapters 4 and 5. The original idea was conceived by the candidate and Prof. A. Brown. The candidate performed all the computations and wrote the draft of the chapter. Prof. A. Brown provided scientific advice and editorial comments.

Chapter 7 summarizes the conclusions found in each of the chapters and the overall content of the thesis. This chapter was written by the candidate and editorially commented by Prof. A. Brown.

The computations employed in the elaboration of the present thesis were provided by different High Performance Computing infrastructures. For chapters 2, 3, 5, and 6, by Compute Canada-Calcul Canada (https://www.computecanada.ca/) as well as by WestGrid (https://www.westgrid.ca/), and for chapter 4, by Sigma2 through the Linux Cluster Stallo at the Arctic University in Norway (https://www.sigma2.no/systems) along with Compute Canada-Calcul Canada. In the case of Compute Canada, the computations were distributed between the CPUs in the Graham and Cedar clusters.

To my parents, Javier and Hilda, and my dearest Lola Arce.

Acknowledgements

I would like to thank:

- My supervisor Prof. Alexander Brown, for his academic guidance throughout my Ph.D. program.

- My parents and family for their love, support, and patience.

- My partner, Alex Gzyl, for his enormous love and support. Thank you for always helping me notice the bright side of life.

- Prof. Dolores Arce, who has been my academic mentor over the last 13 years and, most importantly, a beloved friend and part of my family. You will always be in my heart, Lola.

- Luis Aguilar, Franz Martinez, and Pablo Carpio for their friendship and unconditional support.

- Anna Jordan for her time and meticulous editing of the present thesis, but most importantly, for her kindness, support, and advice.

- The students I had the opportunity to co-mentor and ultimately collaborate with during my Ph.D. program: Xueting Ni, Yidan Xing, Kwami Aku-Dominguez, Megan Joy, and Marina Mondrusova.

- Shyam Parshotam and Antoine Aerts for all the good times and food we shared in Edmonton.

- Compute Canada-Calcul Canada and WestGrid for the technical support and the computational resources provided.

Contents

List of Tables

xiv

List of Figures

Acronyms

1PA One-photon absorption.

2LM Two-level model.

2PA Two-photon absorption.

3PA Three-photon absorption.

5-HT 5-hydroxytryptamine or serotonin.

BFP Blue fluorescent protein.

CASPT2 Complete active-space second-order perturbation theory.

CC2 Coupled cluster singles and doubles.

DFT Density functional theory.

DFTB Tight-binding density functional theory.

eCFP Enhanced cyan fluorescent protein.

EE Electrostatic embedding.

EEF Effective external field.

eGFP Enhanced green fluorescent protein.

ESO Electronic spill-out.

FP Fluorescent protein.

FSM Few-state model.

GFP Green fluorescent protein.

GM Göppert-Mayer units.

HWHM Half width at half maximum broadening factor.

LC-TD-DFTB Long-range corrected TD-DFTB.

MD Molecular dynamics.

MM Molecular mechanical.

MPA Multi-photon absorption.

MPM Multi-photon microscopy.

nCC Non-canonical chromophore.

ONIOM "Our own *n*-layered integrated molecular orbital and molecular mechanics".

PCM Polarized continuum model.

PDB Protein data bank identifier.

PE Polarizable embedding.

QM Quantum Mechanical.

QRT Quadratic-response theory.

R6G Rhodamine 6G.

RFP Red fluorescent protein.

SOS Sum-over states approximation.

TD-DFT Time-dependent density functional theory.

TD-DFTB Time-dependent tight-binding density functional theory.

UFF Universal force field.

wtGFP Wild-type green fluorescent protein.

YFP Yellow fluorescent protein.

Chapter 1

Quantum Mechanical/Molecular Mechanical Studies of Photophysical Properties of Fluorescent Proteins[1]

1.1 Introduction

In the jellyfish *Aequorea victoria* and other sea species, like anemones of the Anthozoa class, fluorescent proteins (FPs) are responsible for their vibrant light emission.[1–3] In the 1960s, the green fluorescent protein (GFP) was isolated from *A. victoria*.[1] Nowadays, engineered GFP derivatives and FPs in general are used in cell and cell organelle tagging, in sensing of proteins' and living systems' activity, and in clinical imaging, among other applications.[4] While FPs are not the only resource in clinical imaging, they are considered to be useful particularly in brain imaging.[5] Their use (e.g., in brain imaging) and exploration are supported by many aspects, ranging from their structural to their photophysical properties, as depicted in Figure 1.1.
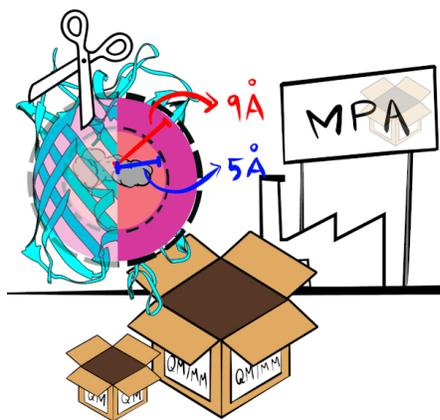
Figure 1.1. Depiction of the study of photophysical (multi-photon absorption) properties of fluorescent proteins using quantum mechanical/molecular mechanical (QM/MM) methods.

---

[1]This chapter has been copied and/or adapted from the publication: Rossano-Tapia, M. and Brown, A. Quantum Mechanical/Molecular Mechanical Studies of Photophysical Properties of Fluorescent Proteins *WIREs Comput Mol Sci.* DOI: 10.1002/wcms.1557

FPs can be decomposed into two complementary pieces, the protein barrel structure and the chromophore (see Figure 1.2). The chromophore, which originates from the autocyclization of three amino acids, is located in the core of the protein.[6] Thus, from the structural perspective, FPs are ideal imaging instruments since they do not require additional molecules (exogenous agents) to develop their fluorescence.
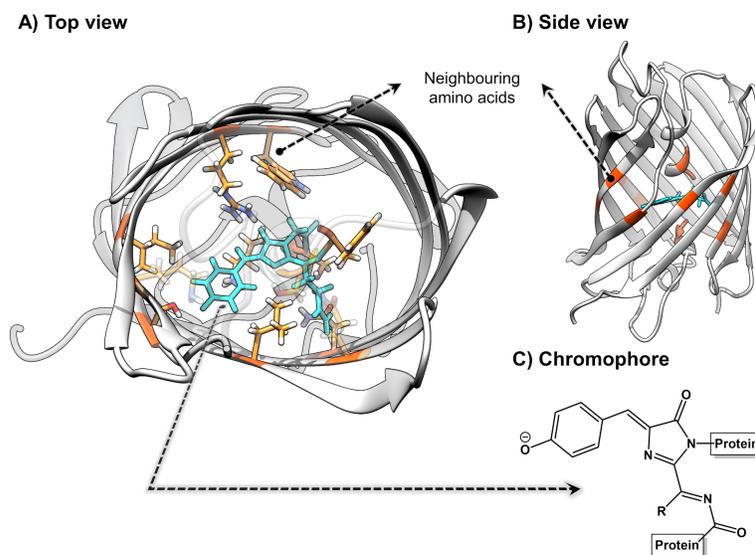
Figure 1.2. Fluorescent proteins (red fluorescent protein DsRed, protein data bank (PDB) identifier 1ZGO,[7] in this particular case) main components; A-B) the protein barrel and C) the chromophore embedded within it. The neighbouring amino acids (within 3 Å) from the chromophore are shown in orange.

Both the protein barrel ("the environment") and the chemical identity of the chromophore contribute to the absorption and emission properties of the FP. Thus, any modifications to the properties of an FP derive from the direct and indirect effects of both the chromophore and the protein barrel. Many FPs have been discovered in nature, and mutants of different types of FPs have been engineered experimentally. Altogether, they are now part of a wide palette of colors that range across the whole visible spectrum. Among the different photophysical properties of FPs, multi-photon absorption (MPA), particularly two-photon absorption (2PA), has gained great relevance and has been explored widely using experimental[5] and computational means.[8–19] 2PA microscopy has notable advantages over its one-photon absorption (1PA) counterpart, including better

focus, less auto-bleaching, and a lower risk of photo-damage in the tissue in the context of in vivo imaging. Importantly, via 2PA microscopy, it is possible to excite FPs with short-wavelength 1PA within the so-called clinical window ($\sim$650–900 nm[20]), which means that it is not necessary for the FP to have a strongly red-shifted 1PA. The clinical window is the far-red zone of the electromagnetic spectrum, where the absorption of water decays and where the risk of photo-damage in the tissue is lower in comparison to that from the use of shorter excitation wavelengths.

Multi-photon absorption (MPA) consists of the simultaneous absorption of two or more photons of the same or different energy. The simultaneous absorption of two photons with the same energy (and same wavelength) is referred to as degenerate two-photon absorption (2PA). In the present thesis, only degenerate processes are studied. Two-photon absorption studies are complementary to one-photon absorption (1PA) ones, as one cannot infer 2PA from the 1PA because they obey different selection rules and depend on different molecular properties. For centrosymmetric molecules, 1PA occurs between states of different parity, however, in the case of 2PA both states must have the same parity.[21] In the case of anionic non-centrosymmetric FP chromophores, their 1PA spectrum is reasonably reflected in the corresponding 2PA spectrum, however, the 2PA spectrum can be noticeably blue-shifted due to a vibrational enhancement.[5] Overall, the 2PA properties of fluorescent proteins cannot be strictly inferred from their 1PA, and it is why 2PA must be measured (or computed) directly.

The computational investigations on the photophysics of FPs and their chromophores range over a wide variety of (i) FPs including the green, red, yellow, and the fruit series; and (ii) computational methodologies, e.g., pure quantum mechanical (QM) methods, classical molecular dynamics (MD), and hybrid quantum mechanical/molecular mechanical (QM/MM) approaches.[22] Most of the QM/MM investigations on the MPA features of FPs referenced in the present introduction and later chapters employ polarizable embedding (PE) approaches. In particular, many of these studies focus on the one

developed by Olsen et al.,[23] which is based on the induced-dipole polarization model, where the environment is represented by multipoles (that can be up to fifth-order) and polarizabilities.[23–27] The Olsen et al. formulation[23] is based upon time-dependent density functional theory (TD-DFT), which makes the method distinctively capable of determining response properties, such as vertical excitation energies (VEEs) and multi-photon absorption features. The total energy in the mechanical (also called classical) and electrostatic embedding (EE) approaches is modeled, within the additive scheme, by the general expression[28]

$$E_{\text{total}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM}-\text{MM}}. \quad (1.1)$$

In the case of mechanical embedding, the interaction energy $E_{\text{QM}-\text{MM}}$ in Equation 1.1 does not allow any kind of polarization between the QM and MM regions, whereas in the case of EE, the electron density of the QM region is polarized by the MM environment. In the models where polarization occurs from the MM environment to the QM region and vice versa, simply called PE schemes or QM/pol-MM, the polarization of the MM region can be addressed through Drude oscillators (also referred to as "charge-on-a-spring"),[29] fluctuating charges,[30] or induced dipole moments. Within the induced dipole moments scheme, the interaction energy contribution to the total energy, E_{total}, includes now the effects of a PE potential as

$$E_{\text{total}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{PE}}. \quad (1.2)$$

In Equation 1.2, E_{PE} is the sum of the interaction energies of the QM region and the environment permanent and induced electrostatics, E_{es} and E_{ind}, respectively. The energy from the induced electrostatics depends on the induced dipole moments, ($\boldsymbol{\mu}$),

$$E_{\text{ind}} \propto \mu \quad (1.3)$$

4

in which μ is given by the product

$$\mu(\mathbf{F}) = \mathbf{BF} \qquad (1.4)$$

In Equation 1.4, $\mathbf{F}$ represents the different electric fields and $\mathbf{B}$, the N$\times$N matrix of the polarizable sites.[23,31–33] Notice that the above equations are oversimplified, as the main purpose of the present introduction is to discuss the outcomes of those methodologies in which QM/MM approaches have been employed and not about the methods per se. The theoretical background of the aforementioned polarizable QM/MM approaches already has been discussed extensively,[34] and will not be repeated here. Furthermore, the outcomes and pitfalls of the induced-dipole polarizable approaches have been addressed recently by Bondanza et al.[35] and Loco et al.[36]

Other aspects of FPs like the formation of the chromophore and the deprotonation/protonation of the chromophore (such as in GFPs) are important, but the primary focus is on photophysical (absorption and emission) properties of FPs. In a similar fashion, FPs' photophysical aspects like Förster resonance energy transfer (FRET), excited state relaxation dynamics via conical intersections, and deactivation of the excited state are not discussed in this chapter nor in the thesis overall. For the latter, the reader may refer to the work previously carried out on the GFP by Martin et al.[37] and on reversibly switchable fluorescent proteins by Smyrnova et al.[38] as well as Morozov and Groenhof.[39]

The present introduction and thesis will review previous research that will be useful for those who are interested in carrying out computations on FPs and exploring the environmental effects on their photophysical properties. Several reviews focused on QM/MM approaches, including FPs, may be useful to the reader. For example, Senn and Thiel[40] discuss the use of QM/MM approaches for the exploration of enzymes. Some key aspects in the preparation of a QM/MM computation, such as the size of the QM region or the choice of the QM level of theory, are provided by Pedraza-González et al.[41] on light-responsive systems, like rhodopsins, and Navizet[42] on bioluminiscent

systems. Bravaya et al.[43] specifically examine FPs and their absorption properties, particularly 1PA, either within QM/MM or in vacuum approaches. Neugebauer addresses the use of QM/MM methodologies specifically applied to the study of GFP,[44] as do Nemukhin and Grigorenko.[45] In addition, Nifosì and Tozzini present a compilation of both computational and experimental work on one- and two-photon absorption (1PA and 2PA, respectively) and emissive properties for different types of green and red fluorescent protein chromophores in vacuum, solvent, and influenced by the action of surrounding residues.[10] Photochemical reactions and other photo-induced processes in FPs are described in the review written by Acharya et al.,[46] although QM/MM schemes are not the primary focus of the methods cited.

1.2   QM/MM Schemes Used in the Study of Fluorescent Proteins

Numerous computational studies have been performed on FPs. As mentioned previously, the employed methods and models are diverse and include QM, MD, QM/MM (both nonpolarizable and polarizable MM) or QM/MM-MD strategies. Pure QM studies typically have been used in the study of the FP chromophores and small clusters constructed from the chromophore and neighbouring residues to include the effects of the environment. Many examples of the QM study of 1PA and 2PA properties for different isolated canonical and non-canonical FP chromophores appear in the literature.[8,9,11,13–15,17,47–52]

The 1PA properties, like VEEs, determined for chromophores in vacuum have been shown to be in reasonable agreement with experimental absorption measurements;[53,54] comparison also has been facilitated through the availability of experimental measurements on isolated chromophores[53–56] or denatured FPs.[8] On the other hand, the comparison of computationally determined two-photon absorption cross-sections (σ^{2PA}) for isolated chromophores to experimental measurements,[53,57] almost exclusively on the full FP[57], has proven more challenging due to both variability in experimental measurements and the strong role played by the environment (see Figure 1.3 and discussion below). Experimental results include an important piece of the puzzle, the protein and, with it,

6

important environmental effects.
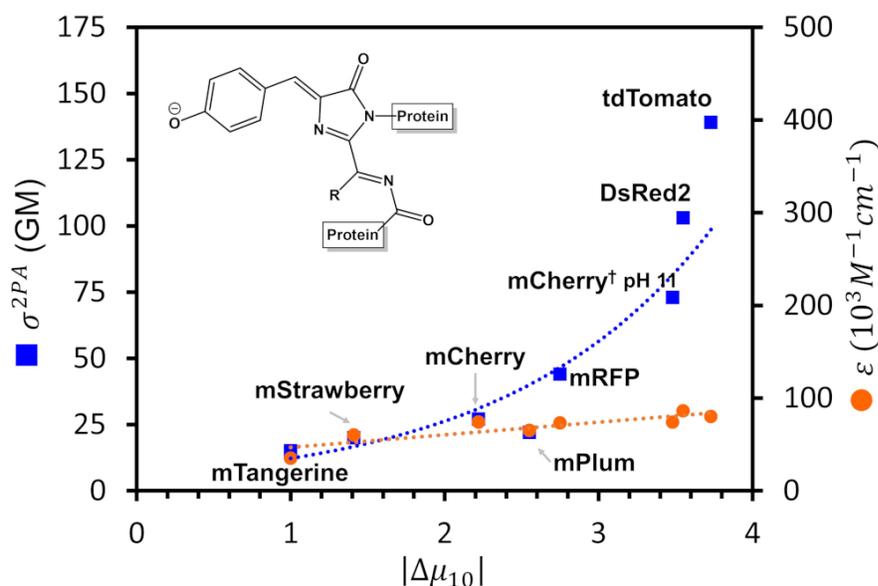
Figure 1.3. Depiction of the strong dependence of σ^{2PA} (blue squares) versus the corresponding 1PA (red dots) as a function of the permanent electric dipole moment difference between the ground and the first excited state $|\Delta\mu_{10}|$.[58] The inset illustrates a simplified version of the RFP chromophore, which is the primary chromophore in all FPs presented. R– corresponds to –CH_2–CH_2–CO–NH_2 in the DsRed FP whereas connections to the protein are indicated as "Protein". [Based on the experimental results and the figure previously reported by Drobizhev et al.[5]]

Experimentally, it has been demonstrated that the protein environment and other non-protein moieties, like water molecules, play a significant role in the MPA properties of FPs. Indeed, the role of the protein environment for a given FP's MPA is larger than for its corresponding 1PA. As an example, the behaviour displayed by the so-called fruit series, which is a group of FPs derived from mutations in the protein sequence of the RFP monomer (mRFP1), can be considered. Figure 1.3 (adapted from experimental data available from Drobizhev et al.[5]) illustrates for a selected set of fruit FPs the strong dependence of the σ^{2PA} on changes in the difference in the permanent dipole moments between the excited and ground states in each FP. The chromophore of the fruit series FPs[59] (fundamentally) is the same as the parent red fluorescent protein (RFP) (see structure provided in Figure 1.3). As observed in Figure 1.3, the effect of the environment on the 1PA cross-sections is not as strong as it is for the 2PA. Nevertheless,

the environmental effects should never be taken for granted in 1PA computations. Based on the above, it is fair to conclude that the gap between the computational data and experimental measurements lies in the inclusion of the environmental effects. However, it is important to mention that within the experimentally measured σ^{2PA}, persistent discrepancies exist due to, most likely, experimental errors.[5] As discussed by Drobizhev et al., such deviations can be of about two orders of magnitude, e.g., the reported σ^{2PA} values for the enhanced green fluorescent protein (EGFP) vary from $\sim$1 to 300 GM.

The above discussion does not suggest that studies on isolated chromophores should be avoided, indeed, they have been shown to be quite useful as first-order approximations to the full FP and for exploratory studies. More importantly, studies on the isolated chromophores have been used for a deeper understanding of the photophysical properties of FPs[11,15] and, in particular for MPA, by using methods that are not implemented with QM/MM (e.g., equation of motion coupled-cluster, EOM-CCSD) or that are too computationally expensive to be used in QM/MM approaches for the computation of non-linear response properties. Furthermore, studies on the isolated chromophores can shed light on the shortcomings of the QM methods typically employed in the computation of 2PA features, e.g., overestimation of energies, inaccurate prediction of excited state dipole moments (as in the case of tight-binding TD-DFT discussed in Chapters 2 and 3), and overestimation of the two-photon absorption cross-sections, σ^{2PA} (Equation 1.5),

$$\sigma^{2PA} = \frac{N\pi^2 a_0^5 \alpha}{c_0} \frac{\omega^2}{\Gamma} \delta^{2PA} \ . \tag{1.5}$$

In Equation 1.5, ω is half the excitation energy, whereas δ^{2PA} is the 2PA transition moment probability. To compare with experiment, the integer value N and the half width at half maximum (HWHM) broadening factor Γ are set to 4 and 0.1 eV, respectively.[60] The role of each one of the latter cross-section components as well as the transition dipole moments has been a matter of discussion in other contributions[15,60] and will be discussed further in Chapters 2 and 3.[61] It is important to bear in mind the limitations

of the models of isolated chromophores, e.g., 2PA results using in vacuum models can exhibit significant discrepancies with respect to those obtained for the chromophore embedded in the protein matrix. In cases like the fruit series, where σ^{2PA} varies by almost two orders of magnitude (see Figure 1.3), it would be impossible to address their 2PA features simply by looking at the isolated chromophore as it is essentially the same in each FP.

Including the environmental effects in the computation of response properties comes with many caveats. One of them is the size of the chromophore–protein system, which in the case of the DsRed RFP (PDB 1ZGO[7]) is >1800 non-hydrogen atoms. Systems of this size, where excited electronic states must be determined, are not accessible to QM methods using present computing capacity and, thus, have been studied through a variety of QM/MM schemes. 1PA properties for the FPs' chromophores have been computed within some of the simplest models for treating an environment like the self consistent field polarizable continuum model (PCM). However, from construction, these models are insufficient to capture the intrinsic anisotropy of the typical protein environment in the computation of second- and higher-order response properties. Indeed, VEEs as well as 1PA and 2PA cross-sections obtained for isolated chromophores (i.e., in vacuum) almost are unaltered upon the inclusion of solvent using the PCM (water). Therefore, they (PCM-like methods) are not a means to improve the results obtained for the isolated chromophore in vacuum[9,12,14,52,62] when the goal is to compare them to properties of the FP.

Other studies have used constrained models, often referred to as "clusters", containing the chromophore and a select set of neighbouring covalently and non-covalently bonded amino acids, as well as water molecules. For these models, both pure QM[16,63–65] and QM/MM[66–68] methods have been employed. Within the QM/MM studies, a diverse selection of FPs have had 1PA features computed, e.g., GFP,[66] mStrawberry and mCherry from the fruit series,[67] as well as mPlum.[68] Cluster models have proven to be useful in

providing insights on the influence of the chromophore–protein interactions through the closest residues on the FP absorption properties. However, how large or small the chosen cluster is can have a significant impact on the obtained results.[18]

The following example highlights the impact that size has on the results when cluster models are employed. Kaila et al. studied fully quantum mechanically the VEE of GFP clusters using a reduced virtual space version of CC2.[69] Their cluster models included the GFP chromophore, either neutral or anionic, surrounded by nine protein residues plus four water molecules (see Figure 1.4). Besides the good agreement of their results with experimental reports, other interesting findings arose, e.g., electrostatic effects had the largest contribution to excitation energies over the steric ones, which suggested the size of their cluster model was adequate to take into account environmental contributions in the description of 1PA properties of the selected GFP models. However, Schwabe et al. subsequently determined that variations in the size of the above cluster models, along with the introduction of explicit interactions between the chromophore and residue Thr203 (missing in the original cluster model) lead to changes in the VEE.[70] Thus, the size of Kaila et al.'s cluster model turns out to not be ideal but circumstantially ideal due to fortuitous agreement with experiment. Hence, cluster models used in the determination of response properties in FPs need to be designed and implemented carefully.
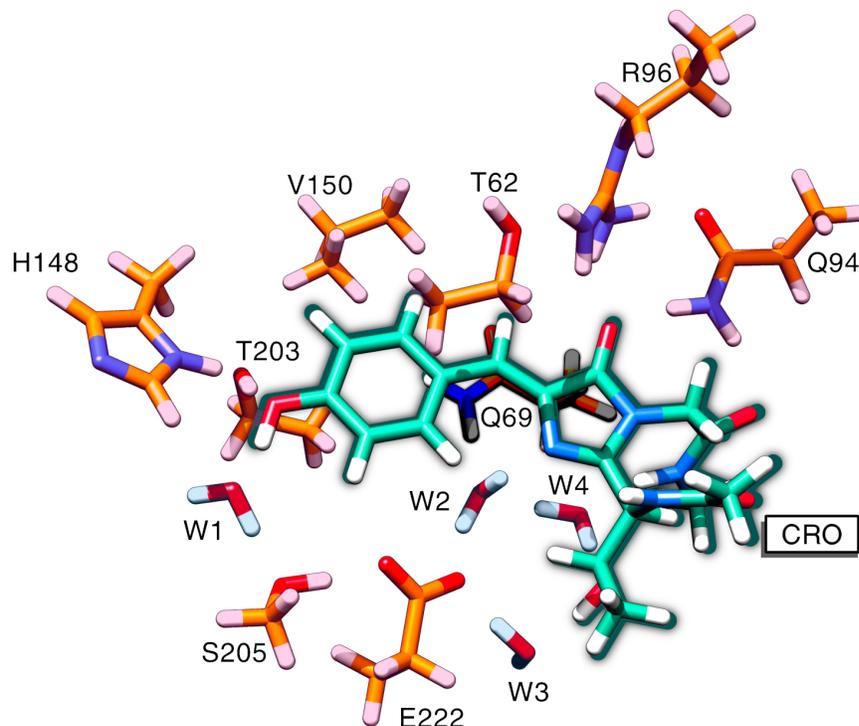
Figure 1.4. Cluster used by Kaila et al. for studying GFP,[69] which included nine truncated residues and four water molecules (W) around the chromophore (CRO). Labeling is almost entirely based on the cluster model of Schwabe et al.[70]

In addition to the chromophore plus selected residues (and, if needed, crystal waters) the entire FP (and, if present, crystal waters) can be modeled using QM/MM approaches. Among many investigations, VEE and the GFP are the most common property and protein, respectively, generally studied through QM/nonpol-MM.[71–73] Hasegawa et al. computed the excitation and emission energies for the green, blue, and cyan FPs using the symmetry-adapted cluster-configuration interaction (SAC-CI)/Amber96 scheme.[72] Sanchez-Garcia et al. computed the VEE and the absorption spectrum, in reasonable agreement with experimental measurements, for the GFP and red fluorescent protein DsRed.M1 using DFT/MRCI, TD-B3LYP, and OM2/MRCI.[73] In some cases, after the inclusion of the protein environment within a certain QM/MM model, it has been observed that the computed excitation energies are not affected to the same extent as other properties, such as the detachment energy.[66] The list above is not complete but introductory, as many more examples will be provided within the rest of the present

chapter.

Besides the level of theory employed in the QM region, the quality and accuracy of the properties computed through QM/MM approaches will depend on different aspects. Some of them are the following:

- The type of embedding of the QM/MM approach, i.e., point charges, as in electrostatic embedding (EE) or a polarizable potential, as in the so-called PE approaches, QM/pol-MM.[70,74,75] From construction, fully polarizable embedding models are able to capture the perturbation in the environment caused by the QM–environment interaction upon the excitation of the QM region. Current PE approaches have pitfalls that undermine the accuracy of the computations; nonetheless, they are more adequate than EE to compute VEE or MPA. Examples of how PE outperforms EE in the computation of MPA features are now more common in the literature. In fact, PE now is considered to be the gold standard in the study of complex systems, as discussed by Bondanza et al.[35] As an example, Filippi et al. found that the use of a non-polarizable force field[74] is among the reasons behind the incorrect description of the shift between the anionic and neutral forms of the GFP absorption energies.[74] Over the following paragraphs, more examples of QM/pol-MM studies of FPs will be discussed.

- The determination of an optimal size for the QM subsystem.[73,76–78] In this regard, Grabarek and Andruniów systematically investigated the 1PA and 2PA spectra for different QM region sizes in QM/pol-MM models of GFP and Citrine.[18] For the systems employed by Grabarek and Andruniów, by including in the QM region the amino acid residues and water molecules found within 3–3.5 Å from the chromophore, the convergence of the σ^{2PA}s is achieved for EE. Whereas, in the case of using a PE for studying GFP models, it is necessary to include in the QM region the chromophore plus residues R96 and E222 (as well as Y203 for YFP-type models) and water molecules H-bonded to the chromophore.

Another example is the work carried out by Filippi et al.[74] in their analysis of two possible GFP protonated states using the following three different QM/MM schemes: CASPT2/Amber99, TD-DFT/Amber99, and QMC/Amber03. The authors observed that when utilizing TD-DFT, spurious charge-transfer effects can arise if the QM region is expanded beyond the chromophore.[74]

- Dynamical contributions introduced by MD.[23,67,73,77,79–83] For the case of GFP, the excitation energy obtained using the crystal structure of GFP can be ~ 0.04 eV larger than that obtained from QM/MM-MD computations.[77] For the case of MPA strengths (δ^{MPA}), the standard deviation of QM/MM-MD results can be about the same order of magnitude as δ^{MPA}.[26]

- The description of the electron density in the QM and QM/MM boundary region, which depends on the choice of QM/MM approach. Particularly, the lack of Pauli repulsion in the QM region wave function in additive QM/MM methods causes an electron spill-out (ESO) in the frontier between the subsystems.[19,27,84] In practice, the ESO artifact appears as a spurious concentration of electron density in the MM region, which worsens in part by the use of highly diffuse basis sets. Marefat Khah et al. tested a solution to this problem on a diverse set of chromophores and embedding environments.[84] Within polarizable QM/MM, the authors introduced the missing Pauli repulsion contribution by employing one-center pseudopotentials to approximate the core and valence electrons of the MM region. The strategy proposed by Marefat Khah et al. prevents the leakage of electron density at a reasonable computational cost in the investigated systems.[84] Excitation energies computed in this way are more accurate than those obtained using a previous PE QM/MM formulation.

Within the context of QM/pol-MM approaches and FPs' photophysics, the formulation proposed by Olsen et al.[23–27] has been explored extensively in the determination of

linear and/or non-linear response properties for different FPs, including wtGFP, eGFP, DsRed, BFP, YFP, and eCFP.[26,70,77,78,83,85,86] Interestingly, studies of MPA have been extended up to four-photon absorption.[26] Results obtained using this QM/pol-MM model are improved with respect to QM/nonpol-MM approaches.[49,85] Using the aforementioned QM/pol-MM model, σ^{2PA}s have been determined for two main groups of FPs listed in Table 1.1, along with some of the experimental data to which computational results have been compared. Bear in mind that Table 1.1 excludes those FPs for which only 1PA has been studied using this PE approach; for 1PA studies using this PE model see Beerepoot et al.[77,83] and Marefat Khah et al.[84]

Table 1.1. One-photon Vertical Excitation Energies (VEE) and Two-Photon Absorption Cross-Sections (σ^{2PA}) Computed Using Polarizable Embedding. In Some Cases Experimental Values, to Which the Original Theoretical Reports Were Compared, are Given in Parenthesis

FP	VEE (eV)	σ^{2PA} (GM)
aGFP	3.01[26(a)], 2.61[85(a)](2.63[87]), 3.10†, 3.03*, [18(b)]	4.4[26], 635[85](5[88]), 3.1†, 2.7*, [18(b)]
nGFP	3.51[26], 3.06[85](3.04[87]). 3.26[†(c)], 3.22*, [18(b)]	1.8[26], 53[85](12[88]), 0.7†, 1.1*, [18(b)]
DsRed	2.826[86(a)](2.22[89])	105.9[86](96[58])

Note: Data for wtGFP and its mutants are grouped as GFP.
[a] TD-CAM-B3LYP/6-31+G(d)
[b] TD-BHandHLYP/cc-pVDZ
† for model with small QM (chromophore only).[18]
* for model with large QM (designated 0.30-noh).[18]

The accuracy of the results of the above-mentioned QM/pol-MM model compared to experimental data is subject to several factors related to both the theoretical approach and the experimental measurements. Some of these factors were discussed earlier in this chapter, while others will be described next. The embedding potential in mechanical and electrostatic embedding approaches, which are the methods typically used in the study of FPs, commonly is assembled from "pre-made" force field parameters and partial charges as implemented in, e.g., AMBER or CHARMM. In the case of the polarizable embedding model discussed here, a polarizable potential is employed, as described in Section 1.1. This potential can be specific either to an FP at a given geometry[90] or parameterized to be applicable to different geometries and FPs.[91–94] The first case consists of fragmenting the system and computing the multipoles as well as the polarizabilities

for each of the resulting fragments, i.e., constituting amino acids and other non-protein moieties (if included) like water molecules. In situations where conformational sampling is done, multipoles and polarizabilities need to be computed for each conformation. The fragmentation and potential assembly can be automated with PyFraME.[95] As expected, the choice of the level of theory used to compute the multipoles and polarizabilities will impact the quality of the potential and, consequently, the properties determined like MPA cross-sections.[31,78,96] Although computing the potential for a specific conformation or set of conformations is an aggregated cost in comparison to other QM/MM schemes, such as those using EE potentials, the payoff is an overall improvement in the accuracy of computed response properties, which are highly sensitive to the description of the electric field of the system.[26] For further details and caveats related to the use of the above mentioned polarizable model, we recommend the reader refer to Steinmann et al.[97]

1.3 Insights Into Fluorescent Proteins Obtained From QM/MM Studies

Excited state optimizations, which are required for the computation of emission energies, have been explored for and beyond the lowest energy transition, $S_0 \rightarrow S_1$, using QM/MM approaches.[98] The computational study of emission properties through QM/MM means has focused on different contexts, such as: (i) the computational "design" of red-shifted chromophores, where Grigorenko et al. discuss the red-shifting of the GFP chromophore emission when sandwiched between two tyrosine residues;[99] (ii) the study of the variation in colors within mutants, e.g., the study of DsRed mutants using SAC-CI;[100] and (iii) the understanding and "prediction" of the red shift in RFPs like mPlum.[80]

Many investigations on FPs have examined the response of certain excited state properties to point mutations within the chromophore cavity.[18] Moron et al.[80] posed an interesting path in the engineering of FPs via the use of statistical modeling tools, in particular for RFPs. Their studies of red-shifted RFPs characterized the interactions between the chromophore and the pocket amino acids. In addition, they found that the

red shift can be tuned through the modification of features in the chromophore, e.g., the acylamine angle or the H-bond interaction with the imidazole ring.[80] Grigorenko et al. described a cationic form of the chromophore of the kindling fluorescent protein (KFP) as the emissive state, which is uncommon within the general behaviour of the GFP family.[101]

In the case of the yellow FP, Park and Rhee[102] determined that several amino acids in the chromophore pocket, including Thr62, Arg96, Tyr145, His148, and Thr203, play active roles in determining its fluorescence energy. The Thr203 amino acid, which predominantly featured a stacking conformation during their simulations, contributes to an emission shift of 0.07 eV. For GFP, eYFP, Dronpa, phiYFP, and mTFP0.7 FPs, it was determined that the inclusion of polarization effects lead to the absorption shifts in these FPs variants.[103] For the KFP, the reasons behind its radiationless transitions were described.[104]

In addition to composing benchmark data on MPA of FPs, efforts have been made using Olsen et al.'s model for describing the roles of the residues closest to the chromophore on the 2PA cross-section and excitation energy of the RFP DsRed.[86] More recently, this polarizable embedding model was employed to determine the 2PA of a curated set of non-canonical chromophores (nCCs) within the RFP protein environment.[105]

The examples above show that insight is obtained from the computational study of FPs using QM/MM approaches. Based on this physical insight, it is possible to determine avenues through which FPs can be engineered in order to "improve" their photophysical properties.

Time-dependent density functional tight-binding (TD-DFTB) has been explored as a less computationally expensive alternative to TD-DFT and coupled cluster methods, although the performance of DFTB and TD-DFTB has been mixed. On one side, TD-DFTB exhibits deviations in the excitation energies of up to ~1 eV in comparison with experimental measurements[106] while, on the other hand, for other systems,

TD-DFTB results agree well with experiment.[107] In particular, TD-DFTB has shown poor performance in the computation of other excited state properties, such as transition and permanent dipole moments and, consequently, prediction of σ^{2PA} within the two-level model (2LM).[61] While performance of TD-DFTB for the excited state properties remains to be improved, DFTB might be used for the ground state structural analysis of FPs.[73,108]

1.4 Summary Comments on the Use of QM/MM for 1PA and MPA in FPs

Here, we discussed the computation of 1PA but, primarily, MPA properties of fluorescent proteins using QM/MM approaches and, in particular, those using a QM/pol-MM scheme.[109] Many photophysical properties (e.g., 2PA cross-sections) of FPs are influenced strongly by the effects of the protein environment that cannot be captured by cluster models. Thus, the use of QM/MM approaches becomes essential. However, their implementation requires accounting for many variables, including the QM region size, the QM method employed in the determination of excited state properties, and/or the nature of the embedding potential. Appealing for a better description of the chromophore–protein interaction, a polarizable potential can be included in the QM/MM models. PE models are critical for the description of 2PA cross-sections, although they also may be important in 1PA contexts. Investigations of non-degenerate 2PA, where the photons are not of equal energy, and higher order MPA remain to be explored computationally, including with QM/MM approaches.

Using QM/MM, many insights about the photophysical behaviour of some FPs have been determined, however, the work is far from being over. Since its formulation in 1976,[22] QM/MM has evolved into a wide variety of robust formulations for the computation of 1PA and MPA properties in complex systems. However, these methods still are underused, particularly in the context of engineering novel FPs or strategies to enhance the MPA performance of existing FPs. QM/MM approaches can be used to

dissect the underlying aspects behind the photophysical features of FPs. This information can be used to engineer FPs with exceptional far-red absorption, multi-photon absorption properties, and quantum yield, as well as resistance to photobleaching and quenching. Even further, statistical models can be created that, along with QM/MM computations, would contribute to the engineering of new FPs. Hence, many exciting developments and applications of QM/MM approaches to FPs remain.

1.5 Thesis Objectives and Outline

The present thesis provides computational insights about the multiphoton absorption processes of FPs and their chromophores, which aim to help in the design of new FPs with enhanced MPA properties (as will be discussed in later chapters) and within a procedural context, in the improvement of the computational strategies widely employed for this purpose. Here, the photophysical properties, such as 1PA, 2PA, and 3PA, of naturally occurring (canonical) and engineered (non-canonical) chromophores are investigated using different QM methods, as well as the ONIOM QM/nonpol-MM method and the QM/pol-MM approach proposed by Olsen et al.,[23] which was formulated for the computation of response properties within a PE potential. Overall, the first part of the thesis addresses the 1PA, 2PA, and 3PA for a set of green- and red-type FP chromophores, as well as for non-canonical FP chromophores and common dyes, for example fluorescein. These chromophores are studied in vacuum, i.e., no environmental effects are taken into account, using TD-DFT and TD-DFTB QM methods. In a subsequent part of the thesis, the environmental effects (in the form of the DsRed protein sequence) on the σ^{2PA} of a set of non-canonical chromophores are studied using Olsen et al.'s QM/pol-MM approach. Finally, the aforementioned polarizable QM/MM method is employed in the investigation of the so-called fruit series.[59] In particular, the case of the FP mCherry is investigated as a means to get insight into the role of neighbouring amino acids in determining its σ^{2PA}. The structure of the thesis described below.

- In Chapter 2, we evaluate the performance of TD-DFTB2 on the computation

of 2PA cross-sections (σ^{2PA}) and its physical components, including the electric dipole moments, for the set of non-canonical amino acids previously studied by Salem and Brown.[12] TD-DFTB2, as implemented in the DFTB+ code,[110] is explored as a less-expensive computational alternative to the methods typically employed for the study of MPA of FPs, such as TD-DFT and coupled-cluster methods.[13,77] The accuracy of TD-DFTB2 in the computation of σ^{2PA} for FP chromophores is contrasted against other computational results, such as the coupled-cluster CC2 method, as well as the TD-CAM-B3LYP and TD-PBE0 methods.[12] The chapter highlights the reasonably good efficiency of TD-DFTB2 in the computation of VEEs, but also, its poor performance in the computation of excited state permanent dipole moments, which causes an inconsistent overestimation of the σ^{2PA} of the studied chromophores.

- In Chapter 3, we evaluate the performance of a long-range corrected version of TD-DFTB (LC-TD-DFTB) implemented in the DFTBaby code,[111] in the computation of the σ^{2PA} for some of the non-canonical and canonical chromophores employed in Chapter 2. The performance of LC-TD-DFTB is evaluated in a similar fashion to TD-DFTB2 using the physical components and the magnitude of the σ^{2PA}s, which are computed within the 2LM. The results we obtained show that LC-TD-DFTB has a poorer performance than TD-DFTB2 in the computation of vertical excitation energies and electric permanent dipole moments. The σ^{2PA}s obtained using LC-TD-DFTB are up to ~100 GM larger than those computed using TD-CAM-B3LYP[12] or TD-DFTB2.[61]

- In Chapter 4, we present the computed 3PA features of the dyes fluorescein, serotonin, and rhodamine 6G, as well as for a set of canonical FP chromophores. In all cases, the σ^{3PA}s are computed for the chromophores in vacuum using TD-CAM-B3LYP via response theory. The 1PA and 3PA data of the dyes

19

is used as a means (and benchmark to an extent) to evaluate the viability of TD-CAM-B3LYP in the computation of 3PA properties of FP chromophores. We found a reasonable agreement between our computational results and the available experimental σ^{3PA}s for the dyes considered in this work. Furthermore, the results we computed for the red and green fluorescent protein (RFP and GFP, respectively) chromophores are in the order of magnitude of the data measured for various RFPs and the GFP. Thus, we consider TD-CAM-B3LYP is an adequate method to sample the 3PA response of FP chromophores. The computations presented here do not account for environmental effects. Future investigations of the 3PA features of FPs chromophores will include the implementation of a polarizable embedding potential.

• In Chapter 5, we discuss for a set of non-canonical chromophores (nCCs) the environmental effects exerted by the protein sequence of the RFP DsRed. Salem et al. [14] previously computed the σ^{2PA}s for the set of nCCs employed here. These previous computations were carried out for the nCCs in vacuum and in water via the polarized continuum model of solvation, PCM (water), using TD-DFT methods. Here, we constructed systems of the form nCC–DsRed, for which we computed their σ^{2PA} using a quantum mechanical/molecular mechanical polarizable embedding, QM/pol-MM. Based on our results and models, the protein has a major effect on the 2PA response of the nCCs. In most of the cases, we observe a reduction of the σ^{2PA} values obtained by Salem et al. [14] This investigation is a preliminary attempt to determine the 2PA response of FPs containing nCCs using a polarizable embedding QM/MM approach. The DsRed protein sequence does not enhance the σ^{2PA} of the nCC embedded in it. However, it could be used as a template to create a tailored environment (made from point mutations in the protein sequence), around each nCC. This is the subject of future investigations.

- In Chapter 6, we present the cross-sections and physical components of the 2PA response for a set of FPs of the so-called mFruits series using the same polarizable embedding model employed in Chapter 5. In particular, we computed the σ^{2PA}s of the mFruits in Figure 1.3 for which a crystal structure was available in the PDB Data Base. Our results are compared with those determined experimentally by Drobizhev et al.[5] (Figure 1.3). Furthermore, for the case of mCherry at pH 11 we investigated the effect on the σ^{2PA} of the neighbouring amino acids located within a radius of 2 Å from the chromophore. The results were employed to create two cluster models, which were extracted from the mCherry FP. Based on the effect that each neighbouring amino acid has on the 2PA of the canonical chromophore in the mCherry at pH 11, we point-mutated these clusters and obtained their σ^{2PA}.

- Chapter 7 contains the concluding remarks of this thesis based on the outcomes obtained in Chapters 2–6. In addition, this chapter presents some perspectives on the role of QM/MM computational approaches in the engineering of FPs with enhanced MPA features.

Chapter 2

Determination of Two-photon Absorption Cross-sections in Fluorescent Protein Chromophores Using Tight-binding Time-dependent Density Functional Theory[2]

2.1 Introduction

Two-photon absorption is the primary photo-excitation process used in multi-photon microscopy (MPM), a technique that plays an important role among the imaging techniques due its non-invasive nature.[112] Out of the large variety of species that exhibit two-photon absorption, the fluorescent proteins (FPs) represent a highly important class, especially as used in biological imaging. After being discovered by Shimomura et al. in the early 1960s,[2] and after their multi-photon features were realized, FPs gained the interest of researchers because of their potential use as probes in fluorometric analysis in MPM. The light-absorption properties of FPs are linked intimately to the associated chromophore. In this sense, by finding, designing, and ultimately synthesizing or engineering chromophores that exhibit notable one- and two-photon absorption (1PA and 2PA, respectively) properties, along with other photophysical characteristics, such as quantum yield or responsiveness to external stimulation, FPs can be used to monitor the concentration of cellular ions (Ca^{2+}, H^+ and Cl^-), as voltage sensors,[4] among other applications.[5,12–14,113,114]

---

[2]This chapter has been copied and/or adapted from the publication: Rossano-Tapia, M.; Brown, A. Determination of two-photon-absorption cross-sections using time-dependent density functional theory tight binding: application to fluorescent protein chromophores. *J. Chem. Theory Comput.* **2019**, *15*, 3153–3161.

Computationally speaking, the methods that account for 2PA properties that exhibit a good performance in terms of accuracy and computing time typically are based on Time-Dependent Density Functional Theory (TD-DFT),[115] although other wave function-based methods are beginning to become viable for studying 2PA.[9,12–14,60,62,116–118] Nevertheless, both TD-DFT and wave function methods preferably are used on small size systems (10–50 heavy atoms) due to their computational costs.[119,120] For this reason, the computation of 2PA cross-sections commonly is limited to the chromophore, excluding extended systems that contain parts of the protein backbone, structurally important waters, and/or the ability to sample a diversity of ground state geometries. The information about how the protein environment affects the 2PA in a candidate chromophore is not conclusive yet, although there are some recent insights.[5,14,16,26,86,121] For instance, Rossano-Tapia et al. studied the effect of the protein environment on the 2PA cross-sections of a series of non-canonical chromophores using a polarizable embedding quantum mechanical/molecular mechanical scheme.[105] To examine these effects in more detail, it would be beneficial to find an efficient computational method that could allow the accurate evaluation of photophysical properties for large sets of chromophores (including at a large number of geometries) and the influence of their environment (by exploring extended chromophore structures with the explicit inclusion of nearby waters and residues). The latter and some other caveats involved in the computational study of the environmental effects on one- and multi-photon absorption response properties of FPs using quantum mechanical/molecular mechanical approaches is discussed in the review published by Rossano-Tapia and Brown.[109]

The present work proposes the evaluation of the time-dependent self-consistent tight-binding method TD-DFTB2, computationally less expensive than common TD-DFT methods[122], for the computation of 2PA cross-sections of a set of chromophores derived from non-canonical amino acids proposed by Salem and Brown[12,14] (see Figure 2.2). The latter is depicted in Figure 2.1. The 1PA and 2PA properties of these chromophores

had been examined previously using the TD-DFT methods, B3LYP/6-31+G(d,p) and CAM-B3LYP/6-31+G(d,p). Both methods describe the trends observed in the experimental values as well as provide a reasonably consistent picture of the underlying physical and (1PA and 2PA) photophysical properties; however, as highlighted recently for the specific case of 2PA, one always must be careful with the choice of functional for one's specific problem of interest, especially for systems with large 2PA, as charge transfer can play an important role, such that the use of range-separated functionals is crucial.[116,123] Moreover, Salem et al. observed that the trends in 2PA properties are preserved using either quadratic response theory or within the sum-over-states formalism through the two-level model (2LM).
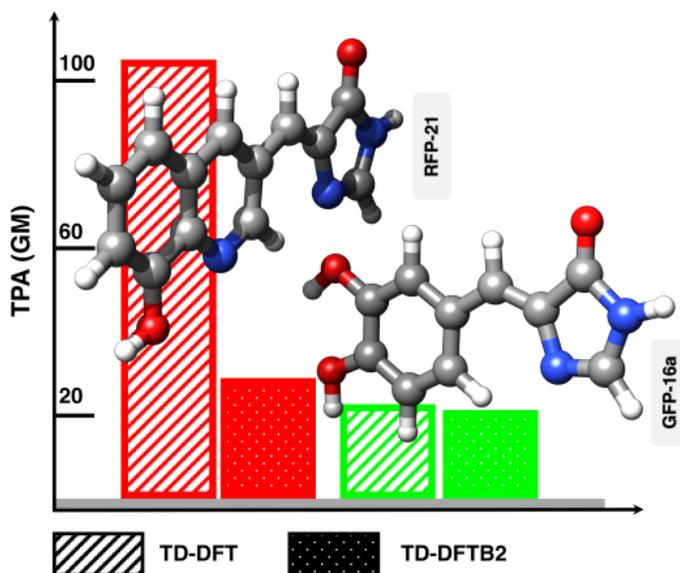
Figure 2.1. Depiction of the deviation of the two-photon absorption (2PA) cross-sections obtained using TD-DFTB versus TD-DFT.[14]

The DFTB method is an approximation to the total energy of density functional theory (DFT).[124] The accuracy of the method depends on the specifics of the approximation used, e.g., second- or third-order (DFTB2[124] or DFTB3,[125,126] respectively). The DFTB3 approximation is an improvement over DFTB2, as it overcomes two main limitations that the second-order approximation exhibits: (i) the chemical hardness varies depending

on the charge state of the species, which allows the atom to have a different chemical hardness than its precursor neutral atom or cation and (ii) the shape of the atoms is not fixed since the atom size depends on the hardness.[126] The principal feature that DFTB possesses is its low computational cost; according to Yang et al., DFTB is 2–3 orders of magnitude faster than purely DFT methods (using small to medium-sized basis sets).[125] Furthermore, the performance of DFTB-based methods[127] is in good agreement with experimental data and the results obtained using DFT methods. The latter according to previous reports on the computation of ground-state properties such as geometrical parameters,[128,129] isomerization energies, and heats of formation.[130–132]

Besides ground-state properties, the time-dependent extension of DFTB, the so-called TD-DFTB,[133] allows the prediction of spectroscopic properties, such as vertical excitation energies for systems that range from small organic and inorganic molecules[134–137] to larger ones, such as fullerenes and polyacenes,[134] as well as dyes.[137,138] Although most of the studies have focused on the excitation energies, Oviedo and Sánchez[139] tested TD-DFTB on the computation of transition dipole moments for different photosynthetic pigments.

In the present chapter, we evaluate the performance of TD-DFTB, particularly the implementation of DFTB2 in the DFTB+ software package,[110] for the computation of 2PA cross-sections for the fluorescent protein chromophores previously studied by Salem et al.[12,14]

2.2 Computational Methods

The chromophores considered here, shown in Figure 2.2, were taken from the work conducted by Salem et al.,[12,14] where new FP chromophores were proposed based on the incorporation of non-canonical amino acids.[140]
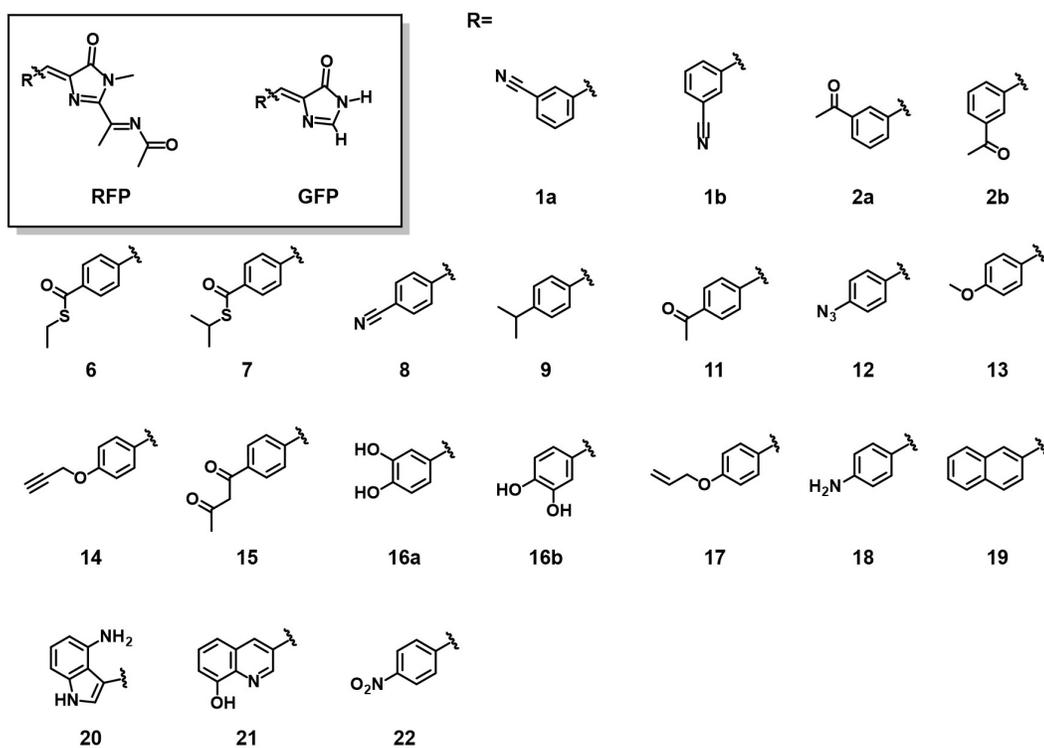
Figure 2.2. Structures of the models used in this work. [12,14] For both RFP-and GFP-like chromophores, the R-moiety is the same in each case.

The set considered here includes 42 structures (Figure 2.2), where 21 are RFP-like, with extended conjugation, and 21 are GFP-like. Although Salem et al. described 50 non-canonical amino acid-based models, [12,14] here we only used those for which the respective mio set parameters [124,126,141] are available. Therefore, molecules containing boron and/or halogens, i.e., chromophores designated as 3, 4, 5, and 10 in the previous work, [12,14] were excluded from the DFTB2 computations. While four molecules in each of the RFP- and GFP-like sets (eight molecules in total) have been excluded, the same molecule numbering is used from the previous work to facilitate comparison.

The aforementioned RFP-and GFP-like structures were optimized by Salem et al. at the PBE0/6-31+G(d,p) level of theory. The validity of structures optimized at a DFT level for being used in TD-DFTB computations can be justified through results shown in the literature; Krüger et al. [129] reported a mean absolute deviation of 0.016 Å and 1.4° (±1% and 2%) with respect to reference values of bond distances and bond angles, respec-

26

tively, obtained using the BLYP[142,143] functional and the cc-pVTZ basis set. The latter deviations are similar to those obtained for DFTB with respect to MP2/6-31G(d).[129] Furthermore, Sattlemeyer et al.[130] showed that structural parameters, such as bond lengths, bond angles, and dihedral angles, computed using DFTB2 are in good agreement with results obtained at the MP2/cc-pVTZ level of theory (mean absolute error [MAE]=0.012 Å, $1.0°$, $2.9°$, respectively), and show a lower MAE in comparison with other semi-empirical methods tested using the same benchmark method.[127]

The TD-DFTB2 computations were performed based on linear response theory in the Casida formulation.[133] All properties, including transition dipole moments (μ_{01}), ground state (μ_0) and excited state (μ_n) permanent dipole moments, excitation energies, and oscillator strengths (OS) were computed using the DFTB+ code v.1.3.[110] The excited electronic states were computed without explicitly indicating the symmetry. The corresponding mio parameters [124,126,141] were obtained from the official DFTB website.[144] In this respect, it is important to mention that, although the aforementioned parameter set was developed using organic and biological molecules as reference models, it did not take into account excited electronic states but only ground-state properties.[124,126,141]

Based on the previous work[12,14] using TD-DFT, the lowest energy bright 2PA state involved (primarily) the HOMO→LUMO excitation. However, for many cases using TD-DFTB2, the HOMO→LUMO transition did not correspond to the lowest energy first excited state (S_1) and, thus, the first 10 excited states were computed for all molecules. However, only the data of the transitions of interest is included here (see Figure A.1). Then, the HOMO→LUMO transition was identified and the state (S_n) associated with it was used for the transition energy and to compute the corresponding transition dipole moment and excited state permanent dipole. The re-ordering of the excited electronic states using TD-DFTB has been observed previously;[136] however, in these cases, the states could be assigned by symmetry, while in the present work this symmetry assignment could not be done, as C_1 symmetry was used in all computations.

All the 2PA cross-sections were obtained within the two-level model (2LM) and calcu-lated through their corresponding two-photon absorption transition moments, δ^{2PA}.[13,14,145,146] The discussion of the general expression for the 2PA transition moment can be found in the work carried out by Monson and Mcclain[145] or by Andrews and Thirunamachan-dran.[146] For the particular case of linearly polarized light, the 2PA transition moment is given by[60,145]

$$\delta^{2PA} = \frac{1}{15} \sum_{\alpha\beta} S_{\alpha\alpha} S_{\beta\beta}^* + 2 S_{\alpha\beta} S_{\alpha\beta}^*, \qquad (2.1)$$

where $S_{\alpha\beta}$ are the two-photon transition matrix elements described as

$$S_{\alpha\beta} = \sum_n \left[\frac{\langle 0 | \mu_\alpha | n \rangle \langle n | \mu_\beta | f \rangle}{\omega_n - \omega} + \frac{\langle 0 | \mu_\beta | n \rangle \langle n | \mu_\alpha | f \rangle}{\omega_n - \omega} \right]. \qquad (2.2)$$

Here $\langle 0 | \mu_\alpha | n \rangle$ corresponds to the transition dipole moment from the ground state $|0\rangle$ to the state $|n\rangle$ along the Cartesian coordinate α, ω_n corresponds to the energy associated to such states, $|f\rangle$ is the final state, and ω is the photon energy given by $\omega = \omega_f/2$. It is necessary to point out that δ^{2PA} can be evaluated via quadratic response theory (for TD-DFT and, in principle for TD-DFTB); nevertheless, we used the so-called sum-over-states expression (SOS) since the DFTB+ code is not capable yet of performing 2PA computations using quadratic response theory. The sum-over-states expression described in Equation 2.2 can be truncated within the few-state-model (FSM) approach and can be expressed in terms of a 2LM as [14,118,147]

$$\begin{aligned} S_{\alpha\beta} &= \frac{2}{\omega_1} \left[\mu_{\beta_{01}} \left(\mu_{\alpha_{11}} - \mu_{\alpha_{00}} \right) + \mu_{\alpha_{01}} \left(\mu_{\beta_{11}} - \mu_{\beta_{00}} \right) \right] \\ &= \frac{2}{\omega_1} \left[\mu_{\beta_{01}} \left(\Delta\mu_\alpha \right) + \mu_{\alpha_{01}} \left(\Delta\mu_\beta \right) \right], \end{aligned} \qquad (2.3)$$

where $\mu_{\alpha_{mn}}$ is the αth component of the dipole moment vector, whereas $\Delta\mu_\alpha$ corresponds to the difference between the excited and ground state permanent dipole moments for the αth component as well. Based on the work carried out by Alam et al.,[147,148] the

2PA transition moment can be expressed in terms of the excitation energy (ω_1), the magnitude of the transition dipole moment $\|\mu_{01}\|$, the magnitude of the difference in permanent dipole moments between the ground and excited states $\|\Delta\mu\|$, and the angle θ between these two vectors:

$$\delta^{2PA} = \frac{16}{15}\left(\frac{\|\mu_{01}\|\,\|\Delta\mu\|}{\omega_1}\right)^2 \left(2\cos^2\theta + 1\right). \qquad (2.4)$$

Finally, the 2PA cross-sections for a Lorentzian lineshape (Γ) can be determined in macroscopic units (10^{-50} cm^4 s molecule^{-1} photon^{-1}or GM[149]) as

$$\sigma^{2PA} = \frac{N\pi^2 a_0^5 \alpha}{c}\frac{\omega^2}{\Gamma}\delta^{2PA}. \qquad (2.5)$$

In Equation 2.5, a_0 is the Bohr radius, α is the fine structure constant, c the speed of light, and ω is the photon energy. The integer value N and Γ were set as 4 and 0.1 eV, respectively, so the results could be compared to single beam experimental data,[150–153] as described by Beerepoot et al.,[60] and moreover, to be consistent with our benchmark data[12,14] that were based on other 2PA computations carried out in analogous fashion.[11,154]

2.3 Results and Discussion

The results we obtained were contrasted mainly with Salem et al.'s data,[12,14] where they tested the performance of B3LYP[155] and CAM-B3LYP functionals for the computation of 2PA cross-sections. The primary comparison considered here involves the data obtained using the B3LYP functional[155] and the 6-31+G(d,p) basis set,[156–160] within the 2LM. The 2PA cross-sections determined at this level of theory have been shown to be in good agreement with experimental values[13] and are consistent with the values obtained via quadratic response theory (QRT).[161–164] In addition, the 2PA cross-sections obtained here via TD-DFTB2 (2LM) can be compared analogously with the TD-CAM-B3LYP (2LM) results and with those obtained previously using TD-B3LYP/6-31+G(d,p) (QRT) (see

Figures A.3 and A.5).

For all the chromophores, the transition involving primarily the HOMO→LUMO excitation was considered for determining the 2PA properties; as mentioned in the Section 2.2, this transition, as determined using TD-DFTB2 does not necessarily correspond to the first excited state of lowest energy. The specific state S_n considered for each molecule is given in Table 2.1 for the RFP-like chromophores, and in Table A.7 for the case of the GFP-like ones.

Table 2.1 also lists the excitation energies, oscillator strengths, 2PA transition moments (δ^{2PA}), and 2PA cross-sections computed for the RFP-like chromophores. Also reported are differences with respect to the previous[12,14] TD-B3LYP (2LM), TD-B3LYP (QRT), and CAM-B3LYP (2LM) results (Table A.2, Figure A.3, and Figure A.4, respectively). Transition dipole moments (μ_{0n}) and the difference between the excited and ground state permanent dipole moments ($\Delta\mu$), which are two of the critical elements for computing the 2PA cross-sections within the 2LM, from both the present TD-DFTB2 and previous TD-B3LYP computations are detailed in Table A.3. Two-photon cross-sections were determined according to Equations 2.4 and 2.5, and then compared to the TD-B3LYP (2LM) results, see Figure 2.3; similar comparisons to TD-B3LYP (QRT) and CAM-B3LYP (2LM) are available in Appendix A.

Table 2.1. Excitation Energies and the Corresponding Excited State S_n for the HOMO→LUMO Transitions, Oscillator Strengths (OS), 2PA Transition moment δ^{2PA} (au), and 2PA Cross-Sections (σ^{2PA}) for the RFP-like Models Obtained at the TD-DFTB2 Level of Theory

Model	S_n	Energy (eV)	OS	δ^{2PA}	σ^{2PA} (GM)
1a	3	2.874	0.45	1808	5.5
1b	3	2.893	0.44	1053	3.2
2a	4	2.851	0.45	1844	5.5
2b	4	2.893	0.43	766	2.3
6	5	2.839	0.50	1420	4.2
7	5	2.773	0.57	765	2.2
8	3	2.816	0.54	1614	4.7
9	3	2.791	0.51	6782	19.4
11	4	2.790	0.55	1141	3.3
12	3	2.632	0.56	22205	56.4
13	3	2.629	0.51	17851	45.2
14	3	2.635	0.54	20188	51.4
15	5	2.784	0.57	741	2.1
16a	2	2.395	0.22	25863	54.4
16b	2	2.443	0.26	22503	49.2
17	3	2.611	0.53	20724	51.8
18	3	2.574	0.51	18469	44.9
19	3	2.631	0.38	34392	87.3
20	1	2.191	0.23	47890	84.3
21	1	2.197	0.04	15599	27.6
22	5	2.804	0.49	1973	5.7
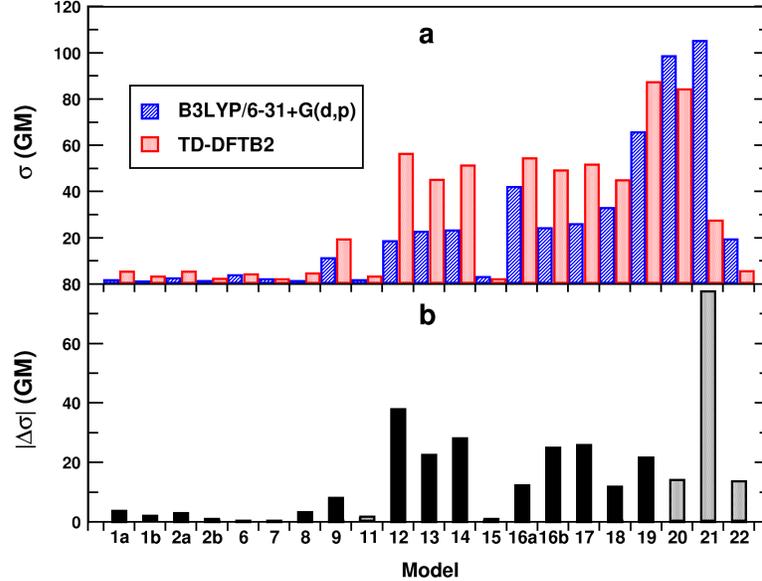
Figure 2.3. For the RFP-like chromophores, (a) comparison of the magnitude of the 2PA cross-sections (σ^{2PA}) obtained using TD-DFTB2 with respect to TD-B3LYP/6-31+G(d,p) values [14] within the two-level model and, (b) the differences between the TD-DFTB2 and TD-B3LYP/6-31+G(d,p) σ^{2PA} values ($\left|\Delta\sigma^{2PA}\right| = \left|\sigma^{2PA}_{DFTB2} - \sigma^{2PA}_{B3LYP}\right|$). Negative differences are shaded in grey whereas the positive differences are given in black.

In general, TD-DFTB2 overestimates the 2PA cross-section σ^{2PA} (except for models

20, 21 and 22 for which the 2PA cross-section is significantly, > 10 GM, underestimated). The difference in σ^{2PA} between both methods does not vary within a well-defined interval, i.e., in some cases, they are within a few units (1–4 GM), e.g., models 1a, 1b, 2a, 2b, 6, 7, 8, 11, and 15, whereas in others, there is a gap of more than 20 GM as in the case of molecule 21. In addition, TD-DFTB2 results do not follow the trend described by B3LYP,[155] e.g., model 13 exhibits a larger 2PA cross-section than 12 using B3LYP,[155] but smaller than with DFTB2. A similar case occurs with 17 and 18, where 18 was expected based on the behaviour observed in the previously reported work,[14] (Table A.1) to be larger than 17 but it is smaller using DFTB2. The above differences are clearly described by Figure 2.4, where poor correlation between these two methods is observed; however, if model 21 with the largest difference between the two methods (see Figure 2.3) is excluded from the fit, the correlation improves ($R^2 = 0.8$).
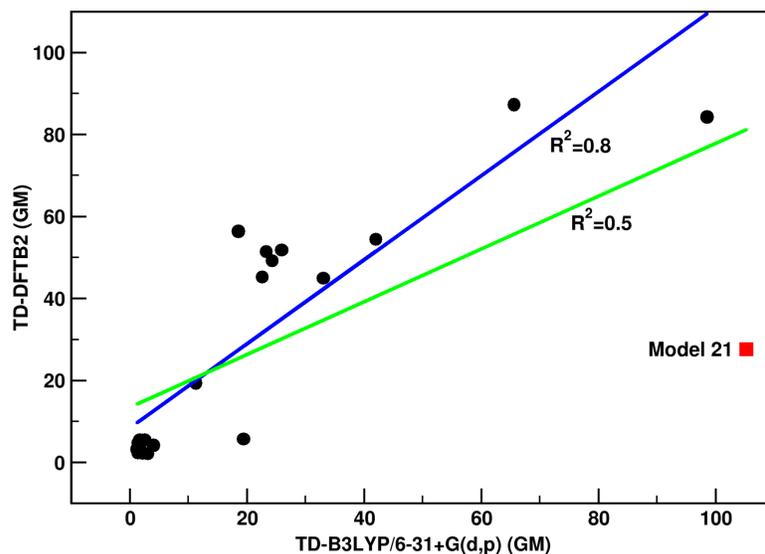
Figure 2.4. Evaluation of the correlation between the 2PA cross-sections computed using TD-DFTB2 and TD-B3LYP/6-31+G(d,p)[14] for the RFP-like models. (i) In green, the correlation considering all models and (ii) in blue, the correlation discarding Model 21, which exhibits the largest deviation among all the models with respect to TD-B3LYP/6-31+G(d,p) cross-sections.

Among all the structures, model 21 has the largest σ^{2PA} difference with respect to the B3LYP[155] (2LM) data ($\Delta\sigma^{2PA}$ of almost -80 GM or 74% error, see Figure 2.3 and Table 2.3). The main reason for this difference is the large underestimation of the

transition dipole moment (Table 2.3) that is directly related to the cross-section, see Equation 2.4. Model 16b suffers similar discrepancies, as its transition dipole also is underestimated (30%); nevertheless in this case, σ^{2PA} is not diminished drastically as in 21. This can be explained through the difference in permanent dipole moments $||\Delta\mu||$, which is overestimated considerably in comparison with 21 and seems to compensate for the small transition dipole moment. The models with the lowest deviation (<10 GM) with respect to B3LYP[155] (2LM) values are those containing nitrile, ketone, thioester, and alkyl groups (see Figure 2.2), whereas the ones with the highest gap (> 10 GM) correspond to those containing hydroxy, ether, azo, amino, nitro, and fused unsaturated rings groups. Related to the above differences is the work reported by Fabian et al.[135] where they evaluated the performance of TD-DFT and TD-DFTB among other semi-empirical methods in the estimation of excitation energies and oscillator strengths for a diverse set of organic molecules, including some with sulfur- and azo-moieties. In this work, the authors describe that the TD-DFTB method can be used to obtain similar results to TD-DFT. For the non-sulfur containing compounds, the MAE of the TD-DFTB2 results with respect to the TD-DFT ones is about 0.256 eV and 0.10 for the excitation energies and oscillator strengths, respectively, as determined based on the reported values. Such MAE was determined by considering both $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions. The MAE values are comparable to the ones obtained in this work which are 0.214 eV and 0.17 for the RFP-like set and 0.142 eV and 0.16 for the GFP-like models, for the excitation energies and oscillator strengths, respectively (Table 2.2) Fabian et al.[135] also mention that the TD-DFT results obtained for the sulfur-containing models show a better agreement with the benchmark data over the azo-containing molecules.

Table 2.2. Mean Absolute Error of the Excitation Energy, Oscillator Strength (OS), δ^{2PA}, σ^{2PA}, $||\Delta\mu||$, $||\mu_{01}||$, and $\cos^2\theta$ of the RFP-like and GFP-like Models Obtained at TD-DFTB2 Level of Theory Within the 2LM

Chromophore family	Energy (eV)	OS	δ^{2PA}	σ^{2PA} (GM)	$||\Delta\mu||$	$||\mu_{01}||$	$\cos^2\theta$
RFP-like	0.214	0.17	6329	0.7	0.4	15.0	0.14
GFP-like	0.142	0.16	4532	17.1	1.0	0.3	0.14

Table 2.3. For Selected Models From Table 2.1, the Percent Error (%) for Excitation Energies, oscillator strengths (OS), the Vector Corresponding to the Difference Between the Excited State Permanent Dipole Moment and the Ground State One ($\|\Delta\mu\| = \left(\sum \Delta\mu_\alpha^2\right)^{1/2}$), and the Transition Dipole Moments (δ^{2PA}) Associated with TD-DFTB2 with Respect to TD-B3LYP/6-31+G(d,p) Values[14]

Model	Energy	OS	$\|\Delta\mu\|$	$\|\mu_{01}\|$	δ^{2PA}
6	−4	−26	33	−12	6
7	−4	−25	30	−12	−1
12	−8	−26	83	−10	205
13	−8	−24	50	−9	100
14	−8	−23	55	−8	121
15	−5	−16	8	−7	−32
16a	−15	−59	61	−30	30
16b	−16	−59	95	−30	103
18	−7	−30	31	−13	36
19	−7	−45	48	−23	33
20	−7	−40	15	−19	−14
21	−19	−85	21	−57	−74
22	−3	−18	−33	−9	−71

The overestimation of 2PA cross-sections obtained within the 2LM, computed through Equation 2.4, results from a combination of factors: a systematic underestimation of both the transition dipole moment and the excitation energies as well as the overestimation of the permanent dipole differences (Table A.4). There are some examples, as in model 11, where the cancellation of errors leads to a 2PA cross-section similar to the reported value (Table A.2). The underestimation of the excitation energies is, in fact, part of the expected behaviour for TD-DFTB2 computations and already has been discussed by other authors.[135,136]

Oviedo and Sánchez[139] screened the transition dipole moments of a group of photosynthetic pigments using the TD-DFTB implementation in the DFTB+ code. From their study, they found that TD-DFTB transition dipoles are correlated well ($R^2 = 0.99$) to TD-DFT ones obtained using the B3PW91 functional and the 6-31+G(d) basis set. In general, the trend for the TD-DFTB2 transition dipole moments determined here is consistent with these observations, i.e., they tended to be smaller than the given B3LYP[155] ones, with the exception of models 1b, 11, 15, 21, and 22 (Table A.5). For the latter molecules, excluding 21, whose differences with TD-DFT have been discussed previously, the drastic underestimation of the μ_{0n} ($R^2 = 0.14$, see Table A.6) did not

have a dramatic effect on the 2PA cross-section (σ^{2PA}) due to a cancellation of errors.

We found that the overestimation of the difference between excited and ground state permanent dipole moments in the RFP-like chromophores (Table A.4) is due to an exaggeration of the excited state permanent dipole moments. Specifically, the ground state dipole moments computed through TD-DFTB2 are modestly underestimated and better correlated ($R^2 = 0.55$) to the TD-B3LYP results (average error =26%; standard deviation =10 D) in comparison with the excited state permanent dipoles which are poorly correlated ($R^2 = 3.0 \times 10^{-3}$) and inconsistently overestimated (average error =73.1%; standard deviation =64 D) to the same benchmark data. This behaviour extends to the GFP-like chromophores, where similar correlations and errors are observed (Tables A.10 and A.11).

Supporting the above conclusions is the fact that the ground state geometries, along with their respective dipole moments have been tested already and shown to be consistent with more accurate methods or experimental data, as shown by Elstner et al. [128] They concluded that within the SCC-DFTB scheme the computed ground state dipoles for a set of small glycine and alanine peptides (length $\leq$ 8 residues) varied around 0.5 D (0.2 au) and tended to be smaller with respect to the ones they obtained at the B3LYP/6-31G(d) level of theory. [128] Also, Cramer determined using charge model 3 (CM3) the ground state dipoles for a set of mostly organic molecules and obtained underestimated values but ones close to benchmark results. [165,166]

The reasons behind the inaccurate estimation of the dipoles, oscillator strengths, and excitation energies lay in the formulation of the TD-DFTB method, including the fact it employs Mulliken analysis in the calculation of atomic charges used in the evaluation of dipole moments. However, Ruger et al. [167] provide some insights regarding the strengths and weaknesses of this TD-DFTB method for obtaining accurate properties.

The case of GFP-based models is not that different from what we have already discussed for the RFP-like chromophores, but some particular details are important.

The analysis of the GFP-type chromophores comprised 10 models (out of 21 total) since there was no benchmark data to compare with for the rest of the set (particularly for the TD-B3LYP/6-31+G(d,p) 2LM level of theory). The percentage error of the TD-DFTB2 excitation energies, OS, $\Delta\mu$, and μ_{0n} for the 10 models is given in Table 2.4. Cross-sections computed with TD-DFTB2 are considerably larger (at least for six out of the ten models) than the 2LM TD-B3LYP results (Figure A.6). The extent of how the other four models (16a, 19, 20, and 22) are underestimated is not comparable with the overestimation associated to models 6, 7, 12, 13, and 14; e.g., both 6 and 19 differ from benchmark data with around the same magnitude, 32 and 20 GM, respectively. However, these differences correspond to errors of 3150% and 64% for each case; the huge percentage error for model 6 is due to the very small 2PA cross-section predicted via TD-DFT. For the actual values of 2PA cross-sections, excitation energies, oscillator strengths, and 2PA transition moments, refer to Table A.7, while for dipoles see Table A.8. In terms of which molecules exhibit the smallest or largest deviation for 2PA cross-sections (σ^{2PA}) from TD-B3LYP values, we expected to find some coincidences between RFP and GFP sets, with GFP models containing sulfur as 6 and 7 having the smallest errors, as happened with their RFP analogues. Unfortunately, 6 and 7 are, in fact, the models with the largest deviation; in addition, instead of being underestimated, the 2PA cross-sections now are overestimated. Model 6 has an error of 33% for the RFP-like set while for the GFP-like one it is 522%. The elements discussed above result in a poor correlation between TD-DFTB2 and TD-B3LYP/6-31+G(d,p) 2PA cross-sections, excitation energies, oscillator strengths, permanent dipole moments, and transition dipole moments ($R^2 = 0.03$, 0.61, 0.76, 0.25, and 0.71, respectively), given in Table A.11

Table 2.4. For a Selected Set of GFP-like Models, Percentage Error Associated with the Excitation Energies, Oscillator Strengths (OS), the Vector Corresponding to the Difference Between the Excited and the Ground State Permanent Dipole Moment ($\|\Delta\mu\| = \left(\sum \Delta\mu_\alpha^2\right)^{1/2}$), and the Transition Dipole Moments (δ^{2PA}) Obtained with TD-DFTB2 with Respect to TD-B3LYP/6-31+G(d,p) Values [12]

Model	Energy	OS	$\|\Delta\mu\|$	$\|\mu_{01}\|$	δ^{2PA}
6	−3	−25	522	−12	3150
7	−3	−23	391	−11	3085
12	−5	−21	168	−9	513
13	−5	−16	85	−6	221
14	−5	−15	107	−5	300
16a	−14	−58	38	−30	−10
18	−3	−20	40	−9	58
19	0	47	−51	21	−64
20	−4	−42	11	−22	−27
22	−1	−18	−9	−9	−31

Models 12, 13, and 14 containing N-derived groups exhibited similar differences when moving from RFP-like to GFP-like models, i.e., they are overestimated by more than 100% from their respective 2LM TD-B3LYP values. That change between GFP-like and RFP-like chromophores points towards the fact that TD-DFTB2 does not show a clear preference in the prediction of excited state properties depending on the type of atoms the molecules contain. Despite the above discrepancies, the rest of the excited state properties behaved as expected, i.e, excitation energies and μ_{0n} tended to be smaller, and $\Delta\mu$ to be larger than TD-B3LYP values. Once again, the large values for the excited state dipoles are primarily involved in the incorrect prediction of 2PA cross-sections.

2.4 Conclusions

Two-photon-absorption cross-sections computed within the two-level model tend to be overestimated by the TD-DFTB2 method in comparison to their TD-DFT counterparts. From the models considered here, it is not possible to conclude if this behaviour is related merely to the nature of the specific chemical species examined (FP-like chromophores) or to a weakness in the formulation of the method when it comes to determining excited state properties. For both the RFP-and GFP-like sets, the models containing nitrogen in the form of azo and amino exhibit large errors in their 2PA cross-sections. In terms of the excited state properties required for the determination of 2PA cross-sections, see

37

Equation 2.5, the deviation results from a systematic underestimation of the excitation energies and transition dipole moments, in addition to a considerable exaggeration of the differences between the excited state permanent dipole moment and the ground state permanent dipole moment. In general, the difference is a consequence of the overestimation of the excited state permanent dipole moment. As DFTB is derived from DFT, particularly because its parametrization is based on the GGA functional PBE, it carries along some problems that DFT exhibits. Gaus et al. [127] and Lundberg et al. [168] have provided more insights in this regard. Table A.1 shows for some RFP-and GFP-like selected models, energies and oscillator strengths computed using TD-PBE/6-31+G(d,p), and their percent of deviation with respect to the results obtained by Salem et al. [12,14]. As expected, energies and oscillator strengths are underestimated as in TD-DFTB. Also, the molecular orbitals were plotted to determine if the switching in them observed in TD-DFTB was directly related to its parent functional; as can be seen from model 22 in Figure A.2, the HOMO$\rightarrow$LUMO is not the transition with the largest weight for the same S_n among the methods. Therefore, we can suggest that some of the inconsistencies observed in TD-DFTB were inherited from PBE.

It was not possible for all the chromophores considered by Salem et al. [12,14] to estimate their 2PA cross-sections using TD-DFTB2 as implemented in the DFTB+ code due to the unavailability of parameters for certain atoms. This had evident consequences on the statistics of the present work but can represent even a worse problem if the span and variety of molecules to be tested wants to be increased. One alternative to TD-DFTB2 will be TD-DFTB3, a method for which halides and some metals such like Mg or Zn are included in the parameters set list. Although the proper parameters exist for DFTB3, the TD-DFTB+ code does not (yet) have a TD-DFTB3 implementation in the version used in the present work. Until further testing can be carried out, thus providing at least systematic errors, TD-DFTB2 cannot be recommended for determining 2PA cross-sections within the 2LM. Thus, implementation of QRT for TD-DFTB may

be premature. However, the prediction of 1PA properties is far more systematic. The second alternative would be moving to implementations of improvements to DFTB that attempt to compensate the lack of long-range contributions.[122,169] For example, the software package DFTBaby, developed by Humeniuk and Mitrić[170], in which the excitation energies can be tuned to approximate CAM-B3LYP (mean error for excitation energies with respect to CAM-B3LYP results of 0.46 eV and 1.13 eV for long-range corrected and non-corrected TD-DFTB respectively) or even, experimental results. The results corresponding to the evaluation of the performance of the DFTB implmentation in the DFTbaby code are discussed in Chapter 3.

Chapter 3

Determination of Two-Photon-Absorption Cross Sections of Fluorescent Protein Chromophores Using Long-Range Corrected Time-Dependent Density Functional Theory Tight Binding

3.1   Introduction

Fluorescent proteins (FPs) were discovered in the 1960s [1,2] and are now popular in clinical imaging. They possess notable multi-photon absorption (MPA) properties, specifically, two-photon absorption (2PA). Their MPA properties have contributed to the increased interest in using them as biological probes. The use of 2PA has many advantages over one-photon absorption (1PA), e.g., deep penetration and less probability of photo-damage in the sampled tissue. FPs are composed of a chromophore, which is located within a barrel-shaped protein. The chromophore in FPs is formed, via an auto catalytic reaction, by three canonical amino acids contained within the FP sequence. This sequence varies within the different FP families (Figure 3.1), e.g., in the green fluorescent protein (GFP), the sequence is composed of serine (Ser), tyrosine (Tyr), and glycine (Gly).

Canonical Chromophores
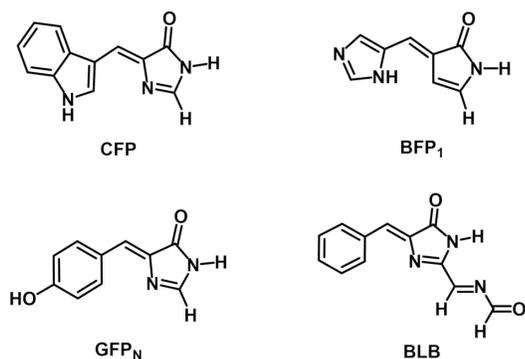
Figure 3.1. A set of canonical chromophores employed in this investigation, part of a more extensive set, previously studied computationally by Salem and Brown.[13]

The variety of FP chromophores can be increased further by replacing the canonical amino acid Tyr by a so-called non-canonical amino acid (nCAA), including the ones developed by Liu and Schultz.[140] Some experimental examples of this approach are the gold fluorescent protein[171] and the red-shift green FPs.[172] On the theoretical side, Salem and Brown computationally studied a set of chromophores made from nCAAs, i.e., non-canonical chromophores (nCC) depicted in Figure 3.2.[12]

RFP-like Non-Canonical Chromophores
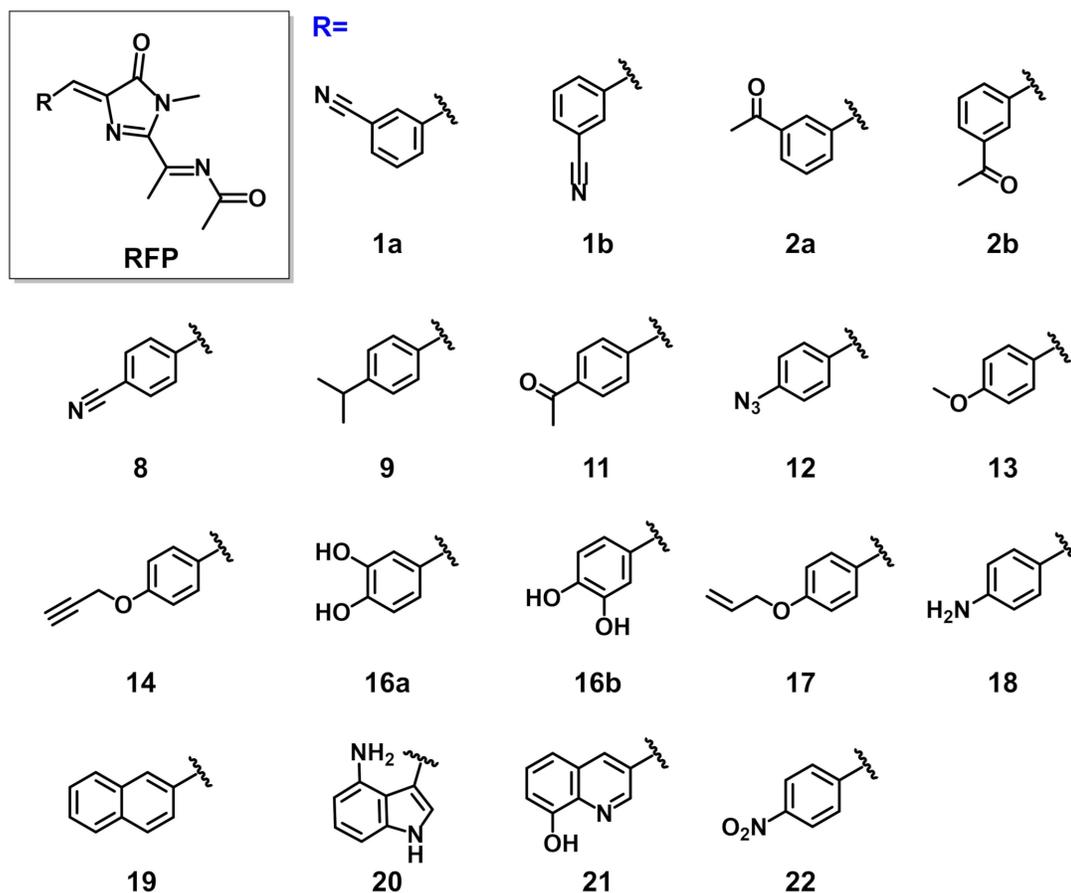
Figure 3.2. A set of red fluorescent protein-like non-canonical chromophores (RFP-like nCCs) studied in this work, part of a more extensive set, previously studied computationally by Salem et al. [14] The template of the RFP chromophore used in the RFP-like nCC models is included in the top-left corner. Labels were kept the same as in the previous work for ease of comparison.

The introduction of nCAAs in the design of FP chromophores ultimately aims to aid in the engineering of novel FPs with improved photophysical features in comparison with the existent FPs built from the set of canonical amino acids. The photophysical properties of FPs lie principally in the chromophore. In turn, the absorption and fluorescence properties of the chromophore are influenced largely by the protein matrix. Over the last decade, several efforts have been focused on describing the important role of the protein on the FPs' photophysical features. [15,26,83,85,86] While computations on the bare chromophore [9,11,13,60] will miss the important influence of the protein, they are still a

useful benchmark to examine the performance of a computational method for obtaining 2PA cross-sections. (Note that throughout the present study, as well as references to previous work, the focus is on degenerate 2PA, where both photons are of the same wavelength).

Two-photon absorption cross-sections in FP's chromophores have been computed using a diverse set of computational methods, including CC2, time-dependent density-functional theory (TD-DFT),[9,11,13,60] and using the time-dependent density functional tight binding (TD-DFTB) method.[61] The TD-DFTB method has been used widely in the study of photophysical properties of a diverse range of systems, including polyacenes and conjugated organic systems,[133,173] as well as inorganic clusters and nanoparticles.[174]

TD-DFTB is considerably less computationally expensive than TD-DFT and semi-empirical methods, including PM6[175],[134] while its performance agrees well with that of TD-DFT. Although, TD-DFTB has been useful in the computation of excitation energies, absorption spectra, and transition dipole moments,[139] it has inherited some of the pitfalls of TD-DFT in describing excited states, for instance, in some cases failing to obtain the proper ordering of the electronic transitions.[173] In the case of FPs, TD-DFTB has been used almost exclusively for their structural analysis.[108] Just in a handful of cases, TD-DFTB has been used for computing the 1PA features of FPs and FP chromophores, e.g., vertical excitation energies (VEE). VEEs previously obtained for FPs using TD-DFTB deviate from experiment up to 1 eV.[106,107]

In a previous investigation, discussed in Chapter 2, we evaluated the performance of the second-order approximation of TD-DFTB (TD-DFTB2) in the computation of 2PA cross-sections (σ^{2PA}) of FP nCCs.[61] The latter analysis was carried out using the TD-DFTB2 linear response implementation of the DFTB+ code.[176] Since second- and higher-order response theory implementations of TD-DFTB2 are not available yet, σ^{2PA}s were computed using the two-level model (2LM) within the sum-over-states approximation (SOS).[145,154,177] The values we obtained for σ^{2PA} using TD-DFTB2 deviated signifi-

cantly from the results previously reported by Salem and Brown.[12,14] As we concluded in Chapter 2, the main reason behind such a deviation is the systematic overestimation of the excited state permanent electric dipole moments. Also in our previous work, as a possible way to improve the σ^{2PA} results, we suggested using the long-range corrected time-dependent density-functional tight binding method (LC-TD-DFTB) implemented in the DFTBaby code.[111,170]

LC-TD-DFTB, as implemented in the DFTBaby code,[111] has been employed previously for the study of excited-state dynamics of chromophore aggregates,[178] exciton formation,[179] optical properties of diverse organic molecules,[180,181] and the study of organic electronics systems.[182] These results suggest that LC-TD-DFTB is a promising method for the study of the dynamics and excited state properties of conjugated organic molecules.[182] LC-TD-DFTB has also been implemented and tested with use of the Polarizable Continuum Model to treat solvation effects.[183] Excited state properties, such as VEEs using LC-TD-DFTB implementations (including those outside DFTbaby), have been shown to be in good agreement with results using TD-DFT methods, such as TD-CAM-B3LYP.[183,184] The performance of LC-TD-DFTB will depend on the elements comprising the studied systems as well as their chemical environment due to the parametrization required. Thus, there can be inconsistency in the deviation from results of more accurate methods, as has been observed for the ground- and excited state structural parameters.[185]

Herein, the LC-TD-DFTB method is used for the computation of σ^{2PA} for a set of canonical and non-canonical FP chromophores previously studied.[13,14] The σ^{2PA}s were computed using the 2LM and compared against TD-DFT and TD-DFTB2 results. The systems employed here include the canonical chromophores of the cyan, blue (neutral), green, and blueberry fluorescent proteins (CFP, BFP_1, GFP_n, and BLB, respectively),[13] as well as eighteen non-canonical chromophores of the red fluorescent protein (RFP) type.[14] Their corresponding structures are provided in Figures 3.1 and 3.2. The set

of non-canonical chromophores studied in this work, as in the case of the TD-DFTB2 computations discussed before, is a subset of the RFP-like non-canonical chromophores studied previously.[13] Bear in mind that such a contraction of the original set of non-canonical chromophores[13] had to be done due to the unavailability of mio parameters for all the atoms contained in the chromophores, such as halogens.

3.2   Computational Methods

All properties employed for the calculation of the corresponding σ^{2PA} were computed using DFTBaby.[111,170] These include vertical excitation energies (VEE), ground- and excited state permanent electric dipole moments, and transition electric dipole moments. The results presented in this chapter were obtained using the parameter set provided in the DFTBaby code. A broader comparison of the results presented here could have been done using the mio[124,141] parameters set, which was employed in the computations using TD-DFTB2;[122] however, using the available version of DFTBaby, the x, y, and z components of the electric dipole moments are not returned by the code when Mulliken transition charges are used to compute the oscillator strengths. Even though the DFTBaby source code is available, we did not attempt to fix the latter bug because the modification falls outside the scope of the present work. Specifically, what we found is that the "Dipole Matrix Elements (in Debye)" are not returned by the DFTBaby code when the input keyword `--oszis=mulliken` is chosen. The "graphical output" could not be employed either since the required libraries are outdated for (the language of the source code) Python v.2.

All σ^{2PA}s were calculated in GM units (or 10^{-50} cm^4 s molecule^{-1} photon^{-1}) according to

$$\sigma^{2PA} = \frac{N\pi^2 a_0^5 \alpha}{c} \frac{\omega^2}{\Gamma} \delta^{2PA}. \tag{3.1}$$

In Equation 3.1, a_0 is the Bohr radius, c the speed of light, α the fine structure

constant, and ω the photon energy (half the excitation energy). The integer value N and the broadening factor Γ (HWMH) are set to 4 and 0.1 eV, respectively, to ease comparison with experiment[60] and the previously reported computational results.[13,14] The 2PA transition moment probability, δ^{2PA}, is given by[145,154,177]

$$\delta^{2PA} = F\delta_F + G\delta_G + H\delta_H \tag{3.2}$$

where the coefficients for polarized light are $F = G = H = 2$ and the elements $\delta_{F,G,H}$ are defined as

$$\delta_F = \sum_{\alpha,\beta} S_{\alpha\alpha}S^*_{\beta\beta}, \qquad \delta_G = \sum_{\alpha,\beta} S_{\alpha\beta}S^*_{\alpha\beta}, \qquad \delta_H = \sum_{\alpha,\beta} S_{\beta\alpha}S^*_{\beta\alpha}. \tag{3.3}$$

with α and β being the axis x, y, or z. For the case of the two-level model (2LM), the transition matrix elements $S_{\alpha\beta}$ can be defined as in Equation 2.2.

The structures of the selected set of chromophores studied here are depicted in Figures 3.1 and 3.2. All computations were performed on the optimized structures reported previously.[13,14] Those structures that were charged or contained sulfur were excluded due to restrictions in the DFTBaby code and missing parameters. In this work, both canonical chromophores and nCCs were included, while in our previous work,[61] see Chapter 2, only nCCs were studied.

3.3   Results and Discussion

The computed σ^{2PA} of CFP, BFP_1, GFP_n, and BLB chromophores are compared in Table 3.1 with previous results obtained within quadratic response theory (QRT) using TD-CAM-B3LYP/6-31+G(d,p)[13] and experimental data. LC-TD-DFTB information corresponding to those transitions S_n for which the oscillator strength (OS) is different from zero (or approximately 0) is included in Table 3.1.

Table 3.1. For the Set of Canonical Chromophores Employed, Vertical Excitation Energies (VEE) and σ^{2PA} Computed Using LC-TD-DFTB Within the 2LM. For Comparison, Previously Reported Values of VEE and σ^{2PA} Computed Using TD-CAM-B3LYP are Included.[13] LC-TD-DFTB Excitation Energies for S_1 are Given in Between Parenthesis. All Computed Values are in Vacuum

Chromo	S_n*	MOs†	VEE (eV)			σ^{2PA} (GM)		
			TD-DFT	LC-TD-DFTB	Exp.‡	TD-DFT	LC-TD-DFTB	Exp.‡
BFP$_1$	2	HOMO → LUMO	3.70	3.88 (3.19)	3.31	3	0.2	12
CFP	2	HOMO → LUMO	3.54	3.80 (3.22)	2.89	8	1.1	23
GFP$_n$	2	HOMO → LUMO	3.72	3.80 (3.13)	3.11	5	5.2	36
BLB	3	HOMO → LUMO	3.15	3.39 (2.81)	3.05	30	55.9	4

*Lowest transition for which OS>0.

† The transition with the largest contribution to the S_n.

‡ Average of experimental measurements reported by Drobizhev et al.,[5] as in Salem and Brown's work.[13]

In all cases, canonical and non-canonical chromophores (Table B.2), the lowest excited state S_1 exhibits an OS equal to 0 or ≈ 0. The latter might indicate the prediction of a non-physical low-lying state by the LC-TD-DFTB method. The wrong prediction of dark states is a known limitation of the method.[61] As for the nature of these states, S_2 and S_3 (for the case of BLB) correspond to the transition HOMO→LUMO, which has been observed to have the largest σ^{2PA}.[9,12–14]

VEEs computed for the canonical chromophores using LC-TD-DFTB (Figure 3.1) are 0.08–0.26 eV larger than those previously obtained using TD-CAM-B3LYP. In comparison with experiment, they are 0.24–0.75 eV larger. In addition, they do not seem to reflect the structural differences of the chromophores; in two out of the four cases, CFP and GFP$_n$, the energies are the same (3.80 eV). In the case of the nCCs, the VEEs for S_2 are also larger than what was reported previously for TD-CAM-B3LYP (Table B.2). Overall, σ^{2PA} results obtained using TD-CAM-B3LYP are underestimated by LC-TD-DFTB by ~1 order of magnitude in comparison to TD-DFT and experimental data, except BLB, whose results are similar to what was obtained before using TD-CAM-B3LYP. BLB's σ^{2PA} using LC-TD-DFTB is noticeably larger than the cross-sections of the rest of canonical chromophores, almost twice (55.9 GM) of what was estimated before using TD-DFT (30 GM) and greater than the experimental reference.

As for the set of nCCs, the σ^{2PA}s obtained using LC-TD-DFTB tend to overestimate

the reported results using the 2LM approximation and QRT with TD-CAM-B3LYP[12,14] and those computed using the 2LM and TD-DFTB2[61] (Figure 3.2).
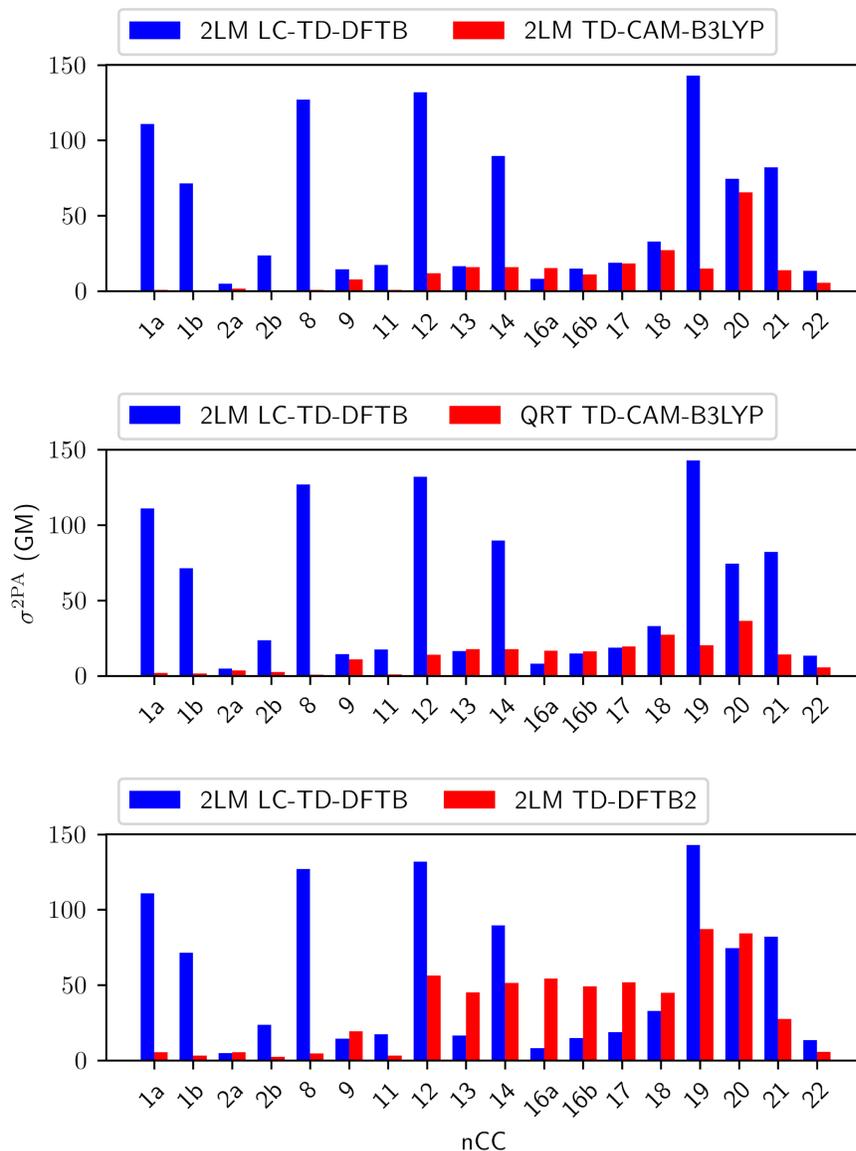
Figure 3.3. Graphical representation of the σ^{2PA} computed using the LC-TD-DFTB implementation in DFTBaby code and results previously obtained using (i) TD-CAM-B3LYP/6-31+G(d,p)[14] within the 2LM (top), (ii) TD-CAM-B3LYP/6-31+G(d,p) within QRT approach (middle),[14] and (iii) TD-DFTB2 within 2LM (bottom).[61] The nCC models can be found in Figure 3.2.

The overestimation of the cross-section does not occur in all the models, in fact, some (8 out of 18) of the LC-TD-DFTB results are in reasonable agreement with the benchmark. However, the overestimation of σ^{2PA} among the rest of models is significant

and lacks any correlation between the overestimation extent and specific chemical aspects of the chromophores.

To get a better understanding of the reasons behind the disagreement between benchmark data and the results obtained in this work, the σ^{2PA} of the nCC set are examined in terms of their 2LM components (Equation 2.2) in a similar fashion to our previous work.[61] The corresponding VEE, OS, ground- and excited state permanent electric dipole moments (μ_{00} and μ_{22}, respectively), transition electric dipole moments (μ_{02}), and the character of the transition are included in the Appendix B (Table B.1).

Among all the 2LM components in Table B.1, the computed excited state permanent dipole moments (μ_{22}) are the ones that deviate the most (mostly overestimated) from the benchmark data, as was observed in our previous investigation on TD-DFTB2 discussed in Chapter 2.[61] In fact, the linear correlation (Figure B.1) between the μ_{nn} values computed using LC-TD-DFTB with respect to the reported results employing TD-CAM-B3LYP and TD-B3LYP is even poorer ($R^2 = 0.04$ and 0.09, respectively) than that for TD-DFTB2 and the same TD-DFT methods ($R^2 = 0.28$ and 0.16, respectively).

In contrast with TD-DFTB2 and somehow unexpected, some of the ground state dipole moments (μ_{00}) deviate from the reference data (Figure B.2). The latter seems to be the reason behind the large deviation of the σ^{2PA} (Equation 2.2) in models 8, 12, 14, 19, and 21 with respect to the TD-DFT results (Figure B.2). For example, μ_{00} for the nCC 1a is 17.1 and 2.34 au for LC-TD-DFTB and TD-CAM-B3LYP, respectively. The VEEs and μ_{02} are in reasonable agreement with their TD-DFT analogues, but the μ_{00} and μ_{22} exhibit large deviations; therefore, the latter two 2LM elements are most likely the reason behind the large overestimation in the σ^{2PA} as compared to the TD-CAM-B3LYP and TD-B3LYP results (Figures B.1 and B.2).

LC-TD-DFTB results also were compared with those computed previously using TD-DFTB2. Overall, the performance of TD-DFTB2 in the computation of σ^{2PA} (and dipole moments μ_{00} and μ_{nn}) is still better than LC-TD-DFTB (Figures B.1 and B.2).

The 2PA cross-sections computed through TD-DFTB2 do not vary abruptly with respect to TD-DFT results as LC-TD-DFTB does (although the agreement is not optimal, as discussed previously in Chapter 2.[61])

3.4   Conclusions

The DFTBaby code[111,170] includes a series of features, such as the option to employ a different parameter set, e.g., the mio parameter set in lieu of the default one. However, it was not possible to test such tools as the source code seems to have bugs and modifying it is out of the scope of this work. Although TD-DFTB2 performs better than LC-TD-DFTB in the computation of σ^{2PA} for the set of nCCs in Figure 3.2, neither of these methods are recommended to be used in the computation of 2PA over standard TD-DFT approaches using range-separated hybrid functionals, such as TD-CAM-B3LYP.[60] Overall, LC-TD-DFTB requires further testing, parameter development, and refinement in its implementation before it can be used as a standard tool for 2PA. The strengths of the method lie in the computation of 1PA properties, such as vertical excitation energies and oscillator strengths. However, care must be taken, as low-lying non-physical transitions might arise, as was demonstrated here.

Chapter 4

Three-photon Absorption Cross-sections of Serotonin, Fluorescein, Rhodamine 6G, and Some Fluorescent Protein Chromophores

4.1 Introduction

Nowadays, multi-photon absorption (MPA) of inorganic, organic, and biochemical materials is exploited in clinical imaging and neuroscience such as in monitoring the brain activity in vivo.[186–188] Degenerate MPA, where all photons have the same wavelength, consists of the simultaneous absorption of N photons of X wavelength, resulting in an electronic excitation of X/N wavelength, i.e., MPA involves N photons of lower energy than the resulting excitation energy. Clinical and in vivo imaging are performed ideally within the near infra-red (NIR), the so-called clinical window ($\sim$650–900 nm),[20,187,189] where the light-absorption of water and other tissue components are minimal. Other studies have determined that deep-tissue brain imaging is optimal between 1600 and 1870 nm.[190] Two- and three-photon absorption (2PA and 3PA, respectively), make the clinical windows accessible, while the chances of photo-damage are decreased and deeper penetration in the tissue is achieved.[187,191,192] In this sense, it is fair to say that MPA plays a significant role in the development of less invasive imaging procedures.

Fluorescent proteins (FPs)[1,2] are a series of proteins composed of a chromophore embedded within a β-barrel protein. Existent FPs have been grouped in families, depending on the structure of their chromophore. Currently, the light absorption and emission of the available FPs span over the entire visible spectrum. The structures for a curated set of FP chromophores are included in Figure 4.1. FPs are among the most widely used tools in clinical imaging and probing, in particular because of their 2PA properties.[4] Extensive experimental and computational efforts have been made on engineering novel

FPs and improving the 2PA performance of those already known.[59,80]
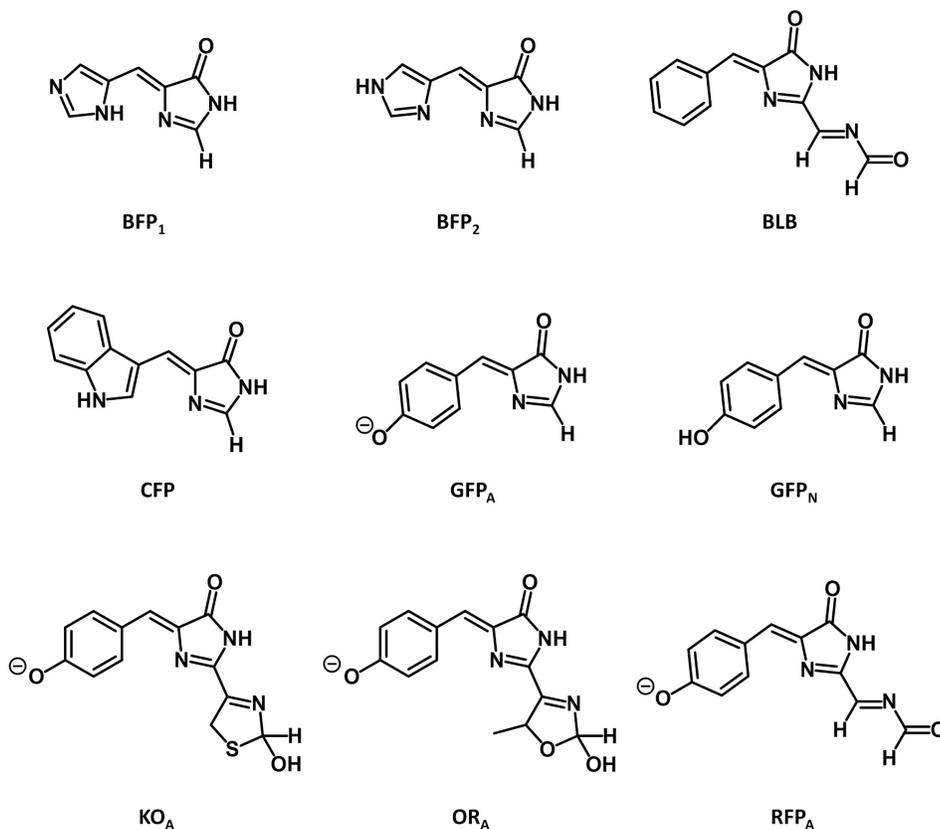
Figure 4.1. Curated set of fluorescent protein chromophores previously studied by Salem and Brown.[13]

Although 2PA is still one of the most exploited MPA techniques for FPs, 3PA promises to be an equally useful alternative.[193] Three-photon absorption increases the signal-to-background ratio (SBR), while using excitation photons of longer wavelength that allow deeper-tissue imaging than 2PA.[191] Measured σ^{3PA}s suggest that FPs are suitable tools for 3PA microscopy.[191,194,195] The 1PA and 2PA properties of FPs have been studied using experimental[5,88,196–198] and computational means.[26,43,85] Experimentally, Cheng et al. previously determined the σ^{2PA}, σ^{3PA}, and σ^{4PA} of the wild-type green fluorescent protein (GFP),[191] while Liu et al. and Deng et al. determined the σ^{3PA} for a group of red fluorescent proteins (RFPs).[194,195] Computationally, FPs' excitation energies and σ^{2PA} have been reported using wave-function[50] and time-dependent density func-

tional theory (TD-DFT) methods.[9,11,13] Moreover, different computational approaches have been used, including quantum mechanical (QM), quantum mechanical/molecular mechanical (QM/MM), and combinations of the latter with molecular dynamics simulations. The computation of MPA strengths (δ^{MPA}) up to 4PA (i.e., δ^{4PA}) have been reported for the neutral and anionic forms of GFP.[26] In particular, Friese et al.[199] observed that the choice of the functional has a larger impact on determining σ^{3PA} than the choice of the basis set. Among the BLYP,[142,143] B3LYP,[155] and CAM-B3LYP[184] functionals, the results obtained using CAM-B3LYP[184] vary the least in Friese et al.'s investigation.[199] Overall, for small molecules, TD-DFT has exhibited good performance in 3PA response theory computations when compared to coupled-cluster methods.[200] In particular, CAM-B3LYP[184] has shown a comparable performance to CC2 in the computation of σ^{2PA},[15,199] thus, it is reasonable to think that it can be useful for the computation of higher-order MPA features.

Xu and Webb showed through the single-intermediate-state (SIS) approximation that the order of magnitude of σ^{3PA}s values of a given molecule is $\sim 1 \times 10^{-82}$ cm^6 s^2 photon^{-2}.[201] Experimentally, σ^{3PA} for the wild-type GFP was determined to be 15.9×10^{-84} cm^6 s^2 photon^{-2} at 1300 nm.[191] In the case of some organic compounds, experimental σ^{3PA} can be as large as 8.8×10^{-76} cm^6 s^2 photon^{-2}.[202] Computationally, Nobis et al.[203] determined, at the TD-CAM-B3LYP/cc-pVDZ//B3LYP/cc-pVTZ level of theory, that the σ^{3PA}s for different excited states (S_n) of an emissive derivative of uridine are about 1×10^{-82} cm^6 s^2 photon^{-2}. Interestingly, Nobis et al. found that experimental 3PA cross-sections correspond to excitation to S_5 and S_6, and not S_1. Previously, 3PA strengths (δ^{3PA} in Equation 4.1) of the order of 1×10^6 au, were computed using the CAM-B3LYP[184] functional and the 6-31+G(d) basis set for the anionic and neutral forms of the green fluorescent protein (GFP) by Steindal et al.[26] Similarly to the case of 2PA investigations,[26] the comparison of computational data for σ^{2PA} to experiment has a number of challenges. Since MPA can occur at higher transition energies than

1PA,[5] the experimental 3PA absorption and excitation spectra will be required for the appropriate comparison between computations and experiment. However, in the context of FPs, experimental σ^{3PA} as well as absorption and excitation spectra are scarce in the literature.[26,191]

Fluorescein and rhodamine 6G commonly have been used as calibration dyes (depicted in Figure 4.2) in experimental MPA measurements of FPs.[5,191,204–206] In particular, fluorescein derivatives have been investigated as dyes in dye-synthesized solar cells,[207] whereas serotonin, also known as 5-hydroxytryptamine or 5-HT, is a neurotransmitter associated with changes of mood and mental diseases. In this context, serotonin has been studied as an intrinsic probe.[208,209] Previously, Jha et al. carried out a systematic investigation on the σ^{2PA} of the dyes Bis-MSB, Coumarin 307, fluorescein, and Rhodamine B in an effort to set a basis for comparisons between computational data and experiment.[210] In this chapter, we report the σ^{3PA} computed for the set of chromophores given in Figure 4.1. In addition, we report the computed σ^{3PA} for dianionic (2-) fluorescein, neutral and cationic (1+) serotonin, as well as for cationic (1+) rhodamine 6G. These probes were chosen because their reported σ^{3PA} provide a benchmark for comparisons between computations and experiment beyond FP chromophores. Here, we aim to evaluate the viability of computational methods in obtaining the σ^{3PA} of small molecules and FP chromophores. The present results also contribute to the creation of a theoretical database of 3PA cross-sections for FPs and dyes used in MPA microscopy.

Figure 4.2. "Through the"eyes" of rhodamine 6g". A cartoon depicting the use of rhodamine 6G as a reference in the computation of 3PA cross-sections of FPs.

4.2 Computational Methods

4.2.1 Geometry Optimization

For fluorescein, rhodamine 6G, serotonin, as well as the set of FP chromophores illustrated in Figure 4.1, we computed the vertical excitation energies and three-photon absorption cross-sections (σ^{3PA}) corresponding to the five lowest energy excited states. Since a fluorophore's absorption and emission properties may depend strongly on its protolytic state and the solvent in which it is immersed,[211–213] specific protolytic states of fluorescein, rhodamine 6G, and serotonin were employed here. Experimental studies of fluorescein, mainly in the multi-photon absorption realm, have been carried out for particular solvents and pH values. In the case of 2PA cross-section measurements, Drobizhev et al.[5] employed solutions of fluorescein at pH 11. At this pH and at $\sim$7 (the physiological pH),[211,214] the dianionic form of fluorescein dominates over other protolytic species.[214] Also, at pH 11, fluorescein exhibits the largest 1PA[211] and the largest fluorescence.[215] Experimental measurements made on rhodamine 6G (also known as rhodamine 590) commonly have employed methanol solutions[5,204,216] that are unlikely to alter its protolytic state (cationic with charge 1+). In the case of serotonin, which has been used as an intrinsic probe of neurotransmitters transport within the brain, its protolytic state depends on the cellular process in which it is involved.[217–219] In solutions

55

with pH $<$ 10, which includes the physiological pH, the cationic 1+ form of serotonin predominates.[209] Here, both the neutral and the 1+ forms of serotonin were studied.

The geometry optimizations of dianionic (2-) fluorescein (referred to as fluorescein^{2-} from now on in the text) as well as both neutral and 1+ forms of serotonin (5-HT and 5-HT$^+$, respectively) were performed in vacuum. For these geometry optimizations, we employed the hybrid Perdew, Burke, and Ernzerhof exchange-correlation density functional (PBE0)[220] and the 6-31+G(d,p) basis set,[156–158,160,221] analogously to the optimization of certain FP chromophores.[13] For the optimization of fluorescein, we used the structure reported in the study carried out by Gerasimova et al. at the B3LYP/aug-cc-pVDZ/PCM(H$_2$O) level of theory[222] as the initial guess structure. Results obtained by Tamulis et al. suggest little structural differences using different basis sets in the optimization of fluorescein,[215] while the selection of functional and basis set can have a considerable impact on the excited state computations for the different protolytic forms of fluorescein.[222] The geometry optimization of the molecules and the corresponding Hessian analysis were carried out in Gaussian16[223] using the default convergence criteria. In particular, the direct geometry optimization of R6G$^+$ using the PBE0/6-31+G(d,p) method failed to converge despite our various efforts. Thus, the B3LYP[155] and CAM-B3LYP[184] functionals along with the 6-31+G(d,p) basis set were used for the geometry optimization of R6G$^+$. The resulting structures were used for the computation of the corresponding excited state properties (Tables C.5, C.6, C.11, and C.12). The optimized structure of R6G$^+$ using the B3LYP[155] functional was submitted to a geometry optimization using the PBE0/6-31+G(d,p) method, however, the resulting structure could not be verified as a true minimum because the Hessian analysis did not converge. Nevertheless, the Cartesian coordinates we obtained are provided in Appendix C. The fluorescent protein chromophore structures (Figure 4.1), optimized using the PBE0/6-31+G(d,p) method, were obtained from Salem and Brown's previous work.[13]

A final note to this section. Initially, a two-step optimization strategy using PM6[175] and PBE0/6-31+G(d,p) was adopted for the geometry optimization of dianionic (2-) fluorescein, serotonin (neutral and cationic [1+]), and cationic (1+) rhodamine 6G. The resulting minimum energy structures most likely correspond to local and not global stationary points of these systems. The structures and the corresponding one-photon absorption features of cationic (1+) fluorescein and cationic (1+) rhodamine 6G significantly disagreed with experimental and computational reports. The origin of such disagreement was the introduction of PM6[175] in the optimization process, as we later concluded. We performed relaxed scans around the dihedral angle θ_F in Figure 4.3 (results not shown here) using the PM6[175] and B3LYP/6-31+G(d,p) methods. The results of these scans showed that the PM6[175] method predicted the B3LYP/6-31+G(d,p) global minimum to be a local minimum instead.

4.2.2   One- and Three-photon Absorption

One- and three-photon absorption computations were carried out in vacuum using time-dependent density-functional theory (TD-DFT) within linear- and cubic-response theory, respectively. All computations were carried out using the Dalton2019.alpha (2019)[224] software package. For all the excited state computations, we employed the long-range corrected functional CAM-B3LYP,[184] along with different basis sets: 6-31+G(d), 6-31+G(d,p),[156–158,160,221] jun-cc-pV(D+d)Z,[225,226] jun-cc-pV(T+d)Z, aug-cc-pVDZ,[227] aug-cc-pVTZ, pcseg-1,[228–230] pcseg-2,[228–230] aug-pcseg-1,[231] and aug-pcseg-2. In the case of fluorescein and serotonin, the spherical and cartesian forms of the Pople-type basis sets were employed. For the rest of the cases, only the spherical form was used such that the frontier molecular orbitals could be visualized in the analysis of results. The basis sets used for each particular molecule are provided in Table 4.1.

Table 4.1. Basis Sets Employed in the 1PA and 3PA Computations for Fluorescein (F^{2-}), Neutral Serotonin (5-HT), Cationic Serotonin (5-HT$^+$), Cationic Rhodamine 6G (R6G$^+$) and the Fluorescent Protein Chromophores (FP chromophores)

	F^{2-}	5-HT	5-HT$^+$	R6G$^+$	FP Chromophores
6-31+G(d)	x	x	x		
6-31+G(d,p)	x	x	x	x	x
jun-cc-pV(D+d)Z	x	x	x		
jun-cc-pV(T+d)Z	x				
aug-cc-pVDZ	x	x	x	x	x
aug-cc-pVTZ	x	x	x		
pcseg-1	x				
pcseg-2	x				
aug-pcseg-1	x	x	x		
aug-pcseg-2	x				

The basis sets 6-31+G(d,p),[156–158,160,221] jun-cc-pV(d+d)Z,[225,226] jun-cc-pV(T+d)Z, pcseg-1,[228–230] aug-pcseg-1,[231] and aug-pcseg-2 were obtained from the New Basis Set Exchange.[226] For all the FP chromophores and other molecules, the microscopic 3PA probabilities (δ^{3PA}s) were computed for the first five lowest energy transitions. The δ^{3PA}s also were computed for the fluorescein structure previously obtained by Gerasimova et al. at the B3LYP/aug-cc-pVDZ/PCM(H$_2$O) level of theory.[222]

Three-photon cross-sections (σ^{3PA}) were obtained from the corresponding δ^{3PA} according to[60,154,200,203,232]

$$\sigma^{3PA} = \frac{N\pi^3 a_0^8 \alpha}{3c^2} \frac{\omega^3}{\Gamma_f} \delta^{3PA} \tag{4.1}$$

for linear polarized light and a Lorentzian lineshape function. Here N is set to be 4, a_0 is the Bohr radius, α is the fine structure constant ($1/137$), and Γ_f is the broadening factor, assumed to be half width at half maximum (HWHM) and set to 0.1 eV, in a similar fashion to σ^{2PA} computations.[13] Notice that ω in Equation 4.1 is the photon energy in a 3PA process, i.e., $1/3$ of the vertical excitation energy (VEE). Within this scheme, the resulting σ^{3PA} are expressed in units of cm^6 s^2 photon^{-2}.

4.3   Results and Discussion

The optimized structures of fluorescein^{2-}, 5-HT, and 5-HT$^+$ using the PBE0/6-31+G(d,p) method in vacuum, as well as the optimized structure of R6G$^+$ using the B3LYP/6-31+G(d,p) method are shown in Figure 4.3. Their corresponding Cartesian coordinates are provided in Appendix C.

a) Serotonin neutral

$\theta(C_{1-4}) = 103.3\,°$

$\theta(C_2 - N_5) = -62.7\,°$
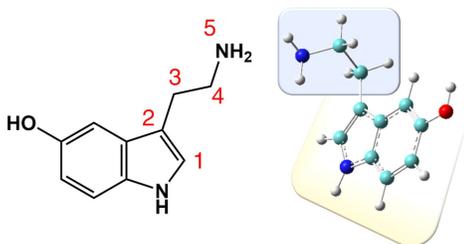
c) Fluorescein 2-

$\theta_F = 96.5\,°$
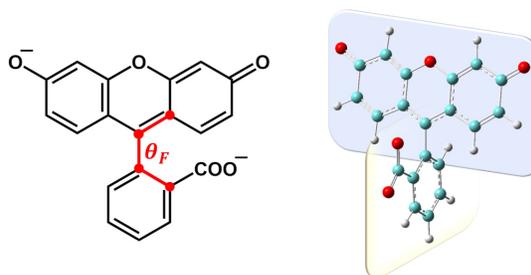
b) Serotonin 1+

$\theta(C_{1-4}) = 126.7\,°$

$\theta(C_2 - N_5) = -38.7\,°$
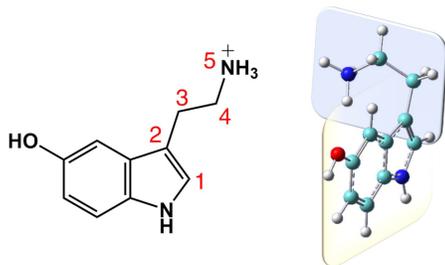
d) Rhodamine 6G 1+

$\theta_{R6G} = 92.5°$
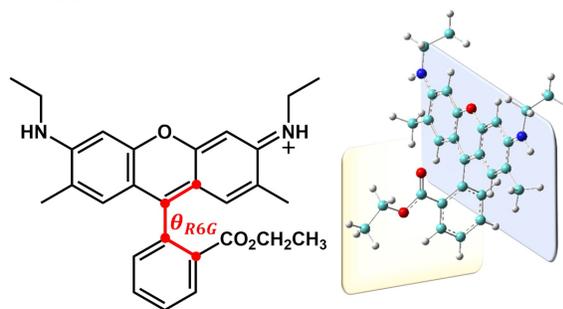
Figure 4.3. Optimized structures of the neutral and cationic (1+) forms of serotonin, dianionic (2-) fluorescein, and cationic (1+) rhodamine 6G, using the PBE0/6-31+G(d,p) [B3LYP/6-31+G(d,p) for rhodamine 6g] method in vacuum. In each case, the most representative dihedral angle(s) is (are) given.

4.3.1   Fluorescein

The fluorescein structure (Figure 4.3c) we obtained is distorted 6° from the structure obtained by Gerasimova et al.[222] and in similar agreement with the structures reported in other computational investigations.[215,233] The reader may refer to Figure C.7

for a qualitative description of the mentioned distortion. The dihedral angle θ_F (Figure 4.3) of the fluorescein structure that we obtained is 96.5°, close to what has been reported previously, 90.3° and 110.0°, using the B3LYP/aug-cc-pVDZ/PCM(H_2O)[222] and B3LYP/TZVP/IEFPCM(H_2O)[233] methods, respectively. Hence, the effect of (implicit) solvent is modest on the structure determined.

Overall, the VEEs obtained for the transition $S_0 \rightarrow S_1$ of fluorescein^{2-} are up to 0.6 eV larger than the experimentally reported 1PA maximum, 2.5 eV (490 nm)[204,211,214,234,235] (Table 4.2). The VEEs closest to experiment for the latter transition are obtained using the CAM-B3LYP[184] functional with the aug-cc-pVTZ (3.063 eV), aug-cc-pVDZ (3.080 eV), and pcseg-2[228–230] (3.104 eV) basis sets. The aug-cc-pVDZ basis set has the best accuracy to computational cost ratio. Using CAM-B3LYP/aug-cc-pVDZ, the transition S_1 (OS=0.72) between the HOMO and LUMO+2 orbitals (Figure C.2) is of $\pi \rightarrow \pi*$ character and takes place mainly within the xanthene ring, as described in previous investigations using the TD-B3LYP/6-311++G(d,p) method.[211] The nature and the energy with respect to S_1 of the second representative transition in the spectrum of the dianion (2-) fluorescein, a shoulder experimentally located at $\sim$470 nm (2.64 eV) and computationally determined at 407 nm (3.05 eV) by Batistela et al., is better described by smaller basis sets than aug-cc-pVDZ (Figure C.1 and Table C.1). As the basis set limit is approached among the basis sets, including diffuse functions, the agreement between the excitation energies and the experimental absorption maximum is improved. However, the basis sets with the largest number of diffuse functions seem to lead to problems for the case of dianionic (2-) fluorescein (Figures C.2 and C.3). For example, the VEE computed using the basis set aug-pcseg-2 is 0.2 eV closer to experiment than aug-cc-pVDZ; however, this basis set and aug-pcseg-1 seem to miss the energy order of the transitions and introduce a number of low-lying Rydberg-type orbitals in comparison with the rest of the basis sets. In comparison with previous computational investigations,[210,211,222,233] our results for the VEEs are slightly farther away from the experimental results. The de-

viation of our results from previous reports is most likely due to the absence of solvation effects during the geometry optimization and the excited state computations. Unfortunately, solvation effects using PCM are not (yet) available for 3PA, which is why they are not considered here. For instance, the VEE of S_1 computed with CAM-B3LYP/aug-ccpVDZ in vacuum has an error of 17%, while Batistela et al. report errors as low as 10% at the B3LYP/LANL2DZ/IEFPCM(H_2O)[211] level of theory and <1% at the TD-B3LYP/aug-cc-pVDZ/non-equilibrium PCM(H_2O)//B3LYP/aug-cc-pVDZ/non-equilibrium PCM(H_2O)[222] level of theory. Here, it is worth commenting that B3LYP[155] displays a better performance in estimating 1PA values of F^{2-},[222,233] however, this functional is not recommended for the determination of σ^{2PA}.[13,60] On the other hand, CAM-B3LYP[184] tends to overestimate experimental excitation energies; however, such a shortcoming is compensated by a better performance in computing MPA response properties in comparison with other functionals.

Fluorescein two- and three-photon absorption cross-sections have been determined using many experimental methods (Table C.13). Previously reported σ^{3PA} vary by up to one order of magnitude,[235] however, most of the measurements are within 1×10^{-83} 3PA units. On average, fluorescein's σ^{3PA} is 2×10^{-83} cm^6 s^2 photon^{-2} at the $\pi \rightarrow \pi^*$ transition of 1395 nm (or the corresponding 1PA, 2.7 eV) (Table 4.2.[191,235] The σ^{3PA}s we obtained for the same type of transition are one order of magnitude smaller than the averaged experimental values. The basis set type and size has a small impact on the value of σ^{3PA} for the lowest energy excitation, as observed in previous investigations computing σ^{3PA}.[199] An increase in the size of the employed basis sets does not translate necessarily into better agreement between the computed results and experiment. Although aug-cc-pVDZ has the best performance in the computation of the VEE, the σ^{3PA} for S_1 obtained using aug-cc-pVDZ (9.97×10^{-85}) and jun-cc-pV(T+d)Z (8.31×10^{-85}) are smaller than the results using the rest of basis sets. According to our computations, except for aug-pcseg-1, S_2 (1PA OS=0.075) has a larger σ^{3PA} than S_1, as in the case of

the experimental 2PA of fluorescein.[204] However, since this transition falls outside of the clinical imaging window, it may be of less practical use than excitation to S_1.

Table 4.2. For the Three Lowest Energy Excited States ($S_{n:1-3}$) of Anionic (2-) Fluorescein, 5-HT and 5-HT$^+$, and Cationic (1+) Rhodamine 6G (R6G$^+$), the Computed Oscillator Strengths (OS), Transition Orbitals (MOs), Vertical Excitation Energies (VEE) in nm, Microscopic 3PA Probabilities (δ^{3PA}) in au, and 3PA Cross-Sections (σ^{3PA}) in cm^6 s^2 photon^{-2}. The Experimental Excitation Energies and σ^{3PA} are Provided for Reference. (NR=Not Reported)

			CAM-B3LYP/aug-cc-pVDZ in vacuum					Exp.
Dye	S_n	OS	MOs	VEE (nm)	VEE (eV)	δ^{3PA} (au)	σ^{3PA}	σ^{3PA}
F^{2-}	1	0.716	HOMO → LUMO+2	403	3.080	3.32E+05	9.97E-85	1.9E-83 (465 nm)
	2	0.024	HOMO → LUMO	380	3.259	3.74E+05	1.34E-84	
	3	0.000	HOMO → LUMO+4	348	3.562	1.60E+06	7.48E-84	
5-HT	1	0.051	HOMO → LUMO+2	271	4.580	5.72E+04	5.64E-85	NR
	2	0.078	HOMO–1 → LUMO+2	254	4.888	6.18E+05	7.50E-84	
	3	0.002	HOMO → LUMO	250	4.954	7.16E+05	8.98E-84	
5-HT$^+$	1	0.010	HOMO → LUMO	281	4.407	1.98E+05	1.75E-84	6.0E-84 (242 nm)
	2	0.074	HOMO → LUMO +1	270	4.592	8.05E+05	8.08E-84	
	3	0.014	HOMO–1 → LUMO	260	4.778	2.44E+06	2.77E-83	
R6G$^+$	1	0.858	HOMO → LUMO	401	3.092	4.55E+06	1.40E-83	
	2	0.000	HOMO–1 → LUMO	334	3.714	1.34E+07	7.01E-83	
	3	0.219	HOMO–2 → LUMO	279	4.444	5.39E+08	4.86E-81	6E-81 (433 nm)

The σ^{3PA}s obtained for the fluorescein structure optimized by Gerasimova et al. are slightly closer to the experimental values and quite similar (6×10^{-85} cm^6 s^2 photon^{-2}) to those obtained for the structure optimized in vacuum using PBE0/6-31+G(d,p) (Tables C.5 and C.6).[222] None of the σ^{3PA}s obtained using Gerasimova et al.'s fluorescein structure are greater than 2×10^{-84} cm^6 s^2 photon^{-2}. The 3PA cross-sections obtained using aug-cc-pVDZ and jun-cc-pV(T+d)Z are 1.35×10^{-84} and 1.69×10^{-84}, respectively. The latter suggests that an optimization using a modest level of theory might be useful to get preliminary insights into the 3PA properties of fluorescein, and maybe, other similar systems.

Modest to significant improvement of the computed VEEs with respect to experimental data can be achieved using solvation models, like the PCM, non-equilibrium PCM, and the Integral-Equation-Formalism Polarizable Continuum Method (IEFPCM).[211,222,233] In the case of multi-photon absorption processes, solvation effects might play a larger role than they do in 1PA, i.e., the 3PA results (σ^{3PA}) reported here for fluorescein might change and, hopefully, improve if these effects are taken into account. Implicit solvent effects are not implemented yet for cubic-response computations. Alternatively, polar-

izable embedding could be used, which might be the subject of a future investigation. The present chapter does not evaluate the environmental effects in any of the 3PA of the molecules employed, including the FP chomophores.

4.3.2 Serotonin

In the case of serotonin, the dihedral angles provided by Lobayan and Schmit[236] were used as an initial structure for the optimization of the 5-HT and 5-HT$^+$ structures. The resulting geometries of 5-HT and 5-HT$^+$ were compared with those described elsewhere, principally for 5-HT.[218,237–242] Here, as has been observed before,[218,236–241] the aminoethyl moiety in both protolytic forms is gauche with respect to the indol ring (Figure C.8), whereas the hydroxyl group has an anti and syn conformation in 5-HT and 5-HT$^+$, respectively (Figure 4.3a and b).

Previously, the order and role of the excited states $\pi \to \pi^*$ and $\pi \to \sigma^*$ in indol and some of its derivatives have been discussed using experimental[237,243] and computational[244,245] approaches. In particular, TD-DFT tends to shift the order of the excited states, placing the energy of $\pi \to \sigma^*$ below $\pi \to \pi^*$ in comparison with CC2 and CASPT2, in which the ordering of the states is reversed.[244,245] Although CC2 and CASPT2 have the best agreement with experimental data in terms of the ordering and energy of the $\pi \to \pi^*$ transitions, it has been suggested that the performance of TD-DFT is suitable to study indol-type systems.[245]

Here, as expected, 1PA computations of 5-HT and 5-HT$^+$ using TD-CAM-B3LYP place the lowest electronic transitions energies close to each other. Also, in the case of 5-HT$^+$, TD-CAM-B3LYP/aug-cc-pVDZ predicts the transition $\pi \to \pi^*$ (4.6 eV) to be 0.2 eV higher in energy than the transition $\pi \to \sigma^*$ (4.4 eV), while their oscillator strengths (OS) are 0.07 and 0.01, respectively (Table 4.2). In a similar fashion to what was obtained for fluorescein, the VEEs computed for 5-HT and 5-HT$^+$ do not experience a strong effect from the basis set choice. Once again, the results obtained with the aug-cc-pVDZ basis set show the best agreement with experiment at a reasonable

computational cost, as previously discussed by Omidyan et al.,[242] in comparison with the rest of basis sets that were employed i.e., 6-31+G(d), 6-31+G(d,p) , jun-cc-pV(D+d)Z, aug-pcseg-1, and aug-cc-pVTZ.

The VEE of the transition $\pi \rightarrow \pi^*$ between the molecular orbitals HOMO and LUMO of 5-HT (S_1 in Table 4.2) at the TD-CAM-B3LYP/aug-cc-pVDZ level of theory, deviates 0.56–0.66 eV from the experimental transition energy determined in vacuum by LeGreve et al., 4.04 eV (307 nm).[237] The same VEE deviates slightly from previous computational results obtained at a higher level of theory.[242,246] The VEE for S_1 in 5-HT is $\sim$0.3 eV larger than that reported using the CC2/aug-cc-pVDZ method[242] and including the explicit interaction between 5-HT and water molecules using the TD-B3LYP/def-TZVP approach.[246] The 5-HT unoccupied molecular orbitals (MOs) obtained using the CAM-B3LYP/aug-cc-pVDZ method are re-ordered with respect to those obtained using basis sets with a fewer number of diffuse functions (Figure C.4), primarily due to the occurence of low-lying Rydberg-type orbitals. In the case of 5-HT$^+$, the VEE we obtained for the transition $\pi \rightarrow \pi^*$ between the molecular orbitals HOMO and LUMO+1 (or S_2 in Table 4.2) also is in good agreement with previous reports.[213,242,247] The transition energy to S_2 is about 0.2 eV larger than the experimental absorption maximum, 4.51 eV (275 nm),[213,247] and 0.4 eV larger than computational results obtained using the CC2/aug-cc-pVDZ method.[242]

Maiti et al. determined the σ^{3PA} at pH = 7 as $\sim$7 $\times$ 10^{-84} cm^6 s^2 photon^{-2}.[248] For a serotonin solution at pH $\approx$ 7, 3PA has been observed at 4.5 eV (277 nm) and has been assigned as a $S_0 \rightarrow S_1$ transition.[249] Here, using the TD-CAM-B3LYP/aug-cc-pVDZ method, we determined a σ^{3PA} of 8 $\times$ 10^{-84} cm^6 s^2 photon^{-2} (HWHM) for the $\pi \rightarrow \pi^*$ transition (to S_2) of 5-HT$^+$. The σ^{3PA}s we obtained using other basis sets vary but are within the same order of magnitude for the same transition.

The agreement between the computed results and experimental data is good, although we predict that it can be improved if solvation effects are taken into account, as in the

case of fluorescein. However, solvation effects are not studied in this work, as has been discussed earlier in this chapter. From the above results, we can suggest that the results obtained from the in vivo studies used as reference in this work most likely correspond to the 5-HT$^+$ form of serotonin. "At physiological pH, the cationic form of extracellular 5-HT is predominantly transported into cells through the plasma membrane 5-HT transporter (SERT), whereas, the neutral form of 5-HT has been suggested to be transported into secretory vesicles through vesicular monoamine transporters (VMATs)."[218] The differences between their VEE values and σ^{3PA} for the two forms of serotonin also suggest that they could be differentiated using their excited state properties.

4.3.3 Rhodamine 6G

In addition to the crystal structure,[250] structural parameters provided for R6G in previous computational studies[251–253] were used as reference for the starting geometry in the geometry optimization of R6G^{1+} in the present work. In the optimized structures using the B3LYP[155] and CAM-B3LYP[184] functionals, the dihedral angle θ_{R6G} between the phenyl and xanthene groups (Figure 4.3) does not vary significantly from previous reports. Here, θ_{R6G} is 92.6° and 92.5° for B3LYP[155] and CAM-B3LYP[184], respectively (Figure C.9). The same angle in the structure obtained using the PBE0 functional is 92.6° (refer to Section 4.2 for further details on the optimization of this structure). In the structure obtained by Watanabe et al. using the B3LYP/6-311++G(d,p) method, it is −92.7°,[251] and −118.3° in the crystal structure reported by Adhikesavalu et al.[250]

The transition to S$_1$ (OS=0.8) computed for R6G$^+$ is of $\pi - \pi^*$ character, in agreement with previous reports.[254] The VEE computed for all S$_n$ are similar between the 6-31+G(d,p) and aug-cc-pVDZ basis sets for each of the two geometries obtained (Tables 4.2, C.5, and C.6). Indeed, the choice of functional for the geometry optimization results in a deviation of <0.1 eV of the VEE to S$_1$. From Kulesza et al.'s[254] contribution, it is observed that TD-DFT largely overestimates (∼0.5 eV) the experimental excitation energies of xanthane-derived dyes in comparison to results from the

CC2 method that deviate less than 0.2 eV in most of the studied cases. Nonetheless, TD-DFT overestimation is fairly consistent among all the results obtained.[254] The VEE to S_1 computed at the B3LYP/6-31+G(d,p)//CAM-B3LYP/6-31+G(d,p) (in vacuum) level of theory is 0.7 eV and 0.4 eV larger than the experimentally obtained values, 2.3 eV (530 nm in ethanol)[234,255] and 2.62 eV (473 nm according to their gas-phase photofragmentation spectrum), respectively.[254] The VEE results we obtained for any S_n using the combination B3LYP//CAM-B3LYP are in good agreement with previously reported computational results.[252–254] Our VEE computations are 0.2, 0.4, and 0.4 eV larger than the results obtained using the TD-CAM-B3LYP/6-31+G(d,p) method and a conductor-like continuum solvation model (CPCM),[253] the results obtained using the TD-B3LYP-D/6-31+G(d,p)/CPCM(acetonitrile) method,[252] and those reported at the CC2/TZVP//B3LYP-D3(BJ)/6-311G(d,p) level of theory,[254] respectively. Furthermore, our results are just 0.05 eV smaller than gas-phase results reported at the ωB97XD/6-311++G(d,p)//B3LYP-D3(BJ)/6-311G(d,p) level of theory.[254]

In R6G$^+$, σ^{3PA} occurs at a higher excitation energy than the 1PA maximum (530 nm).[235] The σ^{3PA} experimentally determined at 1300 nm,[256] 6×10^{-81} 3PA units, most likely corresponds to a higher energy $S_0 \rightarrow S_n$ excitation; the present results suggest the transition $S_0 \rightarrow S_3$ between the orbitals HOMO–2 and LUMO. Here, the σ^{3PA} for the transition to S_3 (4.86×10^{-81}) is in reasonably good agreement with the experimentally measured value (6×10^{-81}). As expected for TD-DFT computations in vacuum for this type of system,[253] the experimental excitation energy to S_3, 3.56 eV,[257] is overestimated by $\sim$0.8 eV. Bear in mind that the magnitude of σ^{3PA} is proportional to the excitation energy (ω) by $(\frac{\omega}{3})^3$, according to Equation 4.1. Thus, the agreement between the computed and experimentally obtained σ^{3PA} is (in part) due to the overestimation of the excitation energies by TD-DFT.

4.3.4 Fluorescent Protein Chromophores

Cheng et al. experimentally determined the 3PA action cross-section, $\eta \times \sigma^{3PA}$, where

η is the fluorescence quantum efficiency, at 1300 nm (2.9 eV in 1PA) of the wild-type GFP as 15.9×10^{-84} cm^6 s^2 photon^{-2}.[191] Liu et al. determined the σ^{3PA} for the RFP tdtomato as 155×10^{-84} and 100×10^{-84} cm^6 s^2 photon^{-2} at 1300 nm and at the 1PA maximum,[5] respectively, as well as measured σ^{3PA} for the red algae fluorescent protein as 414×10^{-84} cm^6 s^2 photon^{-2} at 1700 nm (2.2 eV in 1PA). Adding to the RFP 3PA database, Deng et al. measured the three-photon excitation spectra and σ^{3PA} of eight RFPs (mBanana, mOrange, mStrawberry, eRFP, mRaspberry, mKate, mPlum, and DsRed), where the $\eta \times \sigma^{3PA}$ of DsRed was determined as 32×10^{-84} cm^6 s^2 photon^{-2} at 1620 nm (or 540 nm in 1PA).[195] Because the choice of basis set did not affect significantly the 3PA results in the case of fluorescein and serotonin, 3PA computations of FP chromophores (Figure 4.1) were performed using only two basis sets, 6-31+G(d,p) and aug-cc-pVDZ, with the CAM-B3LYP[184] functional. Excitation energies for these FP chromophore models have not been studied using aug-cc-pVDZ, which has a similar performance to 6-31+G(d,p) in the computation of VEE (see Tables C.1, C.2, C.3, C.4, C.5, C.6, and C.13). The resulting σ^{3PA}s using the CAM-B3LYP/aug-cc-pVDZ method are listed in Table 4.3, along with the corresponding VEEs (nm) and microscopic 3PA probabilities δ^{3PA}. The results using the 6-31+G(d,p) basis set are included in Table C.13. Overall, the σ^{3PA} results for the case of GFP and RFP are in good agreement with experiment. The σ^{3PA} obtained for GFP, 7.2×10^{-83} cm^6 s^2 photon^{-2}, is larger than the experimental cross-section, 1.59×10^{-83} cm^6 s^2 photon^{-2}; nevertheless, both values are within the same order of magnitude. The ratio between the cross-sections of GFP$_N$ and GFP$_A$ is 12:1. The neutral form of the GFP chromophore (GFP$_N$) is the main species at the experimental σ^{3PA} wavelength. The 1PA spectra of GFP exhibits two main peaks within the range of 300 to 500 nm, corresponding to GFP$_N$ (398 nm) and the anionic form GFP$_A$ (478 nm). In a similar fashion, the 2PA spectra of GFP also exhibits two peaks at $\sim$400 and $\sim$420 nm.[88,258] Here, σ^{3PA} for the two protolytic states of GFP match previous observations made on the 1PA and 2PA spectra.

Table 4.3. For the Fluorescent Protein Chromophores in Figure 4.1, the Computed Excitation Energies (VEE) in nm, 3PA Probabilities (δ^{3PA}) in au, and 3PA Cross-Sections (σ^{3PA}) in $cm^6 \ s^2 \ photon^{-2}$. Values of σ^{3PA}s Were Obtained According to Equation 4.1 Using the CAM-B3LYP/aug-cc-pVDZ Method in Vacuum

FP Chromo	S_n	ω (nm)	δ^{3PA} (au)	σ^{3PA}	Exp σ^{3PA}
BFP_1	1	340	1.86E+06	9.31E-84	
BFP_2	1	318	3.39E+06	2.06E-83	
BLB	1	399	1.66E+08	5.12E-82	
CFP	1	354	1.75E+07	7.81E-83	
GFP_A	1	400	1.94E+06	5.98E-84	
GFP_N	1	338	1.41E+07	7.22E-83	$1.59E-83 \ (433)^a$
KO_A	1	481	4.99E+07	8.90E-83	
OR_A	1	475	3.27E+07	6.02E-83	
RFP_A	1	506	2.75E+07	4.17E-83	$8.55E-83 \ (530)^b$

a Wild-type GFP.[191]

b The average of values reported for different RFPs.[194,195]

Within the family of red fluorescent proteins (RFPs), represented by the chromophore RFP_A in Figure 4.1 and Table 4.3, the so-called fruits series[59] are an example of the role that the protein plays in determining the 2PA response within FPs. Drobizhev et al. discussed for a set of FPs from the fruit series and some other RFPs how their σ^{2PA} depends on the electric field exerted by the protein—chromophore interactions.[5] Similar experimental measurements of σ^{3PA} across the mFruit series have not been made but would lead to interesting insights into the role that the protein environment has on the three-photon photophysical properties of FPs.

Here, the σ^{3PA} obtained using the CAM-B3LYP/aug-cc-pVDZ method for the bare chromophore RFP_A is $4.17 \times 10^{-83} \ cm^6 \ s^2 \ photon^{-2}$; which is in reasonable agreement with the average of the experimental σ^{3PA}s in Table 4.3, 8.55×10^{-83}.[195] The obtained cross-section particularly agrees (with an error of 30%) with the measured value for DsRed and mOrange, $3.20 \times 10^{-83} \ cm^6 \ s^2 \ photon^{-2}$ in both cases.[195] However, this might be fortuitous since the protein effects were not introduced in this work. Overall, the method employed here is useful to estimate the σ^{3PA} of RFPs, which proves our initial objective but, without including the protein environment it is insufficient to distinguish one RFP from the other. To address the discrepancy between computation and experiment and to get a better picture of the extent of agreement between them, it is necessary to

take into account the protein environment.[26] Computations using solvation models, like PCM, would only confirm what is already clear, that the MPA response of FPs depends on the protein effects. Thus, computations utilizing PCM were not performed on the chromophores.

Due to insufficient experimental results, we could not assess the performance of our computational strategy to distinguish different families of FPs, i.e., we cannot elaborate on the error between the σ^{3PA} for the GFP- and RFP-type chromophores.

Because of the selection rules that govern them and their intrinsic dependence of different photophysical parameters such as transition, and, for 2PA, permanent dipole moments, 1PA features of FPs and FP chromophores cannot be translated necessarily to the 2PA realm. Similarly, the 2PA properties of these systems are not necessarily a reflection of their 3PA strengths. Figure 4.4 compares for each of the FP chromophores of Figure 4.1 the σ^{2PA}s reported by Salem and Brown [13] with the σ^{3PA}s obtained in this work using the CAM-B3LYP/aug-cc-pVDZ method.[3] The σ^{3PA}s in Figure 4.4 are given in 1×10^{-84} cm^6 s^2 photon^{-2} units, based on the experimental σ^{3PA} value of the wild-type GFP.[191] Figure 4.4 suggests that the trend previously observed in the FPs chromophores of Figure 4.1[13] for 2PA persists for 3PA. CFP and GFP$_N$, two of the chromophores with the largest σ^{2PA} among the group, also exhibit a large σ^{3PA} for excitation to S$_1$. Notice that although σ^{3PA} and σ^{2PA} of BLB are noticeably large with respect to the rest of the chromophores, it might not be what could be observed in the experiment. In a previous investigation, the experimental σ^{2PA} tended to be overestimated by the computational models employed.[13] The large deviation of BLB's σ^{2PA} might occur because (i) the geometry determined in vacuum and using the PCM solvation model deviates drastically from the one the chromophore adopts in the protein and (ii) the protein environment plays a significant role in diminishing the σ^{2PA} of this chromophore.

---

[3]We corrected the σ^{2PA} values reported in the original source [13] by $\frac{1}{4}$ according to Equation C.1.

Nonetheless, BLB at the given geometry[13] is predicted to have the largest σ^{3PA} of the group. The latter observations might change as the environmental effects are introduced. In the case of 2PA (and in fewer cases, for absorption processes of higher order), it has been demonstrated that the protein environment can play a significant role in determining the σ^{NPA} of FPs.[26,105] Experimentally, it has been observed that the 3PA spectra can be (i) blue-shifted with respect to their 1PA counterpart, as in the case of tdtomato RFP, mStrawberry, eRFP, mRaspberry, mKate, and mPlum, or (ii) coincident to it, such as in the case of red algae, mBanana, mOrange, and dsRed.[190,195]
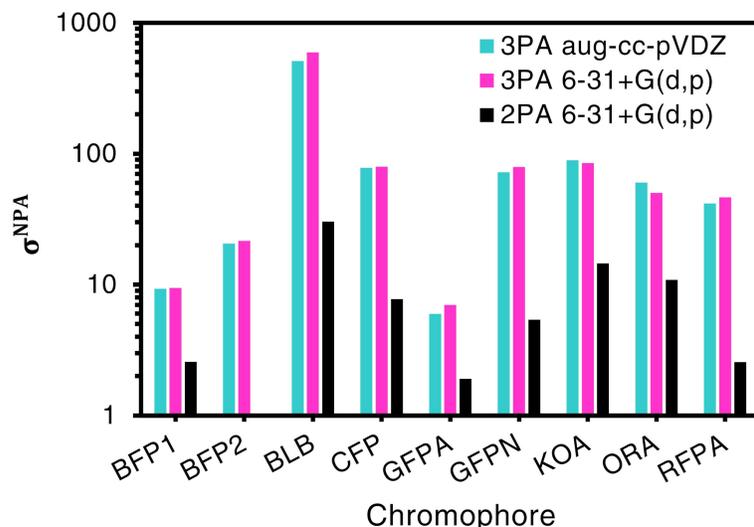
Figure 4.4. Depiction of the 3PA and 2PA ability for each of the chromophores in Figure 4.1 in terms of their σ^{NPA}. Here, σ^{3PA}s obtained using the CAM-B3LYP[184] functional and the aug-cc-pVDZ and 6-31+G(d,p) basis sets are given in 1×10^{-84} cm^6 s^2 photon^{-2} units, whereas σ^{2PA}s previously obtained by Salem and Brown, are expressed in 1×10^{-50} cm^4 s molecule^{-1} photon^{-1} (or GM) units.[13] Notice that the value of σ^{2PA} for BFP$_2$ is 0 GM in the original source,[13] and that the y-axis is given in a logarithmic scale.

4.4 Conclusions

Using the CAM-B3LYP[184] functional, we tested a total of ten basis sets (Table 4.1) for obtaining the σ^{3PA} on the benchmark fluorescein dianionic system, F^{2-}. Complementary studies with a reduced number of basis sets, were carried out on serotonin. Finally, the aug-cc-pVDZ and 6-31+G(d,p) basis sets were used for rhodamine 6G and the FP chromophores. The 3PA response of the systems studied here did not show any significant

70

dependence on the choice of basis set; nevertheless, we found that the CAM-B3LYP/aug-cc-pVDZ method exhibited the best performance in terms of accuracy and computational cost. We showed for a representative set of fluorescent protein chromophores that their experimental σ^{3PA} can be approximated using computational means within the time-dependent density-functional theory framework.

The agreement between our results (on isolated FP chromophores in vacuum) and experimental measurements (on full FPs in solution) are encouraging towards pursuing further investigations on the 3PA response of FPs and FPs chromophores. In this regard, it would be interesting to evaluate a wider set of chromophores, in particular the so-called fruit series for which their σ^{3PA} have been determined already by experimental means. [194,195] In a similar fashion to the case of 2PA, the protein environment seems to play an important role in determining the 3PA properties of the fluorescent proteins and their chromophores. For example, the σ^{3PA} we obtained for the RFP chromophore is comparable to what has been determined experimentally for the RFP, however, it is quite different from what has been measured for other types of RFPs, like the ones in the fruits series. Thus, future investigations would focus on the computational evaluation of the 3PA behaviour of fluorescein, rhodamine 6G, serotonin, the fruit series, and other FPs, taking into account the environment effects through polarizable embedding models, as has been done for the case of 2PA in rhodamine 6G, [259] or for 3PA and 4PA in the case of the GFP. [26]

Chapter 5

Two-photon Absorption Cross-sections in Fluorescent Proteins Containing Non-canonical Chromophores Using Polarizable QM/MM[4]

5.1 Introduction

Among the principal outcomes of the work of Shimomura on the bioluminescent crystal jellyfish[1] was the discovery of the green fluorescent protein (GFP),[2] a barrel-shaped protein, in which a chromophore is located and responsible for its bright green color.[3,6] A set of chromophores from the work carried out by Salem et al.[14] is shown in Figure 5.1.
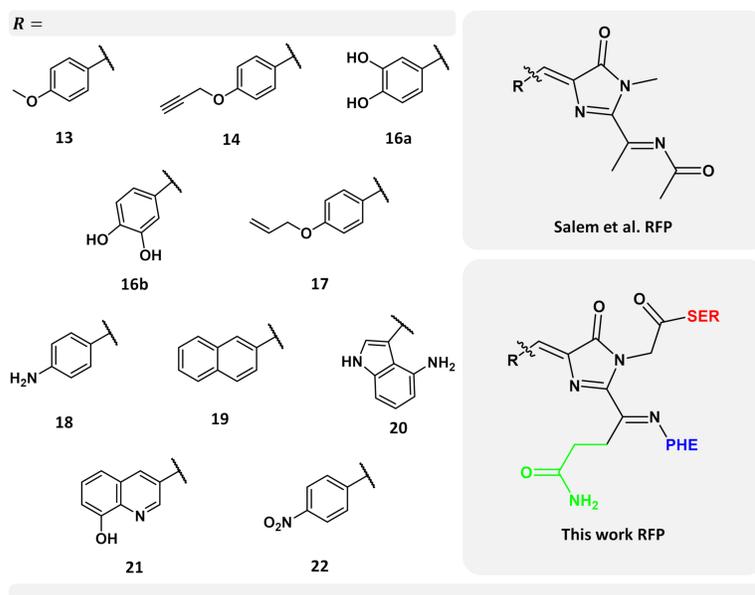
Figure 5.1. Left-hand side: selected set of chromophores from the work carried out by Salem et al.[14] Numerical labels are the same as in their work for ease of comparison. Top right-hand side corner: chromophore model used by Salem et al.[14] Bottom right-hand side corner: chromophore model used in this work. In green, the chromophore side chain $-CH_2-CH_2-CO-NH_2$ excluded in the previous work,[14] while neighbouring amino acids serine and phenylalanine are indicated in red and blue, respectively.

---

[4]This chapter has been copied and/or adapted from the publication: Rossano-Tapia, M.; Olsen, J. M. H.; Brown, A. Two-photon absorption cross-sections in fluorescent proteins containing non-canonical chromophores using polarizable QM/MM. *Front. Mol. Biosci.* **2020**, *7*, 111.

The presence of such a barrel-shaped protein was discovered later not to be exclusive to the crystal jellyfish; indeed, similar fluorescent proteins (FPs) also were found among corals and some others species of the Anthozoa class. FPs of the Anthozoa species exhibit a red shift in their absorption and emission properties with respect to their GFP homologues, therefore, they were called red fluorescent proteins (RFPs). One example is the DsRed RFP, which is found in the anemone Discosoma striata.[89] The chromophore structure in FPs is characterized by an imidazole ring, made from the cyclization of three amino acids; in the case of DsRed, they are glutamine 66, tyrosine 67, and glycine 68 (Gln66, Tyr67, and Gly68, respectively). The RFP chromophore structure is shown in Figure 5.1 (bottom right-hand side corner).

Fluorescent proteins have been used as dyes and clinical markers over the last two decades, and their multi-photon absorption properties have allowed them to be applied in deep-tissue clinical imaging at low phototoxicity.[4,5] Currently, dozens of hues of FPs covering all the colors in the visible spectrum have been engineered, with the purpose of overcoming some of the most common limitations of many FPs, such as low quantum yield, autobleaching, phototoxicity, and weak two-photon absorption (measured in terms of the cross-section, σ^{2PA}). Some strategies in the design of new FPs involve changing the amino acids in the protein sequence, like the way mRFP1 (a monomeric DsRed variant) was tailored in trying to improve DsRed properties, and from which the so-called fruit series was obtained by Shaner et al.[59] However, in 2015 and 2016, the modification of GFP[12] and RFP[14] chromophores, through the substitution of the tyrosine in these chromophores by one of a selected set of non-canonical amino acids previously obtained by Liu and Schultz,[140] was suggested as a means towards improving their two-photon absorption. The resulting chromophores from the incorporation of non-canonical amino acids here are called non-canonical chromophores (nCCs). Incorporation of nCCs has been accomplished and studied previously by experimental means, as discussed for example by Fang et al.[260] and Nediljko and Prajna.[261] Regarding FPs, gold fluorescent

protein is an example of inclusion of non-canonical amino acids for which one-photon absorption properties have been measured experimentally and explored computationally,[171] as will be discussed later in this chapter.

Beyond expanding the color span or improving the structural features of existing FPs, the persistent efforts on creating novel FPs are motivated by their utility as bioimaging tools,[262] including the fact that they do not require any accessorial proteins or fluorophores, as their fluorescence comes from the chromophore embedded in them.

One- and two-photon absorption (1PA and 2PA, respectively) properties of FPs and their chromophores have been addressed using different computational tools, mainly through quantum mechanical (QM) methods based on time-dependent density functional theory (TD-DFT).[9,11,13,62,118] Due to its computational cost, TD-DFT can be applied only to small systems, i.e., the chromophore and, possibly, a few nearby residues, rather than to the entire chromophore–protein system. Outcomes of computational investigations have been compared to experimental data and shown to provide reasonable agreement,[9,62,263] therefore, many 1PA and 2PA studies have been carried out for the FP chromophores in vacuum without considering the role of the protein matrix, such as the work by Salem and Brown,[12] Salem et al.,[14] Nifosì et al.,[62] Nifosì and Luo,[11] Nifosì and Luo,[9] Salem and Brown,[13] List et al.,[264] Beerepoot et al.,[60] and Filippi et al.,[49] among others. Other studies, have included and/or examined the role of the protein in the computation of 1PA and 2PA properties of FPs using combined quantum mechanical/molecular mechanical methods (QM/MM), such as the work carried out by Marques et al.,[65] Moron et al.,[80] Sanchez-Garcia et al.,[265] Sanchez-Garcia et al.,[73] Kaila et al.,[69] Steindal et al.,[85] Kulakova et al.,[106] Beerepoot et al.,[77] List et al.,[86] Schwabe et al.,[70] Steindal et al.,[26] Nåbo et al.,[78] some of which have shown that the protein environment can have a large impact on one- and multi-photon absorption properties (as depicted in Figure 5.2).[26,70,77,86]
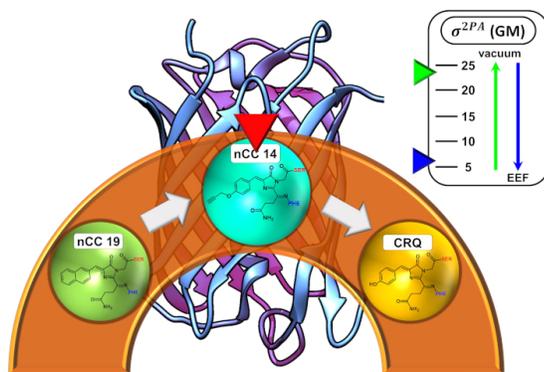
Figure 5.2. Depiction of the dependence of the 2PA cross-sections for an nCC in vacuum versus the protein environment.

QM/MM approaches involve sectioning the entire system into different parts (two or more layers), where each part is modeled using a different computational approach. The part of the system where bond-breaking or -forming takes place, or excited-state processes occur is described through QM methods. The rest of the layer(s) can be treated using an MM model or a less rigorous (computationally less expensive) QM method. QM/MM methods are used to reduce the cost of computations of large systems that would not be feasible to study using pure QM means with current computational resources, for example, those provided through Compute Canada-Calcul Canada.[266] The high performance computing (HPC) resources employed in the elaboration of the present thesis are described in the Preface section.

In particular, for the set of nCCs mentioned above, previous attempts of addressing the environment effects on their 2PA properties have been made using the self-consistent reaction field (SCRF) polarizable continuum model (PCM).[14] Other approaches have used molecular dynamics (MD), along with QM computations on the chromophore alone, such as the study carried out by Şimşek and Brown,[16] which examined 1PA and 2PA properties of gold fluorescent protein.[171] However, continuum models, such as PCM, are not suitable for heterogeneous systems like proteins, and while QM computations on the chromophore carried out on top of MD simulations capture some indirect environment effects through the geometry, they lack the direct electrostatic interactions.

The latter can be included efficiently and accurately using a polarizable embedding (PE) model,[23,24] in particular, the one proposed by Olsen et al., which has been used previously to compute 2PA properties of FPs.[26,86,121] The aforementioned PE model is a fragment-based quantum-classical approach similar to QM/MM, but where the MM part is subdivided into small fragments (usually amino acid residues in the case of proteins) that are represented by atom-centered charges, dipoles, and higher-order multipoles, as well as polarizabilities. These parameters are derived from QM calculations on the individual fragments, which leads to a very high-quality representation of the protein.[96] The PE model was designed to describe general response properties, including 2PA. It allows the QM and MM parts of the system to mutually polarize each other and can account for local-field effects.

The local field acting on a chromophore that is embedded in a polarizable environment generally is different from the externally applied field due to the polarization of the environment by the external field. In the PE model, this effect is modeled through the so-called external effective field (EEF) approach.[121] The EEF effects will not affect the excitation energies but can be important for other properties, such as multi-photon absorption. In fact, studies carried out on the DsRed and GFP proteins, showed that local-field effects play a major role in the determination of multi-photon cross-sections.[26,27,121] Indeed, the results showed that inclusion of local-field effects, through EEF, is required in order to be comparable to full QM results.[121] The failure to include local-field effects in the computation of 2PA cross-sections can lead to a misrepresentation of 2PA features of FPs, as shown by Steindal et al. in the case of GFP.[26] This topic is discussed further later in this chapter.

Using the PE model, including local-field effects, we investigated the 1PA and 2PA properties of a selected set of 10 nCCs taken from a more complete set previously studied by Salem et al.[14] The objective was to provide further insights into the properties of these nCCs within a more realistic context and to provide more information on the perfor-

mance of the PE model in the computation of multi-photon absorption properties in FPs. Figure 5.1 shows the models used here, where –R represents each of the non-canonical amino acids employed. Figure 5.1 also shows the differences between the previous chromophore model[14] and the model used in this work; further discussion about this matter will be provided in the next sections. The selected set of chromophores used here (Figure 5.1) are those, which exhibited the largest 2PA cross-sections and the largest intrinsic cross-section (obtained from the analysis of σ^{2PA} with respect to the tilt and twist angles) in the work of Salem et al., as in the case of nCC 21.

5.2 Computational Methods

5.2.1 Modeling the Protein-chromophore Structures

The main challenge in the construction of nCC-protein model structures is the fact that none of the nCCs shown in Figure 5.1 have been matured in a red fluorescent-type protein experimentally. Thus, there are no experimental protein crystal structures that can be used either directly or as initial structures for geometry optimization. Only one of the nCCs (no. 20) has been expressed successfully in the gold fluorescent protein, and its 1PA properties have been evaluated.[171] However, this protein belongs to the family of GFP derivatives. To overcome this shortcoming, the construction of nCC-protein models consisted of two stages: (i) selecting a protein structure that can be used as a host for the nCC, based on the criteria that the host protein should exhibit red-shifted absorption (like RFPs) and possess a considerable (for FPs) 2PA cross-section (>50 GM) and (ii) replacing the canonical amino acid chromophore with each one of the nCCs shown in Figure 5.1.

TagRFP[267] and some members of the fruit series[59] are among the brightest of the RFP family.[4] However, we used DsRed (PDB:1ZGO)[7] because this protein meets the criteria described above. DsRed is the parent protein of the most common RFPs and has not been tailored in the laboratory around any particular chromophore in the way the fruit series

77

proteins were tailored around the native DsRed chromophore (CRQ). Moreover, the 2PA properties of DsRed have been studied previously through computational QM/MM schemes,[73,86,121,265] and this provides us with a reference to which we can compare the data we obtain here. Experimentally, Drobizhev et al. reported the DsRed 2PA cross-section to be 103 GM for the long-wave absorption band (1050 nm).[5]

To build the nCC–DsRed models, we used a single monomer of 1ZGO and modified the native CRQ chromophore to each of the nCCs (Figure 5.3). Missing hydrogens in 1ZGO were added using pdb4amber,[268] whereas the missing residues 1–5, which are located outside the barrel of the protein, were not added since they were not considered crucial for the purpose of the investigation presented here. Here, we present an initial computational investigation of the protein effect on 2PA cross-sections in RFP-like nCCs. Thus, it is based on a single geometry-optimized structure for each chromophore, i.e., no conformational sampling is included. This way, we get an estimate of the effect of adding the protein matrix into the computation of 2PA cross-sections of the non-canonical amino acids studied here. Moreover, we did not include water molecules inside the cavity or solvent molecules around the protein because understanding the presence of water would require molecular dynamics simulations. The inclusion of statistical sampling could change the quantitative results presented here. However, the latter is beyond the scope of the present investigation but could be the subject of future research.
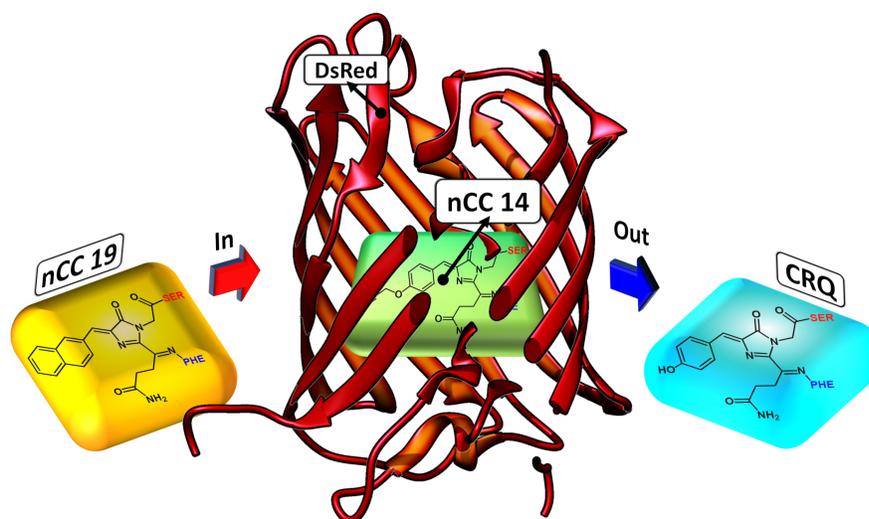
Figure 5.3. Depiction of how nCC–DsRed (1ZGO) models were created. The native chromophore (CRQ) in the DsRed protein matrix is replaced by the non-canonical chromophore (nCC) model.

The nCC–DsRed structures were optimized using a two-layer ONIOM[269–271] QM/MM scheme implemented in Gaussian 16.[223] The QM layer comprised the nCC and the phenylalanine residue bonded to it. The whole phenylalanine amino acid structure was included to avoid cutting double bonds or bonds near the conjugated chain, while retaining the acylimine moiety. Also, the carbonyl group bonded to serine, which previously was excluded by Salem et al.,[14] when studying the isolated chromophores, was included. The rest of the protein is part of the MM layer. All optimizations used mechanical embedding and were carried out in two steps using the default convergence criteria: first, using the semi-empirical PM6[175] method; second, using the long-range corrected functional CAM-B3LYP[184] and the 6-31+G(d,p)[157] basis set. Molecular mechanical parameters, including charges, were obtained from the Amber force field libraries (ff96, ff10, and GAFF) included in Amber 18.[268] In particular, nCC parameters were obtained from the R.E.D. Server[272–275] using the default type of charges, RESP-A1, and the computational method HF/6-31G(d). Hessian analyses were performed on the optimized geometries to verify the absence of imaginary frequencies, and thus that the structure is indeed a minimum.

The CAM-B3LYP[184] functional was chosen based on a comparison of the 14-DsRed

model optimized using three different functionals, i.e., CAM-B3LYP,[184] ωB97XD,[276] and PBE0,[220,277], along with the 6-31+G(d,p) basis set, and all within the ONIOM scheme. In all cases, the Amber force field (libraries mentioned above) was used for the MM region. The optimized structures using CAM-B3LYP[184] and ωB97XD did not show substantial structural differences (see Figure D.1 comparing the overlapped structures). The PBE0 functional suffered from convergence problems, failing to find a stable minimum. In addition to mechanical embedding, optimization of one of the models, nCC 20, was performed using electrostatic embedding and the CAM-B3LYP[184] functional. In this optimization the residues belonging to the QM region, PHE 60 and CRQ 61, along with residues GLN 59 and SER 62 treated by MM, could relax while the rest of the protein was kept fixed.

5.2.2   Two-photon Absorption Cross-section Computations

All computations of 2PA cross-sections were carried out using the Dalton program,[224,278,279] employing the PE model,[23] including effective external field (EEF) effects,[121] to describe the protein environment. The 2PA cross-section is given by[60]

$$\sigma^{2PA} = \frac{N\pi^2 a_0^5 \alpha}{c_0} \frac{\omega^2}{\Gamma} \delta^{2PA} \,, \tag{5.1}$$

where a_0 is the Bohr radius, α is the fine structure constant, c_0 is the speed of light, Γ is the lifetime broadening factor, which is derived from a Lorentzian function and assumed to be 3.675×10^{-3} Hartree (or 0.1 eV) to facilitate comparison to experiment (as well as previous computational results), ω is the excitation energy (Hartree/photon), which for 2PA is half the energy difference between the excited and ground states, and δ^{2PA} is the 2PA transition strength. The resulting σ^{2PA} is given in 10^{-50} cm^4 s molecule^{-1} photon^{-1} or GM (Göppert-Mayer after Maria Göppert-Mayer).[149]

In all cases, the QM region consisted mainly of the nCC. Compared to previous

models,[14,16] two changes were made: (i) any side chains in the native CRQ chromophore of DsRed were preserved and (ii) amino acid residues that are bonded covalently to the chromophore, i.e., serine (SER) and phenylalanine (PHE), were included (see Figure 5.1). The latter are included in the QM region to avoid cutting any bonds near the acylimine moiety and also to include any possible contributions to the 2PA process that the neighbouring amino acids could have.[26] For the MM region, corresponding to the rest of the protein, distributed atom-centered charges, dipoles, quadrupoles, and dipole–dipole polarizabilities for each of the amino acid residues were generated using the PyFraME Python package.[280] PyFraME employs Dalton[278] and LoProp for Dalton[281] to compute the parameters based on a fragmentation scheme. The distributed multipoles and polarizabilities were computed using the LoProp approach[282] employing the CAM-B3LYP[184] functional and the ANO-form of the 6-31+G(d) basis set (named loprop-6-31+G(d) in Dalton). We refer to the work by Steinmann et al.[97] for a tutorial review on the setup, use, and capabilities of the PE model.[23] Figure 5.4 shows a depiction of the QM (SER-nCC-PHE) and MM regions.
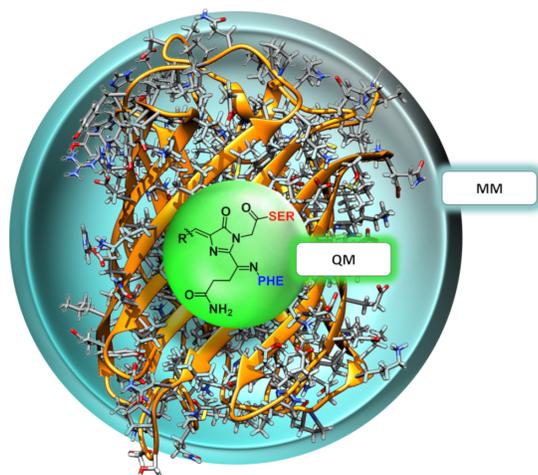
Figure 5.4. Two-layer QM/MM partitioning in each of the nCC–DsRed systems. In the 2PA computations, the QM region includes the chromophore and the neighbouring residues serine (red) and phenylalanine (blue), while the classical MM region includes the protein structure only.

Computation of 2PA cross-sections was carried out using the CAM-B3LYP[184] functional, while different Pople basis sets [6-31G(d), 6-31+G(d), and 6-31+G(d,p)]

and a segmented polarization-consistent basis set (pcseg-2),[228–230] were tested on the 14-DsRed system. After determining the role of the basis set, 2PA cross-sections were computed for the geometry-optimized chromophores (Figure 5.1) (i) in vacuum (i.e., without the protein), (ii) with charges of the atoms closest to the hydrogen link atoms (0.5 or 1.5 Å) redistributed to nearby atoms to avoid overpolarization due to electronic density-point charge proximity (Figure 5.5), and (iii) including or excluding EEF effects. For all computations, hydrogen link atoms were treated using the STO-3G minimal basis set to prevent electronic density from "spilling out" and generating spurious orbitals at the bonding sites, as described by Steinmann et al.[97] The outcome from these analyses will be elaborated in the next section.
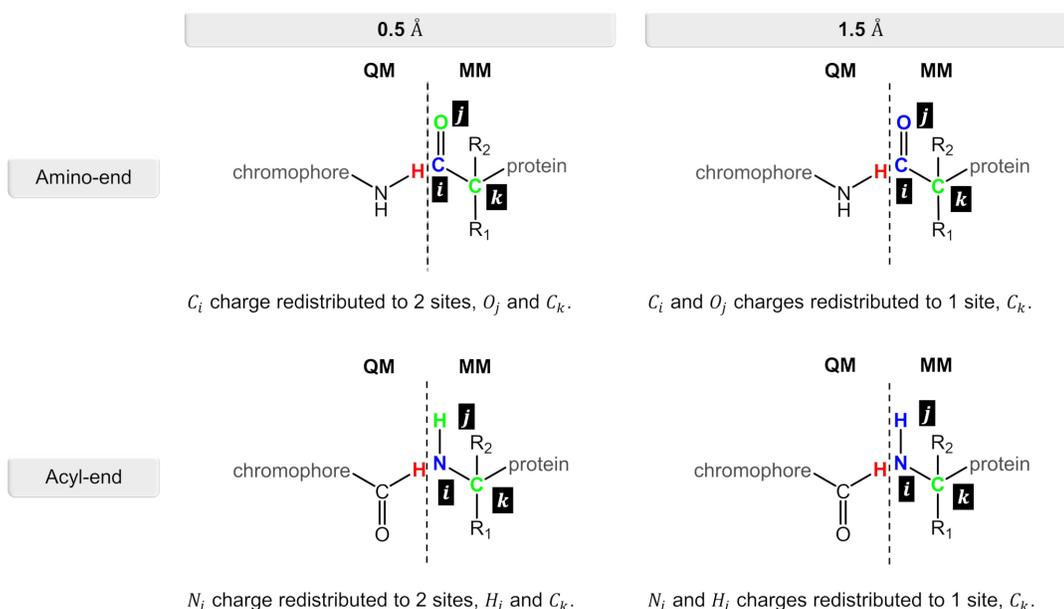
Figure 5.5. Charge redistribution scheme for the N-terminal and C-terminal sides in the nCC–DsRed systems. Atom(s) in blue represent the charge(s) to be redistributed to the atom site(s) in green.

5.3   Results and Discussion

5.3.1   Geometry Optimization

The impact of the protein environment on the optimized geometries of the nCCs was determined by comparing each structure obtained using the ONIOM QM/MM scheme

and the CAM-B3LYP/6-31+G(d,p):Amber method to the experimental structure of the canonical chromophore in crystal DsRed[7] and the corresponding non-canonical chromophore optimized in vacuum using PBE0/6-31+G(d,p).[14] The comparisons are based on the tilt (θ) and twist (φ) dihedral angles, formed by atoms i—j—k—l and k—l—m—n, respectively, as illustrated in Figure 5.6. Through computational studies,[12,14,16] it has been shown that 2PA cross-sections depend on these angles. Based on experimental data, some authors have observed that θ and φ also play important roles in the chromophore's 1PA properties and in the conjugation of the double bonds found between the chromophore and the acylimine group; RFP family chromophores that exhibit a non-planar structure tend to exhibit smaller quantum yields than, for example, the planar DsRed chromophore.[283–285] The results of the θ and φ angle comparisons are summarized in Figure 5.7. Table D.1 contains the values and deviations of the tilt and twist angles for all nCCs in Figure 5.1. The relation between the structural differences of the optimized nCCs and their 2PA cross-sections will be discussed in the next subsection.
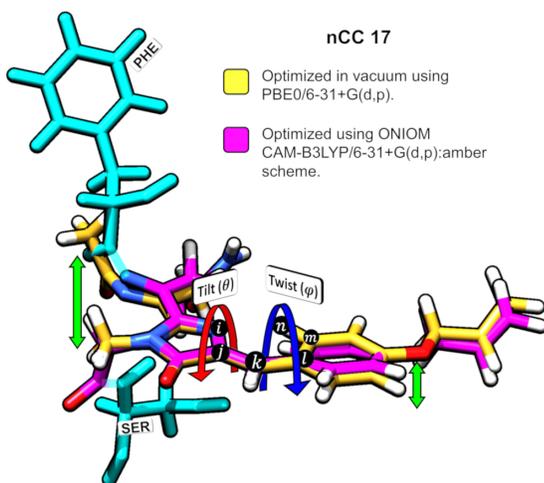
Figure 5.6. Superposition of nCC 17 structures a) optimized using PBE0/6-31+G(d,p)[14] in vacuum and b) optimized using ONIOM [CAM-B3LYP/6-31+G(d,p):Amber]. Tilt (θ) and twist (φ) angles are shown. The green arrows indicate the deviation between the structures.
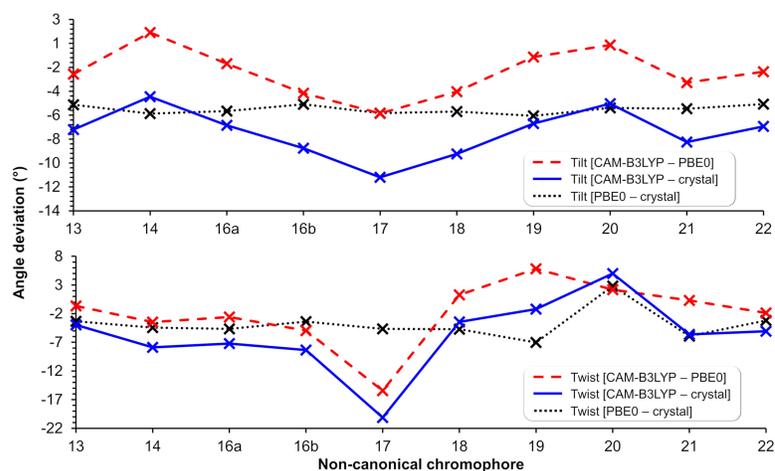
Figure 5.7. Tilt (top) and twist (bottom) angle deviations (in degrees) between the structures optimized using CAM-B3LYP/6-31+G(d,p) (in protein) and PBE0/6-31+G(d,p) [14] (in vacuum), and the canonical chromophore from the crystal structure (PDB ID: 1ZGO). [7] Note: the lines are included to guide the eye through the deviations of tilt and twist angles within methods and nCC–DsRed models.

Apart from applying the ONIOM QM/MM method for the optimization of the nCC–DsRed systems, the QM region used here was also larger than the structures optimized by Salem et al. [14] In our work, the nCCs and the phenylalanine bonded to it were optimized quantum mechanically; previously, only the chromophore structure was optimized, i.e., without any neighbouring amino acids. In addition, Salem et al. excluded the side chain –CH_2–CH_2–CO–NH_2, highlighted in green in Figure 5.1, and the carbonyl group toward the serine amino acid. These differences in optimization strategy and chromophore definition resulted in differences in the θ and φ angles between the CAM-B3LYP[184] (QM/MM) and PBE0 (vacuum) structures that are between −5.4 and 1.4, and −15.5 and 5.8 degrees, respectively. The deviation of the QM/MM-optimized geometries from the crystal structure range from −10.7 to −4.0 and from −20.1 to −1.2 degrees for θ and φ, respectively. In contrast, the θ and φ angles of the nCCs optimized in vacuum using PBE0[14] do not differ by more than 7 degrees from the crystal structure. Moreover, the vacuum to crystal differences for the θ and φ angles for all nCCs vary within rather small intervals, between −4.6 and −5.6, and −3.3 and 7.0 degrees, respectively. These small differences in the PBE0-based vacuum structures, along with the

larger differences observed for CAM-B3LYP-based QM/MM structures with respect to the crystal structure, suggest that the non-canonical amino acid moiety, –R (Figure 5.1), does not determine on its own these two angles, and it is less likely that the –R moiety could provide a realistic picture of what the structure of the chromophore in the protein could be without considering the protein in the optimization process. From the differences in the CAM-B3LYP[184] θ angle compared to the PBE0 results and the angles observed experimentally in the crystal, it is clear that θ will be impacted largely by the presence of nearby amino acids and barely by the nature of the –R substituent (Figure 5.1). The largest CAM-B3LYP[184] to crystal deviations in φ are observed in nCCs 17, 19, 20, and 21, and are due likely to the size of the chromophore. For nCC 17, optimizing it within the volume of the protein cavity played a major role in determining its structure, particularly the θ and φ angles. The embedding of the chromophore in the protein matrix created distortion of the tilt and twist angles in the chromophore, as described above. From the protein perspective, the substitution of the canonical chromophore by the non-canonical model caused structural changes in residues non-covalently bonded to it and located in its immediate surroundings. Some of these structural changes included the expansion of the cavity residues away from the chromophore, specifically those residues close to the –R moiety. The structural changes for two selected residues (to ease visualization), SER 139 and LYS 156, in comparison with 1ZGO are shown in Figure D.11.

Model nCC 20, one of the nCCs with the largest tilt and twist angle deviations from the DsRed crystal structure and 2PA cross-section computed in vacuum[14] and in the protein, was optimized further using electrostatic embedding. The tilt and twist angles of the resulting structure, –1.59 and 11.18 degrees, respectively, are similar to the ones in the nCC 20-DsRed structure optimized using mechanical embedding, 0.27 and 8.29, respectively. Figure D.10 illustrates the small changes in the geometry after optimization using electrostatic embedding.

5.3.2 2PA Cross-sections

One- and two-photon absorption cross-sections in all nCCs were computed both in vacuum and in protein using PE to model the effects of the protein environment. The QM region in these computations included the chromophore and its two neighbouring covalently bonded amino acids, serine and phenylalanine (see Figure 5.1), whereas the rest of the protein was treated classically. The two charge redistribution schemes depicted in Figure 5.5 were evaluated using the 14-DsRed model employing CAM-B3LYP[184] and different basis sets [6-31G(d), 6-31+G(d), 6-31+G(d,p),[156–158,160,221] and pcseg-2[228–230]] in order to establish a suitable approach. The results are provided in Table D.2, together with corresponding molecular orbital (MO) plots in Figures D.4–D.6. For comparison, MO plots of nCC 14 in vacuum are provided in Figure D.3.

Using a point-charge redistribution distance of 1.5 Å, results in an unexpected low intensity transition at around 3.2–3.3 eV, which is due most likely to over-polarization effects. Indeed, an inspection of the MOs (Figure D.5) reveals that this is not a relevant transition, as the main contribution is from an occupied MO that is not localized on the chromophore. Even the intense transition, which is to the second state, involves a main contribution from an occupied MO that has large components outside of the chromophore. Using instead a redistribution distance of 0.5 Å, we find the expected intense $\pi \to \pi^*$ transition as the lowest state. However, for the small 6-31G(d) basis, we find that the two lowest states are quite close in energy, thus resulting in shared intensity between the two transitions. Adding diffuse functions, i.e., using 6-31+G(d), or using the larger pcseg-2[228–230] basis set, increases the separation between the states, thus largely avoids the issue. Comparing the results obtained using 6-31+G(d), 6-31+G(d,p),[156–158,160,221] and pcseg-2,[228–230] we observe very small differences for the two lowest states, but the third state differs significantly. This is not necessarily an issue since we are interested mainly in the lowest intense transition. However, it may be an indication of issues with over-polarization or electron spill-out. Nonetheless, it is clear that the point-charge

redistribution distance of 0.5 Å is superior in this case, therefore, we include only results based on this choice for the following analyses.

To investigate the role of the basis set further, we take a closer look at the MOs. The six highest occupied MOs and six lowest unoccupied MOs of nCC 14 in vacuum and in the protein (14-DsRed) are provided in Figures D.3 and D.4, respectively. A comparison of the MOs reveals rather large differences, and, in particular, the unoccupied MOs depend strongly on whether diffuse functions are used or not and whether they are determined in vacuum or in the protein. For nCC 14 in vacuum, the diffuse functions, which are present in 6-31+G(d) and 6-31+G(d,p),[156–158,160,221] result in Rydberg-like unoccupied MOs, except for the lowest unoccupied MO (LUMO). Such Rydberg-like orbitals would be expected to be much higher in energy when embedded in an environment (if at all present) due to Pauli repulsion. However, since the PE model does not include Pauli repulsion, the use of diffuse functions or large basis sets is not always straightforward. Indeed, for computations in the protein environment, we observe spurious unoccupied MOs when the 6-31+G(d) and 6-31+G(d,p)[156–158,160,221] were used. Similar effects are not observed, at least to the same degree, for the unoccupied MOs obtained using pcseg-2[228–230] or 6-31G(d), which suggests that the diffuse functions have a negative effect on the MOs when the protein is involved.

Typically, the transition of interest in 2PA processes in FPs is to the lowest-lying excited state, S_1. The 1PA results for nCC 14 using the functional and basis sets cited above show that this transition is dominated by the highest occupied MO (HOMO) and the LUMO. Therefore, the presence of spurious MOs beyond the HOMO and LUMO might not be considered important and either 6-31+G(d,p) or pcseg-2[228–230] can be used. However, the role of the rest of the MOs, especially those with unphysical descriptions, on σ^{2PA}, is unknown. They may be important contributors to σ^{2PA}, as the expression for the 2PA transition moment involves a sum over all excited states, thus, in principle, involving all MOs. The excitation energies, oscillator strengths, and main MO contributions for S_1

for all nCCs in vacuum and embedded within the protein matrix are given in Table D.3. The excitation to S_1 in most of the remaining nCCs, besides nCC 14, also involves mainly the HOMO and LUMO. However, in nCC 19, it is HOMO–1 that dominates, while in nCCs 21 and 22 it is primarily HOMO–2. In these cases, the troublesome scenario discussed above appears to be present even with the redistribution distance of 0.5 Å. The highest occupied and lowest unoccupied MOs of models 19, 21, and 22 are provided in Figures D.7, D.8, and D.9, respectively. These cases further emphasize that care must be taken when evaluating σ^{2PA} using the PE model and QM/MM approaches in general, particularly when diffuse functions or large basis sets are used. Besides possible issues present at the QM/MM interface where bonds have been broken, we also suspect that this is a symptom of over-polarization or electron spill-out because of the proximity of the point-charges surrounding the electronic density.

Two-photon absorption cross-sections for the nCCs in vacuum and embedded in the DsRed protein (nCC–DsRed) are shown in Table 5.1. For the latter, two different approaches were considered, including or excluding local-field effects (denoted PE(+EEF) and PE(−EEF) respectively).

Table 5.1. For All Non-canonical Chromophores (nCCs) in Vacuum (Figure 5.1) and the nCC–DsRed (Non-canonical Chromophore With DsRed Protein) Systems, the Two-photon Absorption Cross-sections (σ^{2PA}) Computed Using the CAM-B3LYP[184] Functional Along With the 6-31+G(d,p) or pcseg-2[228–230] Basis Sets. For the nCC–DsRed Systems, the PE Model is Used to Include the Effects from the Protein Either with Effective External Field Effects [PE(+EEF)] or Without [PE(–EEF)]. For Comparison, σ^{2PA} Results Reported by Salem et al. are Included[14]

nCC	vacuum 6-31+G(d,p)†	6-31+G(d,p)§	pcseg-2§	PE(–EEF) 6-31+G(d,p)	PE(+EEF) 6-31+G(d,p)	pcseg-2
13	19.2	59.7	56.1	23.4	5.7	5.8
14	19.7	58.7	55.4	23.8	5.9	5.8
16a	17.2	32.2	31.1	10.5	2.6	2.7
16b	15.4	48.0	46.4	6.9	1.8	1.9
17	21.7	67.8	64.1	63.7	16.2	16.3
18	29.0	88.1	83.0	55.6	14.1	16.9
19	20.5	76.8	71.8	12.2	3.1	3.1
20	43.9	85.6	82.9	43.3	10.9	15.9
21	15.0	70.6	67.6	4.2	1.1	1.3
22	3.0	6.6	6.8	7.3	2.0	1.5

$\dagger$ Salem et al. results[14] obtained for smaller versions of the chromophores than the ones in Figure 5.1.
$\S$ Results obtained in this work for the isolated chromophores using the QM/MM-optimized geometries.

As discussed in the previous subsection, the chromophore geometries and, in particular, the tilt and twist angles obtained in this work differ from the ones determined by Salem et al.[14] due to the inclusion of the protein effects in the geometry optimization. Moreover, the present models use extended structures. These differences cause the 2PA cross-sections computed in the present work for the chromophores in vacuum to be on average three times larger than the cross-sections previously reported for the same set of chromophores[14] (left-hand side of Table 5.1). However, for the nCC–DsRed models, the effect that the disruption of the planarity, through the increase of tilt and twist angles, has on σ^{2PA} seems to be attenuated by the introduction of the protein. Figure 5.8 illustrates the variation of σ^{2PA} among the different methods.
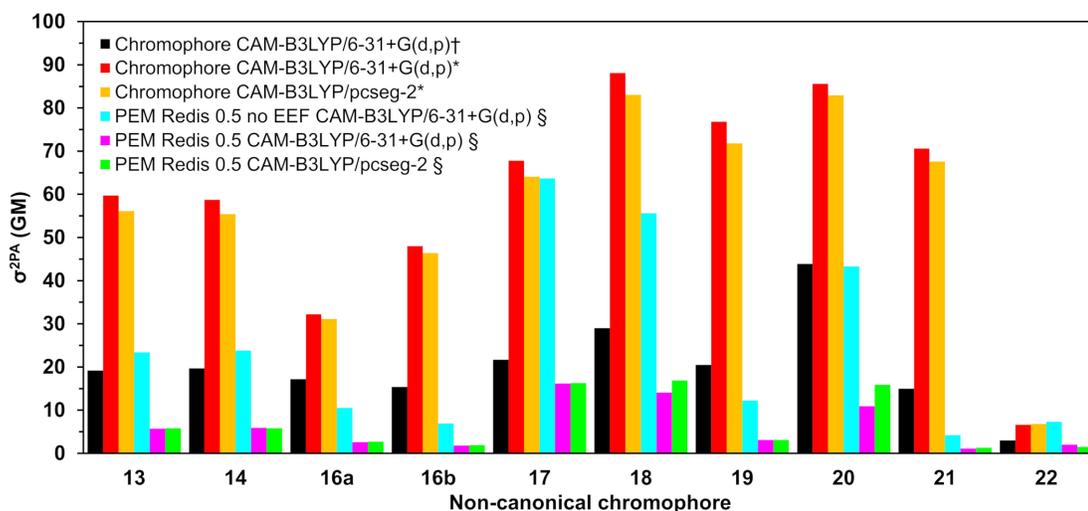
Figure 5.8. Two-photon absorption cross-sections for all non-canonical chromophores shown in Figure 5.1 computed †) in vacuum by Salem et al. using CAM-B3LYP/6-31+G(d,p),[14] in this work using QM/MM-optimized geometries ∗) in vacuum, and §) in protein (using the PE model).

The inclusion of local-field effects through the EEF approach [PE(+EEF)] led to an additional reduction in σ^{2PA} on top of the reduction already induced by the direct electrostatic interactions [PE(−EEF)], in comparison with the values obtained in vacuum. For the computations including EEF effects, the σ^{2PA} values are not affected by the basis set choice to any significant extent. In most cases, the difference between results determined using 6-31+G(d,p) and pcseg-2[228–230] is 0.1–0.2 GM, and the largest deviation is 5 GM for nCC 20. Previously, List et al.[121] tested the inclusion of EEF effects on the computation of 1PA and 2PA properties of DsRed. Their study showed that σ^{2PA} of DsRed (30 GM) was 3.5 times smaller than the σ^{2PA} obtained for the same system treated without EEF effects. Here, the σ^{2PA} values obtained for the nCC–DsRed chromophores align with the previous observations, i.e., they are reduced from 3.7 and up to 4.1 times upon inclusion of EEF. In addition, the results obtained here, including EEF effects in comparison with the results obtained by Salem et al.[14] in vacuum and without taking into account the protein in the optimization of the chromophores, are up to 14 times smaller. Unfortunately, and somewhat disappointingly, none of the nCC–DsRed models investigated in this work surpass the σ^{2PA} of CRQ-DsRed, previously computed by List

et al. [121] The latter is true for all models, including nCCs 20 and 21, which showed the largest cross-sections in previous studies (44 and 15 GM, respectively) of the isolated chromophores. In fact, nCC 21 has one of the smallest σ^{2PA}, whereas nCC 20 exhibits a σ^{2PA} similar to that of nCCs 17 and 18. The 2PA cross-section for the nCC 20-DsRed system, optimized using electrostatic embedding, was computed including EEF effects and using the pcseg-2 [228–230] basis set. The result, 21.9 GM, although larger than what was obtained for nCC 20-DsRed, optimized using mechanical embedding, i.e., 15.9 GM, is still inferior to what has been predicted previously in vacuum both in this work and by Salem et al. [14] It also is smaller than the 2PA cross-section computed for DsRed by List et al. [121] through a PE scheme. The fact that the difference in 2PA cross-section is rather modest suggests that a refinement of the structures using electrostatic embedding will not have a large impact on the results presented in Table 5.1, although it stresses the importance of the protein when determining the chromophore structure and multi-photon absorption properties. Similar optimizations of the rest of the systems using electrostatic embedding were not attempted.

5.4   Conclusions

The inclusion of protein effects in the geometry optimization of the nCC–DsRed systems studied here suggest that the identity of the substituent R– in the non-canonical chromophore (see Figure 5.1) does not have a significant impact on the geometry of the chromophore if such optimizations are carried out in vacuum, i.e., without the protein environment. More realistic pictures of the conformation of the non-canonical chromophores studied in this work needed to be addressed by including the protein environment in QM/MM strategies.

Although the nCC–DsRed systems evaluated in this work involve the same computational models of the non-canonical chromophores previously proposed and investigated by Salem et al., [14] their 2PA cross-sections here have here been evaluated in a more complete way, taking into account protein environment effects, which have not been con-

sidered before for these systems. We found that both direct electrostatic interactions and local-field effects have a large impact on the 2PA cross sections. For all chromophores, the 2PA cross sections decreased in comparison with previous studies carried out in vacuum, which highlights the critical role of the environment in the design of new FPs with large 2PA cross-sections. The results obtained in this work suggest that the choice of basis set should be done carefully when using QM/MM models, as diffuse functions can result in spurious molecular orbitals, whose impact on the σ^{2PA} computation has not been evaluated extensively.

In this work, the DsRed protein was chosen as the protein host for the set of non-canonical chromophores. Future work could involve the evaluation of 2PA properties of selected nCCs in other RFP hosts and/or a tailored environment, where amino acids surrounding the chromophore can be modified or substituted to tune its 1PA and 2PA properties.[15,86,103] Water molecules in the immediate surroundings of the chromophore can play a role in the absorption properties of fluorescent proteins.[16,68,286–288] However, we do not expect that the 2PA cross-sections obtained here would change so drastically that they surpass the 2PA cross-section limit exhibited by existing FPs. That said, it would be useful to establish their role, if present, on 2PA as well as examine the effects from solvent and conformational sampling.

Chapter 6

One and Two-photon Absorption Cross-Sections of the Fruits Series Fluorescent Proteins using QM and QM/pol-MM approaches

6.1 Introduction

The one-photon absorption (1PA) of fluorescent proteins (FPs) and how it can be tuned among different regions of the visible spectrum has been studied extensively, resulting in a wide variety of FPs' hues. The 1PA of FPs is commonly tuned by mutating the amino acids surrounding and/or contained in the chromophore.

The spectral properties of green fluorescent protein (GFP) are affected by mutations of the neighbouring residues, as they are involved in a hydrogen-bond network, which facilitates the deprotonation of the chromophore and stabilizes the resulting anion. The disruption of the H-bond network near the GFP chromophore occurs when the residues Lys61, Glu222, Thr203, and Ser205 are mutated.[289] In addition to point-mutations, the pH also can affect the spectroscopic properties of FPs. The pH can impact the conformation of the chromophore amino acids, such as in the case of the cis-trans conformation of tryptophan in the nowGFP chromophore,[289] which determines the chromophore deprotonation and, consequently, its 1PA properties. In fact, the replacement of tyrosine with tryptophan results in a shorter fluorescence wavelength.[289]

Despite the deep knowledge gained about the 1PA photophysical behaviour of FPs, it cannot be used to predict the multi-photon absorption (MPA) in FPs. Excellent MPA characteristics are the reasons behind the popularity of FPs in clinical imaging.[5,198] Furthermore, despite the structural and MPA information of FPs that has been obtained using experimental and computational approaches, there are no conclusive guidelines yet that can be used to engineer unambiguously red-shifted FPs with enhanced MPA

properties. For TagRFP, the 2PA and 2PA brightness ($\eta\sigma^{2PA}$, where η is the fluorescence quantum yield) within 700–1000 nm are 315 GM and 130 GM, respectively, whereas for tdTomato within 1000–1100 nm they are 216 GM and 120 GM, respectively.[197] In this context, the engineered FPs have a "raw" σ^{2PA} >300 GM or >200 GM to account, to an extent, for any foreseeable 2PA brightness depletion due to the experimental fluorescence quantum yield.

Some findings that could be used to enhance the 2PA ability of FPs and their chromophores have been previously discussed in experimental and computational investigations, and some specific examples will be mentioned next. The mutation of the residue positions near the chromophore in the mFruits series causes a rearrangement of the hydrogen-bond network in some cases and, in general, creates a more hydrophobic pocket around the chromophore.[58,284] Drobizhev et al. experimentally determined the σ^{2PA} for the mFruits and correlated this to the difference between the permanent ground- and excited-state electric dipole moments. Furthermore, they explained the hues of the mFruits through the quadratic Stark effect exerted by the protein on the chromophore.[58] Computationally, the extent of the impact the environment has on the 2PA has been examined on, e.g., the red- and yellow-fluorescent proteins.[15,86,103]

Herein, we investigate the two-photon absorption of a series of RFPs, including some exemplars of the mFruits series, and their chromophores. Through the computation of the σ^{2PA}, ground- and excited-state permanent dipole moments, as well as the transition electric dipole moments of the DsRed, mCherry (at pH = 8 and pH = 11), mOrange, mPlum, and mStrawberry RFPs, we aim to expand the available 2PA computational data for FPs, taking into account the environmental effects using a QM/pol-MM approach.[70] Computationally, the role of the different protonation states of the neighbouring residues around the chromophore in FPs has not been extensively explored.[74] Here, we discuss the impact of the different protolytic states of a selected set of amino acids around the mCherry chromophore at pH = 8 and pH = 11.

In addition, we aim to shed light on the possible pathways to enhance the 2PA ability of FPs. The results presented contribute to the creation of the guidelines necessary to engineer FPs with enhanced 2PA in a more rational manner. Also, the present results will help determine a suitable environment for the RFP-type nCC chromophores suggested by Salem et al., which have not been expressed experimentally.[14] Computationally, these nCCs have been studied previously by Salem et al. in vacuum[14] and by Rossano-Tapia et al. within the DsRed protein environment.[105] Motivated by the extraordinary 2PA that some of the FPs of the mFruits series exhibit (see Figure 1.3), we studied the 2PA of the RFP-form of the non-canonical amino acid 20 (nCC 20), whose structure is depicted in Figure 2.2, within the protein environment of the RFP mCherry at pH = 11. Bear in mind that the latter was decided based on the results published by Drobizhev et al. in 2009,[58] which suggest that the mCherry pH = 11 σ^{2PA} (75 GM) is three times that of mCherry at pH = 8 (25 GM). However, in a recent publication, the σ^{2PA} of mCherry has been determined experimentally as 33 and 24 GM at pH = 11.4 and 7.4, respectively,[290] suggesting that the 2PA difference between the two pH environments might not be as prominent as determined previously.

6.2 Computational Details

6.2.1 Optimization of the Protein Structures

The crystal structures of the proteins studied here, DsRed (PDB: 1ZGO),[7] mCherry (PDB:2H5Q), mOrange (PDB: 2H5O), mStrawberry (PDB: 2H5P),[284] and mPlum (PDB: 2QlG),[291] were obtained from the RCSB Protein Data Bank.[292] Any missing hydrogens in the crystal structures were added using the PDB2PQR[293] software at pH = 8 for DsRed, mOrange, mPlum, and mStrawberry. In the case of mCherry, the protonation of the crystal structure was done at pH = 8 as well as at pH = 11.

Subsequently, the protonated crystal structures were optimized using the QM/MM approach ONIOM,[269–271] as implemented in Gaussian16.[223] The optimization of the

protein structures were carried out in two steps. While treating in all cases the MM region with the ff96 Amber parameters, the QM region was described first using the semi-empirical method PM6[175] (PM6:Amber). Subsequently, a second optimization was performed on the resulting structures using the CAM-B3LYP[184] functional and the 6-31+G(d,p)[157] basis set (CAM-B3LYP/6-31+G(d,p):Amber). The MM parameters of the chromophores were obtained from the R.E.D. Server[272–275] using the default setup, i.e., RESP-A1 charges and the HF/6-31G(d) method, in a similar fashion to our previous investigation discussed in Chapter 5.[105]

For the geometry optimization of the protein structures, we included the chromophore and the covalently bonded neighbouring residues, Ser and Phe, in the QM region, as depicted on the left-hand side of Figure 6.1. The rest of the protein was treated using MM. In particular, the structure of mCherry at pH = 8 was optimized further using an enlarged QM region (right-hand side of Figure 6.1), which was defined based on a series of σ^{2PA} computations performed on the CAM-B3LYP/6-31+G(d,p):Amber optimized structure. Results of this preliminary stage are included in Table E.8 as part of the Appendix E.

Experimental findings suggest that the protonated form of Glu 210 is present in mCherry and mStrawberry at pH < 10. The former places Glu 210 interacting with the chromophore's imidazolinone moiety through a hydrogen bond.[284] PDB2PQR[293] predicts Glu 210 to be protonated in mCherry, but deprotonated in mStrawberry. In addition, the pK_a of the glutamic acid amino acid (4.15) suggests that the deprotonated form of Glu 210 should be present in the protein models at pH = 8. Therefore, both protonated forms of Glu 210 in mCherry (Figure 6.1) were studied here. The 2PA findings regarding the role of this amino acid could guide us through a more accurate modeling of mStrawberry at pH = 8. In the case of mOrange, Glu 210 is predicted by PDB2PQR[293] to be deprotonated, which agrees with experimental observations.[284]

a. SER 64
b. Chromophore 63
c. PHE 62
d. ARG 90
e. WAT 343
f. GLU 210
g. LYS 65
h. SER 141
i. WAT 256
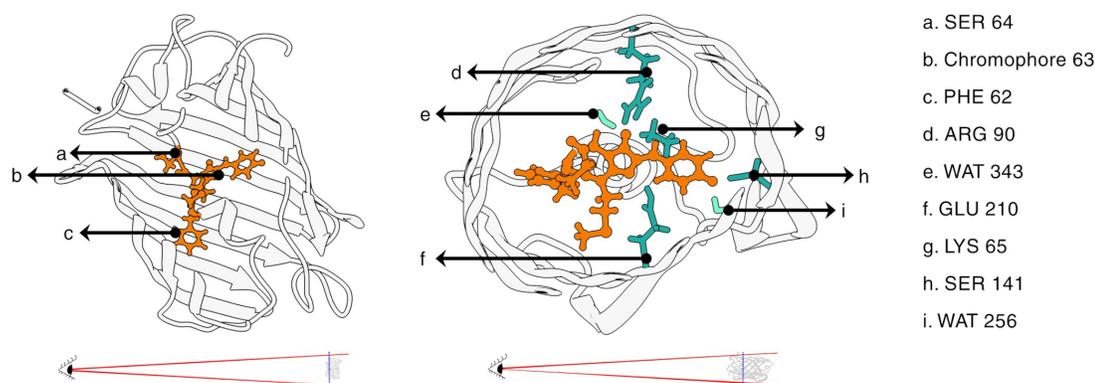
Figure 6.1. The protonated structure of mCherry (PDB:2H5Q)[284] is employed here to depict the system partition used in the QM/MM geometry optimization of the protein structures. Left-hand side: the QM region is defined by the chromophore as well as the residues Ser 64 and Phe 63, which are bonded covalently to the chromophore. Right-hand side: enlarged QM region used for the optimization of mCherry at pH = 8. The protein structures were created using the Chimera UCSF v.1.15 software.[294]

6.2.2   One- and Two-Photon Absorption Computations

Linear- and quadratic-response features were computed using the CAM-B3LYP[184] functional along with the pcseg-2 basis set.[228–230] All the response computations performed within the protein environment were carried out using the polarizable embedding model formulated by Olsen et al. and implemented in the Dalton software.[278,279,295] QM/MM response computations included the effective external field (EEF) effects.[121] For the cluster models, all computations were performed fully quantum mechanically using the Dalton software.[278,279,295]

6.2.3   Mutation of mCherry

The optimized structures of mCherry-11 and mCherry—nCC-20 were point-mutated at position 192 or 202 in the source PDB file.[284] The mCherry crystal structure obtained from the PDB Data Base does not include the coordinates of residues 1–8, as listed in the FASTA amino acid sequence. In addition, the residues Met 71, Tyr 72, Gly 73, and Ser 74 in the FASTA sequence, are labeled as CH6 66 and Ser 69 in the PDB:2H5Q file, thus, resulting in a shift of 10 positions in the residue index between our protein sequence and the source PDB:2H5Q file. The point-mutation I192S was performed using the Shapovalov and Dunbrack Jr. rotamer library[296] implemented in Chimera UCSF v.1.15.[294]

From the resulting mutated structures, we obtained four cluster models containing the residue combinations chromophore + Ser141 + Ser192 and chromophore + Ser192, where chromophore corresponds to CH6 (canonical mCherry chromophore) or nCC-20. These residue combinations were obtained as follows. The chromophore + Ser141 + Ser192 clusters were optimized using the Universal Force Field (UFF),[297] as implemented in the Avogadro suite v.1.2.0.[298] From the latter, the chromophore+Ser192 were extracted without carrying out any further geometry optimizations. The x, y, z, coordinates of the four point-mutated clusters are provided in the Appendix E.

6.3 Results and Discussion

6.3.1 Full FP Models

In their work, Drobizhev et al. refer to the RFPs TagRFP, mOrange, and mBanana as the best candidates for clinical microscopy within the 2PA region 1000–1100 nm because of their high 2PA brightness. Here, we only examined those FPs for which their crystal structure was available in the PDB protein data base, i.e., DsRed and mOrange.

In general, the structures we obtained after the optimization are in agreement with what has been observed experimentally, including the conformation towards the chromophore of the residues Glu 210 and Lys 65.[284] The arrangement of a selected set of neighbouring amino acids around the chromophore of mCherry at pH = 11 are shown in Figure 6.2.
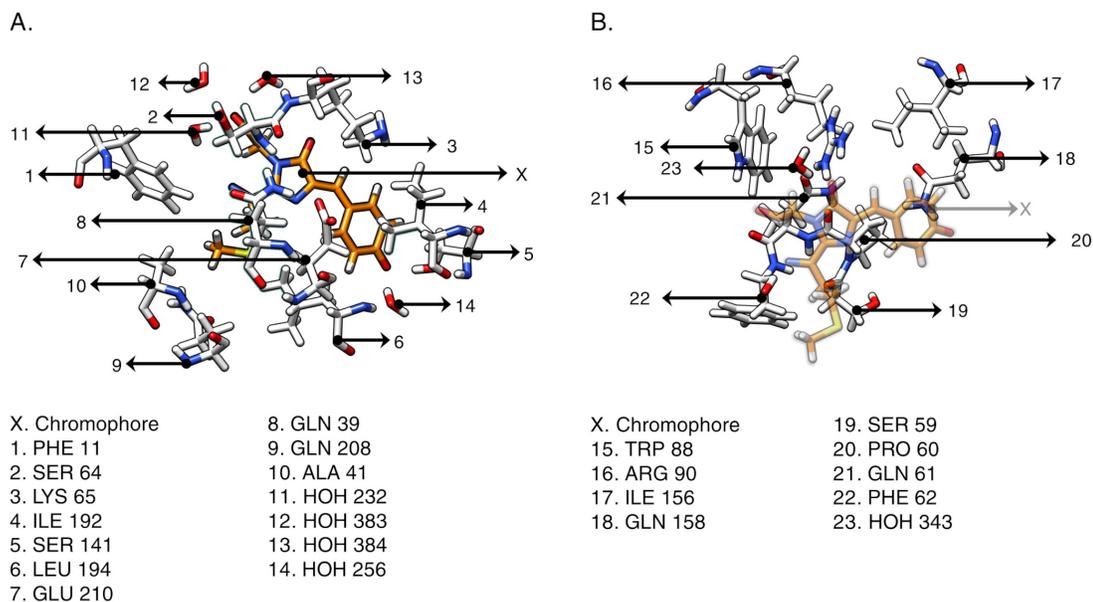
Figure 6.2. For the optimized structure of mCherry at pH = 11, the neighbouring residues for which the effect on the 2PA elements of the chromophore were evaluated (Table 6.2).[294]

X. Chromophore	8. GLN 39
1. PHE 11	9. GLN 208
2. SER 64	10. ALA 41
3. LYS 65	11. HOH 232
4. ILE 192	12. HOH 383
5. SER 141	13. HOH 384
6. LEU 194	14. HOH 256
7. GLU 210	

X. Chromophore	19. SER 59
15. TRP 88	20. PRO 60
16. ARG 90	21. GLN 61
17. ILE 156	22. PHE 62
18. GLN 158	23. HOH 343

To carry out a more fair comparison, we only studied the saturated form of the mOrange chromophore that leads to a typical RFP-like chromophore, i.e., we did not model the 2-hydroxy-dihydrooxazole possibly formed between the residues Phe 61 and Thr 62 according to experimental observations.[284] Nevertheless, the VEE computed for the saturated mOrange model employed is still slightly blue-shifted with respect to DsRed (Table 6.1), an effect attributed mainly to the formation of 2-hydroxy-dihydrooxazole.[284] Thus, we assume that the mOrange model employed here, i.e., without the oxazole ring, might have little impact on our results (Table 6.1). However, it will be interesting to study the 1PA and 2PA of mOrange containing the oxazole ring to verify its role.

Table 6.1. The Vertical Excitation Energies (VEEs) in eV and 2PA Cross-Sections (σ^{2PA}) in GM for the First Lowest Excited State S_1 of a Set of FP Chromophores in Vacuum and Within the Protein Environment. The Last Two Columns Correspond to the Previously Determined Experimental Absorption Energies (in eV) and σ^{2PA}s (in GM).[197] ND=No Data

Chromophore	Chromophore—Protein		In vacuum		Exp.	
	VEE	σ^{2PA}	VEE	σ^{2PA}	VEE	σ^{2PA}
DsRed	2.63	1	2.61	3	2.36[§,ζ], 2.21*	63[§], 55*, 73[ζ]
mCherry pH = 11	2.71	3	2.63	7	2.20*	33*
mCherry pH = 8[†]	2.40	2	2.71	9	2.30[§,ζ], 2.11*	5.5[§], 24*, 6.4[ζ]
mCherry pH = 8[‡]	2.64	2	2.61	7	”	”
mCherry pH = 8	2.38	3	2.61	8	”	”
mOrange	2.67	3	2.66	9	2.30[§,ζ]	37[§], 47[ζ]
mPlum a	2.61	1	2.63	4	2.05*, 2.24[ζ]	15*, 2.9[ζ]
mPlum b	2.22	0	2.63	10	”	”
mStrawberry	2.64	12	2.65	6	2.32[§,ζ]	5.2[§], 6.8[ζ]
nCC 20[ξ]	2.63	3	2.89	60	ND	ND

† Larger QM region in the ONIOM optimization.

‡ Glu 210, located above the chromophore, is deprotonated.

ξ Optimized within the mCherry pH = 11 protein barrel.

§ $\eta\sigma^{2PA}$ reported by Drobizhev et al. [197]

∗ σ^{2PA} reported by Drobizhev et al. [290]

ζ $\eta\sigma^{2PA}$ reported by Drobizhev et al. [5]

Glu 210 in Figure 6.2 is analogous to the Glu 22 residue in the wild-type green fluorescent protein (wtGFP). In the wtGFP and some of its mutants, the Glu 222 residue actively participates in the formation of a complex H-bond network that favors the deprotonation of the chromophore and stabilizes the resulting anion.[258,289] Thus, Glu 210 becomes a residue of special interest. In Table 6.1, the mCherry at pH = 8, where Glu 210 is protonated, has a lower VEE than the model where it is deprotonated (0.2 eV or about 40 nm), which is in agreement with previous experimental observations. Shu et al. attribute the red-shift of mCherry to the hydrogen-bond interaction between Glu 210 and the chromophore.[284] Whether the residue Glu 210 is protonated or not seems to have little effect by itself on the σ^{2PA} magnitude.

The computation of σ^{2PA} for DsRed has been done before at physiological pH (we assume is ≈7) and considering the factor N (in Equation 1.5 for σ^{2PA}) as 8, where here it has been set equal to 4.[86] Besides the value of N, some of the reasons that could have resulted in the difference between the value of σ^{2PA} we obtained (Table 6.1) with respect

to that reported before by List et al.,[86] 105.9 GM, are discussed next.

- The differences in the optimized structures, in particular, the conformation of the chromophore. Although the planarity in our model resembles that of List et al.'s model (Figure E.1), the acylimine moiety in our model is deviated from the cis-conformation[73] obtained by them.[86] Such a difference is focused on the angles $\alpha - \beta - \gamma - \delta$ and $\beta - \gamma - \delta - \epsilon$, depicted in Figure E.1. The difference between the dihedral angles in the two structures might be one of the main reasons behind the large disagreement of our result with respect to the one obtained previously. List et al. report a $\sigma^{2PA} = 16.6$ GM or 8.3 GM if N = 4 (as employed here) for the DsRed chromophore in vacuum (Figure E.1),[86] which is larger but still within the same order of magnitude to what we obtained, 3 GM (Table E.1). To investigate further the impact of these angles on the σ^{2PA} on the DsRed chromophore, we computed, at the TD-CAM-B3LYP/6-31+G(d,p) level of theory, the σ^{2PA} and VEE for the DsRed chromophore structure obtained here and that obtained by List et al. Both chromophores were truncated right after the acylimine moiety. As expected, the cis-like chromophore reported previously is red-shifted (2.58 eV) with respect to the chromophore we obtained (2.65 eV). Furthermore, the σ^{2PA} of List et al.'s (truncated) chromophore is larger, 20.4 GM, to that we obtained for our chromophore model, 3.1 GM. A comprehensive computational analysis in vacuum of the dependence between the σ^{2PA} with respect to the dihedral angle of the acylimine moiety of the RFP-type chromophore, has been reported previously.[14]

- The discrepancy between the σ^{2PA} obtained by List et al. and the one presented in Table 6.1 can be attributed also to the protolytic states of the neighbouring amino acids and the resulting local electric field, however, we believe that it is not the case, as the protolytic state of the residues at pH = 8 is similar at pH = 7. The value of σ^{2PA} without EEF effects is 2 GM, i.e., [PE(-EEF)], similar to that of [PE(+EEF)] in Table 6.1. Although the σ^{2PA} will not change for our optimized

101

DsRed structure, as the EEF effects are considered, the σ^{2PA} obtained previously[86] might do so, since the magnitude of the MPA strengths can vary up to 300% when the EEF effects are taken into account.[26]

- The size of the QM region has an impact on the σ^{2PA}, as has been discussed by Grabarek and Andruniów.[18] Also, through a series of computations we performed on the mCherry (pH = 11), we found that a region containing 70 atoms (PHE 62, Chromo 63, and SER 64) has a $\sigma^{2PA} = 1$ GM (Table 6.1), whereas a region comprised of 159 atoms (residues in Figure 6.1 + WAT 374 and WAT 275 H-bind to GLU 210) exhibits a $\sigma^{2PA} = 4.5 \times 10^{-3}$ GM. However, in the case of DsRed, we employed a similar QM region (Phe 60 + Chro 61 + Ser 62 = 70 atoms) to the one employed previously (some atoms of Gln 59 + Phe 60 + Chro 61 + Ser 62 = 63 atoms)[86] for this protein. Therefore, it is fair to say that the size of the QM region does not contribute largely to the difference in the σ^{2PA} we obtained and that reported previously.[86]

To be able to directly compare our 2PA results with those reported previously, it might be necessary to compute the σ^{2PA} of our DsRed model at the acylimine conformation, shown in Figure 2.2.[86] Even further, it will be pertinent to analyse the dependence between σ^{2PA} with respect to the angle $\beta - \gamma - \delta - \epsilon$, taking into account the environmental effects, in an analogous fashion to what has been done previously for the chromophore in vacuum.[14] Notice that the values of $\alpha - \beta - \gamma - \delta$ and $\beta - \gamma - \delta - \epsilon$ are not the only factors contributing to the 2PA of the rest of FPs studied here. Although, the $\alpha - \beta - \gamma - \delta$ and $\beta - \gamma - \delta - \epsilon$ angles of some of the FPs, e.g., mCherry and mOrange, are similar to those reported previously[86] (refer to Table E.7), their σ^{2PA} is not noticeably enhanced (Table 6.1).

The σ^{2PA} of the mFruits and their chromophores listed in Table 6.1 (except for the nCC 20) are plotted in Figure 6.3 against the difference between the excited- and ground-state permanent electric dipole moment ($\Delta\mu_{10}$). In Figure 6.3, the 2PA features

obtained within the protein environment (right-hand side of Figure 6.3) do not exhibit the quadratic dependence of σ^{2PA} with respect to $\Delta\mu_{10}$, as expected from experimental data.[5,58] However, our results need to be refined further, as we need to find the optimal size of the QM region at which σ^{2PA} converges.[18] However, those results will be part of a future publication.

a) Bare chromophores   b) Chromophores within the protein environment

a. DsRed   e. mCherry pH=8[‡]   i. mStrawberry
b. mCherry pH=11   f. mOrange   j. nCC 20[ξ]
c. mCherry pH=8[†]   g. mPlum a
d. mCherry pH=8   h. mPlum b

† A larger QM region was employed in the ONIOM optimization.
‡ Glu 210, located above the chromophore, is deprotonated.
ξ Optimized within the mCherry pH=11 protein environment.
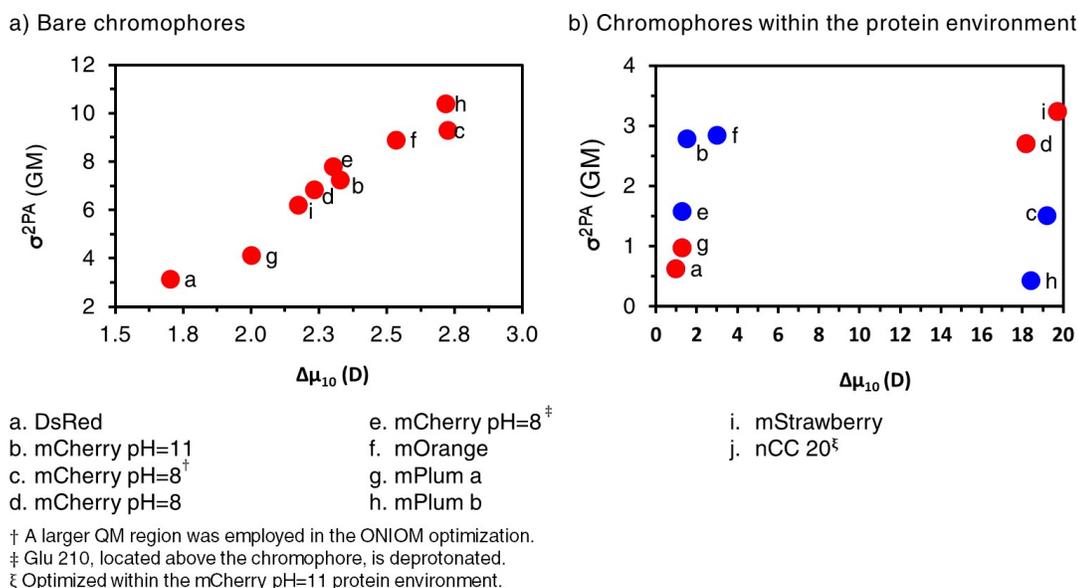
Figure 6.3. Depiction of the correlation between the 2PA cross-section (σ^{2PA}) and the difference between the excited- and ground-state permanent electric dipole moment ($\Delta\mu_{10}$) for the fluorescent proteins listed in Table 6.1 and their corresponding chromophores.

6.3.2   2PA of Cluster Models

The effect of a selected set of neighbouring residues on the σ^{2PA} of mCherry at pH = 8 and pH = 11 is analyzed through a series of clusters comprised of the chromophore and the residues of interest. The VEEs, OS, percentage of deviation of the cluster σ^{2PA} with respect to that of the chromophore in vacuum, ground-state electric dipole moment, transition electric dipole moment, and the difference between the excited- and ground-state permanent electric dipole moment ($\Delta\mu_{10}$) are listed in Tables E.8 and E.9. A shortened version of Table E.8 is presented as Table 6.2. We are aware that the results in these tables will change when the rest of the protein is taken into account, as they lack the collective effects of the whole protein. Nevertheless, our cluster-based approach

will guide us to determine what are the residues that most likely play a major role in the 2PA of the mCherry chromophore. A more "realistic" approach to the one implemented here will require, for example, the substitution of the residues of interest with residues that might be innocuous, such as Gly, and determination of the σ^{2PA} of the resulting structures.[86]

With respect to the chromophore in vacuum, the residues Glu 210, Ser 141, and Arg 90 (Figure 6.2) have the largest impact on the σ^{2PA} of mCherry at pH = 8 and pH = 11 (Tables E.8 and E.9). It is interesting that the role of Lys 65 strongly depends on its protolytic state. The σ^{2PA} of the chromophore + the charged form of Lys 65 is 44% smaller than the σ^{2PA} of the bare chromophore. On the other hand, the σ^{2PA} of the neutral (deprotonated) form of Lys 65 + chromophore cluster is only 11% smaller than in the case of the chromophore in vacuum.

In the model of mCherry protonated at pH = 11, the residue Lys 65 points away from the (formerly) tyrosine ring in the chromophore, whereas in its protonated form in mCherry at pH = 8, Lys 65 seems to be attracted to it. Such a dependence of the Lys 65 residue conformation and the pH at which mCherry is found has been discussed previously.[289] In mCherry at pH = 8, the protonated Lys 65 is located on top of the anionic ring of the tyrosine. Such an interaction between Lys 65 and the tyrosine of the chromophore translates into a depletion of 44% of the bare chromophore σ^{2PA} (Table E.9). The decrease of the σ^{2PA} of mCherry (pH 8) as Lys 65 interacts with the bare chromophore can be due to a "containment" of the electron density within a smaller surface, which has a direct impact on the electric dipole moment. Indeed, in previous investigations, it has been discussed that the presence of an analogous residue to Lys 65, Lys 61 in NowGFP[289], decreases the pK_a of the proton in the tryptophan of the chromophore with a sequence threonine-tryptophan-glycine, i.e., the electronic density is drawn to the cycle site, thus, making the proton more susceptible to be removed. On the other hand, it seems that the presence of a charged amino acid, such as Lys 163 in

the DsRed FP, in close contact to the charged phenolate moiety of the FP chromophore, is linked to an increase of its 2PA in comparison to those FP cases, where such an amino acid is absent or not interacting with the chromophore, as in mCherry, mStrawberry, and mOrange.[197]

Table 6.2. Effect of Neighbouring Residues in the mCherry FP (pH = 11) on the 2PA of the RFP-Type Chromophore. The bare chromophore has a $\sigma^{2PA} = 13$ GM. A more comprehensive version of this table is provided in Table E.8 as part of Appendix E

Residue	Water	VEE (eV)	OS (au)	% σ^{2PA}	μ_{00} (D)	μ_{01} (D)	$\Delta\mu$ (D)
Chro	NA	2.6	1.1	0.0	10.4	10.4	2.3
16	23	2.9	1.1	−41.1	20.0	9.9	1.8
16		2.9	1.1	−34.0	20.5	10.0	1.9
2	13, 23	2.8	0.8	−25.4	7.6	10.3	2.1
15	11, 13, 23	2.7	0.9	−23.9	11.7	10.3	2.2
2	13, 23, 283*, 374*	2.7	1.0	−23.1	7.8	10.3	2.1
21		2.5	0.9	−16.4	11.7	9.8	2.2
19		2.7	0.7	−14.9	12.1	9.8	2.2
18		2.6	1.1	23.9	7.6	10.7	2.4
5		2.6	1.2	39.6	4.6	10.8	2.4
3, 5, 7, 18		2.7	1.0	44.0	14.2	10.6	2.4
5	14	2.6	1.2	62.7	4.7	10.9	2.6
5, 7, 18		2.7	1.0	82.8	19.1	10.8	2.7
5, 18		2.6	1.2	85.8	3.6	11.1	2.7
5, 18	14, 324*	2.6	1.2	130.6	5.5	11.0	2.9

From examining Table 6.1 and Table 6.2, it is possible to notice that the inclusion of the environmental effects in the computation of 2PA using a polarizable embedding model, has a dramatic diminishing effect on the σ^{2PA} for the set of FPs studied here, except for mStrawberry (Table 6.1).

6.4   Conclusions and Future Work

Here we present some of the results that are part of an ongoing more comprehensive investigation performed on a series of RFPs, for which their 2PA cross-sections have been determined experimentally.[5,58,290] The results presented here aim to help in the rationalization of the 2PA behaviour of FPs through DsRed, mCherry, mStrawberry, mPlum, and the nCC 20. These insights aim to aid in the engineering of novel FPs with enhanced 2PA.[299,300]

The σ^{2PA} computed for the FPs (Table 6.1) are not particularly in good agreement

with experimental data, however, this does not come as a surprise as we did not take into account yet the convergence of σ^{2PA} with respect to the size of the QM region in the QM/pol-MM computations. Besides finding the optimal size of the QM region at which σ^{2PA} converges, it will be pertinent to look into the conformation of the acylimine at which σ^{2PA} is maximum,[14] as both factors seem to have a significant impact on the 2PA of DsRed (see discussion in Section 6.3.1).

Here, we also discussed the role that some of the neighbouring amino acids around the chromophore in mCherry at pH = 8 and pH = 11 play in determining its σ^{2PA}. In particular, we found that residues Ser 141 and Gln 158 (Figure 6.2) have the largest impact on the σ^{2PA} of the mCherry cluster models studied here.

To test the performance of the nCC 20 further as a possible FP chromophore, we computed its σ^{2PA} within the protein environment of mCherry at pH = 11. The results we found (σ^{2PA} = 3 GM) are similar to the results we obtained in a previous investigation, where nCC 20 was embedded in the DsRed protein environment.[105] However, these findings are not conclusive, as we did not examine different sizes of QM regions systematically. The results we obtained in regard to the $\Delta\mu_{10}$ will help us to tailor a protein environment around the nCC 20. In a future publication, we will present the 2PA resulting from point mutations (Section 6.2.3) around the nCC 20 within the environment of mCherry at pH = 11. In addition, the 1PA and 2PA features will be investigated with respect to the non-covalent interactions between the neighbouring amino acids and the chromophores as a means to determine how their polar character impacts the electronic density distribution and thus, the electric dipole moment and 2PA of the chromophore. Something that is not commonly discussed is the protolytic state of the amino acids around the FPs chromophore.[74] Here we found that the deprotonated form of Lys 65 has a larger impact (–44%) on the σ^{2PA} of mCherry at pH = 11 in comparison to when it is protonated in mCherry at pH = 8 (–11%). An even further comprehensive study of the 2PA features of the FPs analyzed here will require the introduction of dynamical effects

using molecular dynamics simulations.

Chapter 7

Conclusions and Future Work

The two-level model approximation for the computation of σ^{2PA}s of FP chromophores is an alternative to response theory computations, as Salem et al. have shown;[14] however, it is sensitive to variations in the magnitude of the vertical excitation energy or electric dipole moments, according to Equations 2.2, 2.5, 3.2, and 3.3. In Chapter 2, we discussed how the overestimation of the excited-state permanent electric dipole moments are most likely the reason behind the overestimation of the σ^{2PA}s of the series of RFP-type chromophores we studied. Based on the results presented in Chapters 2 and 3, we concluded that the semi-empirical methods TD-DFTB2 and LC-TD-DFTB require further developments to be used in the computation of multi-photon absorption properties. These would include obtaining parameters for a wider range of elements, including halides, parameters that are specific for computing ground- and excited-state electric dipole moments; as well as transition electric dipole moments; and higher-order (second- and third-) response theory implementations of the TD-DFTB methods.

In Chapter 4, we discussed for a series of dyes and chromophores in vacuum, the qualitative agreement between the σ^{3PA}s we computed and the experimental benchmark data.[195] We expect that the results we presented here will vary (hopefully improve) and quantitatively reproduce the experimental data, if environmental effects are taken into account. Future work could include the computation of σ^{3PA}s of the canonical chromophores listed in Table 4.3 using a polarizable embedding model, such as the one employed in Chapter 5.[23] In the context of cubic-response theory of fluorescent proteins and their chromophores, the σ^{3PA}s obtained here for fluorescein, rhodamine 6g, and serotonin will serve as computational benchmark data for future computations.

The set of nCCs discussed in Chapter 2 (Figure 2.2) had been investigated previously

using a series of TD-DFT and wave-function methods.[14] The σ^{2PA}s computed previously in vacuum suggest that some of these nCCs might be suitable as FP chromophores, however, the environmental effects remained to be explored. Therefore, we investigated the 2PA of a curated set of nCCs (Figure 5.1) within the DsRed protein sequence to take into account the environmental effects. The resulting σ^{2PA}s, presented in Chapter 5, are generally smaller than those of existing FPs and the same nCCs chromophores in vacuum. However, these results are not conclusive and most likely will change if a different protein is employed as the host environment in lieu of DsRed and/or molecular dynamics simulations are introduced to account for structural sampling.

Motivated by the results obtained in Chapter 5, we investigated the σ^{2PA}s of the nCC 20 within the environment of the mCherry FP at pH = 11. To have reference data to which we could compare the results of nCC 20—mCherry and to get a better understanding of the performance of the QM/pol-MM approach employed in Chapter 5, we investigated the 2PA features of a curated set of FPs of the mFruits series in Chapter 6. The mFruits we employed have been studied extensively using experimental means.[5] From our work, we conclude that the σ^{2PA}s computed for the mFruit proteins in Table 6.1 require further refinement to be suitable for comparison with experiment. That is, different sizes of the QM region need to be investigated to determine at which of them the σ^{2PA} of each mFruit converges. From the results obtained for the cluster models of mCherry at pH = 8 and pH = 11, it is possible to say that among the residues investigated, Ser 141, Glu 210, and Arg 90 have the largest effect on modifying the σ^{2PA} of the mCherry chromophore. In addition, we can conclude that the residues located around the charged oxygen of the phenolate moiety of the chromophore tend to enhance the σ^{2PA} of the chromophore.[5] In particular, the mutations around the nCC 20 within the mCherry (pH = 11) protein that possibly could enhance its σ^{2PA} without disrupting the H-bond network are Pro 60 and Ile 192. Based on the former, the hydrophobic residue Ile 192 was replaced by the polar amino acid serine. The results obtained from this mutation

will be part of a future publication.

In the introduction of this thesis, we discussed the important role that the environment, typically comprised of the protein barrel and crystal waters, plays in determining the multi-photon absorption features of FP chromophores. One of the challenges of the computational study of FPs is to obtain results that reproduce the experimental observations quantitatively. Although this major challenge might be difficult to achieve if the environmental effects are neglected, the examples provided over Chapters 2–4 and 6 reinforce the idea of using isolated chromophores to test the performance of computational methods in obtaining σ^{MPA}. Furthermore, these models are useful to perform preliminary computations before a more complex method, e.g., QM/pol-MM, is employed. Thus, the use of in vacuum and/or cluster models will prevail at least until the existing QM/pol-MM models become more intuitive to implement or until (less likely) the entire protein can be studied at a QM level.[36]

Bibliography

[1] Shimomura, O. The discovery of aequorin and green fluorescent protein. *J. Microsc.* **2005**, *217*, 3–15.

[2] Shimomura, O.; Johnson, F. H.; Saiga, Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, aequorea. *J. Cell. Physiol.* **1962**, *59*, 223–239.

[3] Shimomura, O. Structure of the chromophore of aequorea green fluorescent protein. *FEBS Lett.* **1979**, *104*, 220–222.

[4] Chudakov, D. M.; Matz, M. V.; Lukyanov, S.; Lukyanov, K. A. Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol. Rev.* **2010**, *90*, 1103–1163.

[5] Drobizhev, M.; Makarov, N. S.; Tillo, S. E.; Hughes, T. E.; Rebane, A. Two-photon absorption properties of fluorescent proteins. *Nat. Methods.* **2011**, *8*, 393–399.

[6] Cody, C. W.; Prasher, D. C.; Westler, W. M.; Prendergast, F. G.; Ward, W. W. Chemical structure of the hexapeptide chromophore of the aequorea green-fluorescent protein. *Biochemistry.* **1993**, *32*, 1212–1218.

[7] Tubbs, J. L.; Tainer, J. A.; Getzoff, E. D. Crystallographic structures of discosoma red fluorescent protein with immature and mature chromophores: linking peptide bond trans-cis isomerization and acylimine formation in chromophore maturation. *Biochemistry.* **2005**, *44*, 9833–9840.

[8] Timerghazin, Q. K.; Carlson, H. J.; Liang, C.; Campbell, R. E.; Brown, A. Computational prediction of absorbance maxima for a structurally diverse series of

engineered green fluorescent protein chromophores. *J. Phys. Chem. B*. **2008**, *112*, 2533–2541.

[9] Nifosì, R.; Luo, Y. Predictions of novel two-photon absorption bands in fluorescent proteins. *J. Phys. Chem. B*. **2007**, *111*, 14043–14050.

[10] Nifosì, R.; Tozzini, V. In *Fluorescent Proteins I: From Understanding to Design*; Jung, G., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; pp 3–40.

[11] Nifosì, R.; Luo, Y. Origin of the anomalous two-photon absorption in fluorescent protein dsred. *J. Phys. Chem. B*. **2007**, *111*, 505–507.

[12] Salem, M. A.; Brown, A. Two-photon absorption of fluorescent protein chromophores incorporating non-canonical amino acids: td-dft screening and classical dynamics. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25563–25571.

[13] Salem, M. A.; Brown, A. Two-photon absorption in fluorescent protein chromophores: tddft and cc2 results. *J. Chem. Theory Comput.* **2014**, *10*, 3260–3269.

[14] Salem, M. A.; Twelves, I.; Brown, A. Prediction of two-photon absorption enhancement in red fluorescent protein chromophores made from non-canonical amino acids. *Phys. Chem. Chem. Phys.* **2016**, *18*, 24408–24416.

[15] Beerepoot, M. T. P.; Friese, D. H.; Ruud, K. Intermolecular charge transfer enhances two-photon absorption in yellow fluorescent protein. *Phys. Chem. Chem. Phys.* **2014**, *16*, 5958–5964.

[16] Şimşek, Y.; Brown, A. Two-photon absorption properties of gold fluorescent protein: a combined molecular dynamics and quantum chemistry study. *J. Phys. Chem. B*. **2018**, *122*, 5738–5748.

[17] Grabarek, D.; Andruniów, T. Illuminating the origins of two-photon absorption

properties in fluorescent protein chromophores. *Int. J. Quantum Chem.* **2020**, *120*, e26086.

[18] Grabarek, D.; Andruniów, T. What is the optimal size of the quantum region in embedding calculations of two-photon absorption spectra of fluorescent proteins? *J. Chem. Theory Comput.* **2020**, *16*, 6439–6455.

[19] Grabarek, D.; Andruniów, T. Removing artifacts in polarizable embedding calculations of one-and two-photon absorption spectra of fluorescent proteins. *J. Chem. Phys.* **2020**, *153*, 215102.

[20] Weissleder, R. A clearer vision for in vivo imaging. *Nat. Biotechnol.* **2001**, *19*, 316–317.

[21] Friedrich, D. M. Two-photon molecular spectroscopy. *J. Chem. Educ.* **1982**, *59*, 472.

[22] Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

[23] Olsen, J. M. H.; Aidas, K.; Kongsted, J. Excited states in solution through polarizable embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

[24] Olsen, J. M. H.; Kongsted, J. In *Chapter 3 -molecular properties through polarizable embedding*; Sabin, J. R., Brändas, E., Eds.; Advances in Quantum Chemistry; Academic Press, 2011; Vol. 61; pp 107–143.

[25] Steindal, A. H.; Olsen, J. M. H.; Frediani, L.; Kongsted, J.; Ruud, K. Parallelization of the polarizable embedding scheme for higher-order response functions. *Mol. Phys.* **2012**, *110*, 2579–2586.

[26] Steindal, A. H.; Beerepoot, M. T. P.; Ringholm, M.; List, N. H.; Ruud, K.; Kong-sted, J.; Olsen, J. M. H. Open-ended response theory with polarizable embedding: multiphoton absorption in biomolecular systems. *Phys. Chem. Chem. Phys.* **2016**, *18*, 28339–28352.

[27] Reinholdt, P.; Kongsted, J.; Olsen, J. M. H. Polarizable density embedding: a solution to the electron spill-out problem in multiscale modeling. *J. Phys. Chem. Lett.* **2017**, *8*, 5949–5958.

[28] Cao, L.; Ryde, U. On the difference between additive and subtractive qm/mm calculations. *Front. Chem.* **2018**, *6*, 89.

[29] Dick, B. G.; Overhauser, A. W. Theory of the dielectric constants of alkali halide crystals. *Phys. Rev.* **1958**, *112*, 90–103.

[30] Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force fields: application to liquid water. *J. Chem. Phys.* **1994**, *101*, 6141–6156.

[31] List, N. H.; Olsen, J. M. H.; Kongsted, J. Excited states in large molecular systems through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20234–20250.

[32] Menger, M. F. S. J.; Caprasecca, S.; Mennucci, B. Excited-state gradients in polarizable qm/mm models: an induced dipole formulation. *J. Chem. Theory Comput.* **2017**, *13*, 3778–3786.

[33] Caprasecca, S.; Jurinovich, S.; Viani, L.; Curutchet, C.; Mennucci, B. Geometry optimization in polarizable qm/mm models: the induced dipole formulation. *J. Chem. Theory Comput.* **2014**, *10*, 1588–1598.

[34] Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. An empirical polarizable

force field based on the classical drude oscillator model: development history and recent applications. *Chem. Rev.* **2016**, *116*, 4983–5013.

[35] Bondanza, M.; Nottoli, M.; Cupellini, L.; Lipparini, F.; Mennucci, B. Polarizable embedding qm/mm: the future gold standard for complex (bio)systems? *Phys. Chem. Chem. Phys.* **2020**, *22*, 14433–14448.

[36] Loco, D.; Lagardère, L.; Adjoua, O.; Piquemal, J.-P. Atomistic polarizable embeddings: energy, dynamics, spectroscopy, and reactivity. *Acc. Chem. Res.* **2021**, ASAP, DOI: 10.1021/acs.accounts.0c00662.

[37] Martin, M. E.; Negri, F.; Olivucci, M. Origin, nature, and fate of the fluorescent state of the green fluorescent protein chromophore at the caspt2//casscf resolution. *J. Am. Chem. Soc.* **2004**, *126*, 5452–5464.

[38] Smyrnova, D.; Marín, M. d. C.; Olivucci, M.; Ceulemans, A. Systematic excited state studies of reversibly switchable fluorescent proteins. *J. Chem. Theory Comput.* **2018**, *14*, 3163–3172.

[39] Morozov, D.; Groenhof, G. Hydrogen bond fluctuations control photochromism in a reversibly photo-switchable fluorescent protein. *Angew. Chem. Int. Ed.* **2016**, *55*, 576–578.

[40] Senn, H. M.; Thiel, W. Qm/mm methods for biomolecular systems. *Angew. Chem. Int. Ed. 48*, 1198–1229.

[41] Pedraza-González, L.; Marín, M. d. C.; De Vico, L.; Yang, X.; Olivucci, M. In *QM/MM Studies of Light-responsive Biological Systems*; Andruniów, T., Olivucci, M., Eds.; Springer International Publishing: Cham, 2021; pp 1–75.

[42] Navizet, I. In *QM/MM Studies of Light-responsive Biological Systems*; An-

druniów, T., Olivucci, M., Eds.; Springer International Publishing: Cham, 2021; pp 227–270.

[43] Bravaya, K. B.; Grigorenko, B. L.; Nemukhin, A. V.; Krylov, A. I. Quantum chemistry behind bioimaging: insights from ab initio studies of fluorescent proteins and their chromophores. *Acc. Chem. Res.* **2012**, *45*, 265–275.

[44] Neugebauer, J. Subsystem-based theoretical spectroscopy of biomolecules and biomolecular assemblies. *ChemPhysChem.* **2009**, *10*, 3148–3173.

[45] Nemukhin, A. V.; Grigorenko, B. L. In *QM/MM Studies of Light-responsive Biological Systems*; Andruniów, T., Olivucci, M., Eds.; Springer International Publishing: Cham, 2021; pp 271–292.

[46] Acharya, A.; Bogdanov, A. M.; Grigorenko, B. L.; Bravaya, K. B.; Nemukhin, A. V.; Lukyanov, K. A.; Krylov, A. I. Photoinduced chemistry in fluorescent proteins: curse or blessing? *Chem. Rev.* **2017**, *117*, 758–795.

[47] Gross, L. A.; Baird, G. S.; Hoffman, R. C.; Baldridge, K. K.; Tsien, R. Y. The structure of the chromophore within dsred, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 11990–11995.

[48] Send, R.; Kaila, V. R. I.; Sundholm, D. Benchmarking the approximate second-order coupled-cluster method on biochromophores. *J. Chem. Theory Comput.* **2011**, *7*, 2473–2484.

[49] Filippi, C.; Zaccheddu, M.; Buda, F. Absorption spectrum of the green fluorescent protein chromophore: a difficult case for ab initio methods? *J. Chem. Theory Comput.* **2009**, *5*, 2074–2087.

[50] Nanda, K. D.; Krylov, A. I. Two-photon absorption cross sections within equation-of-motion coupled-cluster formalism using resolution-of-the-identity and

cholesky decomposition representations: theory, implementation, and benchmarks. *J. Chem. Phys.* **2015**, *142*, 064118.

[51] Christiansen, O.; Koch, H.; Jörgensen, P. The second-order approximate coupled cluster singles and doubles model cc2. *Chem. Phys. Lett.* **1995**, *243*, 409–418.

[52] Amat, P.; Granucci, G.; Buda, F.; Persico, M.; Tozzini, V. The chromophore of asfp595: a theoretical study. *J. Phys. Chem. B.* **2006**, *110*, 9348–9353.

[53] Drobizhev, M.; Callis, P. R.; Nifosì, R.; Wicks, G.; Stoltzfus, C.; Barnett, L.; Hughes, T. E.; Sullivan, P.; Rebane, A. Long-and short-range electrostatic fields in gfp mutants: implications for spectral tuning. *Sci. Rep.* **2015**, *5*, 13223.

[54] Langeland, J.; Kjær, C.; Andersen, L. H.; Brøndsted Nielsen, S. The effect of an electric field on the spectroscopic properties of the isolated green fluorescent protein chromophore anion. *ChemPhysChem.* **2018**, *19*, 1686–1690.

[55] Webber, N. M.; Meech, S. R. Electronic spectroscopy and solvatochromism in the chromophore of gfp and the y66f mutant. *Photochem. Photobiol. Sci.* **2007**, *6*, 976–981.

[56] Lin, C.-Y.; Boxer, S. G. Mechanism of color and photoacidity tuning for the protonated green fluorescent protein chromophore. *J. Am. Chem. Soc.* **2020**, *142*, 11032–11041.

[57] Clark, T. B.; Orr, M. E.; Flynn, D. C.; Goodson, T. Synthesis and optical properties of two-photon absorbing gfp-type probes. *J. Phys. Chem. C.* **2011**, *115*, 7331–7338.

[58] Drobizhev, M.; Tillo, S.; Makarov, N. S.; Hughes, T. E.; Rebane, A. Color hues in red fluorescent proteins are due to internal quadratic stark effect. *J. Phys. Chem. B.* **2009**, *113*, 12860–12864.

[59] Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N. G.; Palmer, A. E.; Tsien, R. Y. Improved monomeric red, orange and yellow fluorescent proteins derived from discosoma sp. Red fluorescent protein. *Nat. Biotechnol.* **2004**, *22*, 1567–1572.

[60] Beerepoot, M. T. P.; Friese, D. H.; List, N. H.; Kongsted, J.; Ruud, K. Benchmarking two-photon absorption cross sections: performance of cc2 and cam-b3lyp. *Phys. Chem. Chem. Phys.* **2015**, *17*, 19306–19314.

[61] Rossano-Tapia, M.; Brown, A. Determination of two-photon-absorption cross sections using time-dependent density functional theory tight binding: application to fluorescent protein chromophores. *J. Chem. Theory Comput.* **2019**, *15*, 3153–3161.

[62] Nifosì, R.; Amat, P.; Tozzini, V. Variation of spectral, structural, and vibrational properties within the intrinsically fluorescent proteins family: a density functional study. *J. Comput. Chem.* **2007**, *28*, 2366–2377.

[63] Laino, T.; Nifosì, R.; Tozzini, V. Relationship between structure and optical properties in green fluorescent proteins: a quantum mechanical study of the chromophore environment. *Chem. Phys.* **2004**, *298*, 17–28.

[64] Send, R.; Suomivuori, C.-M.; Kaila, V. R. I.; Sundholm, D. Coupled-cluster studies of extensive green fluorescent protein models using the reduced virtual space approach. *J. Phys. Chem. B.* **2015**, *119*, 2933–2945.

[65] Marques, M. A. L.; López, X.; Varsano, D.; Castro, A.; Rubio, A. Time-dependent density-functional approach for biological chromophores: the case of the green fluorescent protein. *Phys. Rev. Lett.* **2003**, *90*, 258101.

[66] Bravaya, K. B.; Khrenova, M. G.; Grigorenko, B. L.; Nemukhin, A. V.; Krylov, A. I. Effect of protein environment on electronically excited and ionized

states of the green fluorescent protein chromophore. *J. Phys. Chem. B*. **2011**, *115*, 8296–8303.

[67] Laurent, A. D.; Mironov, V. A.; Chapagain, P. P.; Nemukhin, A. V.; Krylov, A. I. Exploring structural and optical properties of fluorescent proteins by squeezing: modeling high-pressure effects on the mstrawberry and mcherry red fluorescent proteins. *J. Phys. Chem. B*. **2012**, *116*, 12426–12440.

[68] Faraji, S.; Krylov, A. I. On the nature of an extended stokes shift in the mplum fluorescent protein. *J. Phys. Chem. B*. **2015**, *119*, 13052–13062.

[69] Kaila, V. R. I.; Send, R.; Sundholm, D. Electrostatic spectral tuning mechanism of the green fluorescent protein. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4491–4495.

[70] Schwabe, T.; Beerepoot, M. T. P.; Olsen, J. M. H.; Kongsted, J. Analysis of computational models for an accurate study of electronic excitations in gfp. *Phys. Chem. Chem. Phys.* **2015**, *17*, 2582–2588.

[71] Petrone, A.; Caruso, P.; Tenuta, S.; Rega, N. On the optical absorption of the anionic gfp chromophore in vacuum, solution, and protein. *Phys. Chem. Chem. Phys.* **2013**, *15*, 20536–20544.

[72] Hasegawa, J.-Y.; Fujimoto, K.; Swerts, B.; Miyahara, T.; Nakatsuji, H. Excited states of gfp chromophore and active site studied by the sac-ci method: effect of protein-environment and mutations. *J. Comput. Chem.* **2007**, *28*, 2443–2452.

[73] Sanchez-Garcia, E.; Doerr, M.; Thiel, W. Qm/mm study of the absorption spectra of dsred.m1 chromophores. *J. Comput. Chem.* **2010**, *31*, 1603–1612.

[74] Filippi, C.; Buda, F.; Guidoni, L.; Sinicropi, A. Bathochromic shift in green fluorescent protein: a puzzle for QM/MM approaches. *J. Chem. Theory Comput.* **2012**, *8*, 112–124.

119

[75] Murugan, N. A.; Kongsted, J.; Rinkevicius, Z.; Ågren, H. Color modeling of protein optical probes. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1107–1112.

[76] Hu, L.; Söderhjelm, P.; Ryde, U. On the convergence of qm/mm energies. *J. Chem. Theory Comput.* **2011**, *7*, 761–777.

[77] Beerepoot, M. T.; Steindal, A. H.; Ruud, K.; Olsen, J. M. H.; Kongsted, J. Convergence of environment polarization effects in multiscale modeling of excitation energies. *Comput. Theor. Chem.* **2014**, *1040-1041*, 304–311.

[78] Nåbo, L. J.; Olsen, J. M. H.; Martínez, T. J.; Kongsted, J. The quality of the embedding potential is decisive for minimal quantum region size in embedding calculations: the case of the green fluorescent protein. *J. Chem. Theory Comput.* **2017**, *13*, 6230–6236.

[79] Kjellgren, E. R.; Olsen, J. M. H.; Kongsted, J. Importance of accurate structures for quantum chemistry embedding methods: which strategy is better? *J. Chem. Theory Comput.* **2018**, *14*, 4309–4319.

[80] Moron, V.; Marazzi, M.; Wanko, M. Far red fluorescent proteins: where is the limit of the acylimine chromophore? *J. Chem. Theory Comput.* **2019**, *15*, 4228–4240.

[81] Schwabe, T.; Olsen, J. M. H.; Sneskov, K.; Kongsted, J.; Christiansen, O. Solvation effects on electronic transitions: exploring the performance of advanced solvent potentials in polarizable embedding calculations. *J. Chem. Theory Comput.* **2011**, *7*, 2209–2217.

[82] Eriksen, J. J.; Olsen, J. M. H.; Aidas, K.; Ågren, H.; Mikkelsen, K. V.; Kongsted, J. Computational protocols for prediction of solute nmr relative chemical shifts. A case study of l-tryptophan in aqueous solution. *J. Comput. Chem.* **2011**, *32*, 2853–2864.

[83] Beerepoot, M. T. P.; Steindal, A. H.; Kongsted, J.; Brandsdal, B. O.; Frediani, L.; Ruud, K.; Olsen, J. M. H. A polarizable embedding dft study of one-photon absorption in fluorescent proteins. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4735–4743.

[84] Marefat Khah, A.; Reinholdt, P.; Olsen, J. M. H.; Kongsted, J.; Hättig, C. Avoiding electron spill-out in qm/mm calculations on excited states with simple pseudopotentials. *J. Chem. Theory Comput.* **2020**, *16*, 1373–1381.

[85] Steindal, A. H.; Olsen, J. M. H.; Ruud, K.; Frediani, L.; Kongsted, J. A combined quantum mechanics/molecular mechanics study of the one-and two-photon absorption in the green fluorescent protein. *Phys. Chem. Chem. Phys.* **2012**, *14*, 5440–5451.

[86] List, N. H.; Olsen, J. M. H.; Jensen, H. J. A.; Steindal, A. H.; Kongsted, J. Molecular-level insight into the spectral tuning mechanism of the dsred chromophore. *J. Phys. Chem. Lett.* **2012**, *3*, 3513–3521.

[87] Creemers, T. M. H.; Lock, A. J.; Subramaniam, V.; Jovin, T. M.; Völker, S. Photophysics and optical switching in green fluorescent protein mutants. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2974–2978.

[88] Drobizhev, M.; Makarov, N. S.; Hughes, T.; Rebane, A. Resonance enhancement of two-photon absorption in fluorescent proteins. *J. Phys. Chem. B.* **2007**, *111*, 14051–14054.

[89] Matz, M. V.; Fradkov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelov, M. L.; Lukyanov, S. A. Fluorescent proteins from nonbioluminescent anthozoa species. *Nat. Biotechnol.* **1999**, *17*, 969–973.

[90] Nørby, M. S.; Olsen, J. M. H.; Kongsted, J.; Jensen, H. J. A. Multipole moments for embedding potentials: exploring different atomic allocation algorithms. *J. Comput. Chem.* **2016**, *37*, 1887–1896.

[91] Beerepoot, M. T. P.; Steindal, A. H.; List, N. H.; Kongsted, J.; Olsen, J. M. H. Averaged solvent embedding potential parameters for multiscale modeling of molecular properties. *J. Chem. Theory Comput.* **2016**, *12*, 1684–1695.

[92] Nørby, M. S.; Steinmann, C.; Olsen, J. M. H.; Li, H.; Kongsted, J. Computational approach for studying optical properties of dna systems in solution. *J. Chem. Theory Comput.* **2016**, *12*, 5050–5057.

[93] Reinholdt, P.; Kjellgren, E. R.; Steinmann, C.; Olsen, J. M. H. Cost-effective potential for accurate polarizable embedding calculations in protein environments. *J. Chem. Theory Comput.* **2020**, *16*, 1162–1174.

[94] Bolnykh, V.; Olsen, J. M. H.; Meloni, S.; Bircher, M. P.; Ippoliti, E.; Carloni, P.; Rothlisberger, U. Extreme scalability of dft-based qm/mm md simulations using mimic. *J. Chem. Theory Comput.* **2019**, *15*, 5601–5613.

[95] Olsen, J. M. H. Pyframe: python tools for fragment-based multiscale embedding. 2018; https://doi.org/10.5281/zenodo.1443314.

[96] Olsen, J. M. H.; List, N. H.; Kristensen, K.; Kongsted, J. Accuracy of protein embedding potentials: an analysis in terms of electrostatic potentials. *J. Chem. Theory Comput.* **2015**, *11*, 1832–1842.

[97] Steinmann, C.; Reinholdt, P.; Nørby, M. S.; Kongsted, J.; Olsen, J. M. H. Response properties of embedded molecules through the polarizable embedding model. *Int. J. Quantum Chem.* **2019**, *119*, e25717.

[98] Grigorenko, B. L.; Nemukhin, A. V. Modeling trans-cis chromophore isomerization for the asfp595 kindling protein. Genetically Engineered and Optical Probes for Biomedical Applications IV. 2007; pp 165–169.

[99] Grigorenko, B. L.; Nemukhin, A. V.; Polyakov, I. V.; Krylov, A. I. Triple-decker motif for red-shifted fluorescent protein mutants. *J. Phys. Chem. Lett.* **2013**, *4*, 1743–1747.

[100] Hasegawa, J.-y.; Ise, T.; Fujimoto, K. J.; Kikuchi, A.; Fukumura, E.; Miyawaki, A.; Shiro, Y. Excited states of fluorescent proteins, mko and dsred: chromophore-protein electrostatic interaction behind the color variations. *J. Phys. Chem. B.* **2010**, *114*, 2971–2979.

[101] Grigorenko, B. L.; Polyakov, I. V.; Savitsky, A. P.; Nemukhin, A. V. Unusual emitting states of the kindling fluorescent protein: appearance of the cationic chromophore in the gfp family. *J. Phys. Chem. B.* **2013**, *117*, 7228–7234.

[102] Park, J. W.; Rhee, Y. M. Emission shaping in fluorescent proteins: role of electrostatics and π-stacking. *Phys. Chem. Chem. Phys.* **2016**, *18*, 3944–3955.

[103] Nifosì, R.; Mennucci, B.; Filippi, C. The key to the yellow-to-cyan tuning in the green fluorescent protein family is polarisation. *Phys. Chem. Chem. Phys.* **2019**, *21*, 18988–18998.

[104] Shelaev, I.; Mironov, V.; Rusanov, A.; Gostev, F.; Bochenkova, A.; Sarkisov, O.; Nemukhin, A.; Savitsky, A. The origin of radiationless conversion of the excited state in the kindling fluorescent protein (kfp): femtosecond studies and quantum modeling. *Laser Phys. Lett.* **2011**, *8*, 469–474.

[105] Rossano-Tapia, M.; Olsen, J. M. H.; Brown, A. Two-photon absorption cross-sections in fluorescent proteins containing non-canonical chromophores using polarizable qm/mm. *Front. Mol. Biosci.* **2020**, *7*, 111.

[106] Kulakova, A. M.; Khrenova, M. G.; Nemukhin, A. V. Simulation of spectra of red fluorescent protein mutants. *Moscow Univ. Chem. Bull.* **2018**, *73*, 212–215.

[107] Wanko, M.; García-Risueño, P.; Rubio, A. Excited states of the green fluorescent protein chromophore: performance of ab initio and semi-empirical methods. *Phys. Status Solidi B.* **2012**, *249*, 392–400.

[108] Sun, Q.; Li, Z.; Lan, Z.; Pfisterer, C.; Doerr, M.; Fischer, S.; Smith, S. C.; Thiel, W. Isomerization mechanism of the hcred fluorescent protein chromophore. *Phys. Chem. Chem. Phys.* **2012**, *14*, 11413–11424.

[109] Rossano-Tapia, M.; Brown, A. Quantum mechanical/molecular mechanical studies of photophysical properties of fluorescent proteins. *WIREs Comput Mol Sci.* **2021**, e1557.

[110] Aradi, B.; Hourahine, B.; Frauenheim, T. Dftb+, a sparse matrix-based implementation of the dftb method. *J. Phys. Chem. A.* **2007**, *111*, 5678–5684.

[111] Humeniuk, A.; Mitrić, R. Dftbaby: a software package for non-adiabatic molecular dynamics simulations based on long-range corrected tight-binding td-dft(b). *Comput. Phys. Commun.* **2017**, *221*, 174—202.

[112] Zipfel, W. R.; Williams, R. M.; Webb, W. W. Nonlinear magic: multiphoton microscopy in the biosciences. *Nat. Biotechnol.* **2003**, *21*, 1369–1377.

[113] Goulding, A.; Shrestha, S.; Dria, K.; Hunt, E.; Deo, S. K. Red fluorescent protein variants with incorporated non-natural amino acid analogues. *Protein Eng. Des. Sel.* **2008**, *21*, 101–106.

[114] Villa, J. K.; Tran, H.-A.; Vipani, M.; Gianturco, S.; Bhasin, K.; Russell, B. L.; Harbron, E. J.; Young, D. D. Fluorescence modulation of green fluorescent protein using fluorinated unnatural amino acids. *Molecules.* **2017**, *22*, 1194.

[115] Runge, E.; Gross, E. K. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* **1984**, *52*, 997.

[116] Grabarek, D.; Andruniów, T. Assessment of functionals for tddft calculations of one-and two-photon absorption properties of neutral and anionic fluorescent proteins chromophores. *J. Chem. Theory Comput.* **2019**, *15*, 490.

[117] Kamarchik, E.; Krylov, A. I. Non-condon effects in the one-and two-photon absorption spectra of the green fluorescent protein. *J. Phys. Chem. Lett.* **2011**, *2*, 488–492.

[118] Salem, M. A.; Gedik, M.; Brown, A. In *Handbook of Computational Chemistry*; Leszczynski, J., Kaczmarek-Kedziera, A., Puzyn, T., G. Papadopoulos, M., Reis, H., K. Shukla, M., Eds.; Springer International Publishing: Cham, 2017; pp 1875–1893.

[119] Laurent, A. D.; Jacquemin, D. Td-dft benchmarks: a review. *Int. J. Quantum Chem.* **2013**, *113*, 2019–2039.

[120] Dreuw, A.; Head-Gordon, M. Single-reference ab initio methods for the calculation of excited states of large molecules. *Chem. Rev.* **2005**, *105*, 4009–4037.

[121] List, N. H.; Jensen, H. J. A.; Kongsted, J. Local electric fields and molecular properties in heterogeneous environments through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 10070–10080.

[122] A., N. T.; Della Sala, F. Range separated functionals in the density functional based tight-binding method: formalism. *Phys. Status Solidi B.* **2011**, *249*, 237–244.

[123] Beerepoot, M. T. P.; Alam, M. M.; Bednarska, J.; Bartkowiak, W.; Ruud, K.; Zaleśny, R. Benchmarking the performance of exchange-correlation functionals for predicting two-photon absorption strengths. *J. Chem. Theory Comput.* **2018**, *14*, 3677–3685.

[124] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*. **1998**, *58*, 7260–7268.

[125] Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. Extension of the self-consistent-charge density-functional tight-binding method: third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction. *J. Phys. Chem. A*. **2007**, *111*, 10861–10873.

[126] Gaus, M.; Cui, Q.; Elstner, M. Dftb3: extension of the self-consistent-charge density-functional tight-binding method (scc-dftb). *J. Chem. Theory Comput.* **2011**, *7*, 931–948.

[127] Gaus, M.; Cui, Q.; Elstner, M. Density functional tight binding: application to organic and biological molecules. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 49–61.

[128] Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. Energetics and structure of glycine and alanine based model peptides: Approximate scc-dftb, am1 and pm3 methods in comparison with dft, hf and mp2 calculations. *Chem. Phys.* **2001**, *263*, 203–219.

[129] Krüger, T.; Elstner, M.; Schiffels, P.; Frauenheim, T. Validation of the density-functional based tight-binding approximation method for the calculation of reaction energies and other data. *J. Chem. Phys.* **2005**, *122*, 114110.

[130] Sattlemeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. L. Comparison of scc-dftb and nddo-based semiempirical molecular orbital methods for organic molecules. *J. Phys. Chem. A*. **2006**, *110*, 13551–13559.

[131] Zheng, G.; Irle, S.; Morokuma, K. Performance of the dftb method in comparison to dft and semiempirical methods for geometries and energies of c20-c86 fullerene isomers. *Chem. Phys. Lett.* **2005**, *412*, 210–216.

[132] Otte, N.; Scholten, M.; Thiel, W. Looking at self-consistent-charge density functional tight binding from a semiempirical perspective. *J. Phys. Chem. A.* **2007**, *111*, 5751–5755.

[133] Niehaus, T. A.; Suhai, S.; Della Sala, F.; Lugli, P.; Elstner, M.; Seifert, G.; Frauenheim, T. Tight-binding approach to time-dependent density-functional response theory. *Phys. Rev. B.* **2001**, *63*, 085108.

[134] Niehaus, T. A. Approximate time-dependent density functional theory. *J. Mol. Struct.* **2009**, *914*, 38–49.

[135] Fabian, J.; Diaz, L. A.; Seifert, G.; Niehaus, T. Calculation of excitation energies of organic chromophores: a critical evaluation. *J. Mol. Struct.* **2002**, *594*, 41–53.

[136] Trani, F.; Scalmani, G.; Zheng, G.; Carnimeo, I.; Frisch, M. J.; Barone, V. Time-dependent density functional tight binding: new formulation and benchmark of excited states. *J. Chem. Theory Comput.* **2011**, *7*, 3304–3313.

[137] Nishimoto, Y. Time-dependent density-functional tight-binding method with the third-order expansion of electron density. *J. Chem. Phys.* **2015**, *143*, 094108.

[138] Barone, V.; Carnimeo, I.; Scalmani, G. Computational spectroscopy of large systems in solution: the dftb/pcm and td-dftb/pcm approach. *J. Chem. Theory Comput.* **2013**, *9*, 2052–2071.

[139] Oviedo, M. B.; Sánchez, C. G. Transition dipole moments of the qy band in photosynthetic pigments. *J. Phys. Chem. A.* **2011**, *115*, 12280–12285.

[140] Liu, C. C.; Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **2010**, *79*, 413–444.

[141] Niehaus, T. A.; Elstner, M.; Frauenheim, T.; Suhai, S. Application of an approximate density-functional method to sulfur containing compounds. *J. Mol. Struct.* **2001**, *541*, 185–194.

[142] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

[143] Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

[144] `https://www.dftb.org/parameters/download/`.

[145] Monson, P. R.; Mcclain, W. M. Polarization dependence of the two-photon absorption of tumbling molecules with application to liquid 1-chloronaphthalene and benzene. *J. Chem. Phys.* **1970**, *53*, 29–37.

[146] Andrews, D. L.; Thirunamachandran, T. On three-dimensional rotational averages. *J. Chem. Phys.* **1977**, *67*, 5026–5033.

[147] Alam, M. M.; Chattopadhyaya, M.; Chakrabarti, S.; Ruud, K. Chemical control of channel interference in two-photon absorption processes. *Acc. Chem. Res.* **2014**, *47*, 1604–1612.

[148] Alam, M. M.; Chattopadhyaya, M.; Chakrabarti, S. Solvent induced channel interference in the two-photon absorption process-a theoretical study with a generalized few-state-model in three dimensions. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1156–1165.

[149] Mayer Göppert, M. Uber elementarakte mit zwei quantensprungen. *Ann. Phys.* **1931**, *401*, 273–295.

[150] Guo, J.-D.; Wang, C.-K.; Luo, Y.; Ågren, H. Influence of electron-acceptor strength on the resonant two-photon absorption cross sections of diphenylaminofluorene-based chromophores. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3869–3873.

[151] Sun, M.; Chen, J.; Xu, H. Visualizations of transition dipoles, charge transfer, and electron-hole coherence on electronic state transitions between excited states for two-photon absorption. *J. Chem. Phys.* **2008**, *128*, 064106.

[152] Silva, D. L.; Murugan, N. A.; Kongsted, J.; Rinkevicius, Z.; Canuto, S.; Ågren, H. The role of molecular conformation and polarizable embedding for one-and two-photon absorption of disperse orange 3 in solution. *J. Phys. Chem. B.* **2012**, *116*, 8169–8181.

[153] Matczyszyn, K.; Olesiak-Banska, J.; Nakatani, K.; Yu, P.; Murugan, N. A.; Zalesny, R.; Roztoczynska, A.; Bednarska, J.; Bartkowiak, W.; Kongsted, J.; Ågren, H.; Samoc, M. One-and two-photon absorption of a spiropyran-merocyanine system: experimental and theoretical studies. *J. Phys. Chem. B.* **2015**, *119*, 1515–1522.

[154] Cronstrand, P.; Luo, Y.; Ågren, H. In *Response Theory and Molecular Properties (A Tribute to Jan Linderberg and Poul Jørgensen)*; Jensen, H., Ed.; Adv. Quantum Chem.; Academic Press, 2005; Vol. 50; pp 1–21.

[155] Becke, A. D. Density-functional thermochemistry. Iii. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[156] Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-consistent molecular-orbital methods. Ix. An extended gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.

[157] Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-consistent molecular orbital methods. Xii. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

[158] Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta.* **1973**, *28*, 213–222.

[159] Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; De-Frees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. Xxiii. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665.

[160] Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. Iii. The 3-21+ g basis set for first-row elements, li–f. *J. Comput. Chem.* **1983**, *4*, 294–301.

[161] Olsen, J.; Jorgensen, P. Linear and nonlinear response functions for an exact state and for an mcscf state. *J. Chem. Phys.* **1998**, *82*, 3235–3264.

[162] Tretiak, S.; Chernyak, V. Resonant nonlinear polarizabilities in the time-dependent density functional theory. *J. Chem. Phys.* **2003**, *119*, 8809–8823.

[163] Sałek, P.; Vahtras, O.; Guo, J.; Luo, Y.; Helgaker, T.; Ågren, H. Calculations of two-photon absorption cross sections by means of density-functional theory. *Chem. Phys. Lett.* **2003**, *374*, 446–452.

[164] Frediani, L.; Rinkevicius, Z.; Ågren, H. Two-photon absorption in solution by means of time-dependent density-functional theory and the polarizable continuum model. *J. Chem. Phys.* **2005**, *122*, 244104.

[165] Cramer, C. J. *West Sussex, England: J. Wiley*, 2nd ed.; John Wiley & Sons, 2004; pp 322–323.

[166] Kalinowski, J. A.; Lesyng, B.; Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. Class iv charge model for the self-consistent charge density-functional tight-binding method. *J. Phys. Chem. A.* **2004**, *108*, 2545–2549.

[167] Ruger, R.; van Lenthe, E.; Heine, T.; Visscher, L. Tight-binding approximations to time-dependent density functional theory a fast approach for the calculation of electronically excited states. *J. Chem. Phys.* **2016**, *144*, 184103.

[168] Lundberg, M.; Nishimoto, Y.; Irle, S. Delocalization errors in a hubbard-like model: Consequences for density-functional tight-binding calculations of molecular systems. *Int. J. Quantum Chem.* **2012**, *112*, 1701.

[169] Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. A long-range-corrected time-dependent density functional theory. *J. Chem. Phys.* **2004**, *120*, 8425.

[170] Humeniuk, A.; Mitrić, R. Long-range correction for tight-binding td-dft. *J. Chem. Phys.* **2015**, *143*, 134120.

[171] Bae, J. H.; Rubini, M.; Jung, G.; Wiegand, G.; Seifert, M. H.; Azim, M.; Kim, J.-S.; Zumbusch, A.; Holak, T. A.; Moroder, L.; Huber, R.; Budisa, N. Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. *J. Mol. Biol.* **2003**, *328*, 1071–1081.

[172] Zhang, S.; Ai, H.-w. A general strategy to red-shift green fluorescent protein-based biosensors. *Nat. Chem. Biol.* **2020**, *16*, 1434–1439.

[173] Rüger, R.; Niehaus, T.; van Lenthe, E.; Heine, T.; Visscher, L. Vibrationally resolved uv/vis spectroscopy with time-dependent density functional based tight binding. *J. Chem. Phys.* **2016**, *145*, 184102.

[174] Cuny, J.; Tarrat, N.; Spiegelman, F.; Huguenot, A.; Rapacioli, M. Density-func-

131

tional tight-binding approach for metal clusters, nanoparticles, surfaces and bulk: application to silver and gold. *J. Phys. Condens. Matter.* **2018**, *30*, 303001.

[175] Stewart, J. J. P. Optimization of parameters for semiempirical methods v: modification of nddo approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.

[176] Hourahine, B. et al. Dftb+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.* **2020**, *152*, 124101.

[177] McClain, W. M. Excited state symmetry assignment through polarized two-photon absorption studies of fluids. *J. Chem. Phys.* **1971**, *55*, 2789–2796.

[178] Titov, E.; Humeniuk, A.; Mitrić, R. Exciton localization in excited-state dynamics of a tetracene trimer: a surface hopping lc-tddftb study. *Phys. Chem. Chem. Phys.* **2018**, *20*, 25995–26007.

[179] Darghouth, A. A. M. H. M.; Correa, G. C.; Juillard, S.; Casida, M. E.; Humeniuk, A.; Mitrić, R. Davydov-type excitonic effects on the absorption spectra of parallel-stacked and herringbone aggregates of pentacene: time-dependent density-functional theory and time-dependent density-functional tight binding. *J. Chem. Phys.* **2018**, *149*, 134111.

[180] Tran, T.; Prlj, A.; Lin, K.-H.; Hollas, D.; Corminboeuf, C. Mechanisms of fluorescence quenching in prototypical aggregation-induced emission systems: excited state dynamics with td-dftb. *Phys. Chem. Chem. Phys.* **2019**, *21*, 9026–9035.

[181] Schreck, M. H.; Röhr, M. I. S.; Clark, T.; Stepanenko, V.; Würthner, F.; Lambert, C. A self-assembled unit comprising 12 squaraine dyes built up from two star-shaped hexasquarainyl-benzene molecules. *Chem. Eur. J.* **2019**, *25*, 2831–2839.

[182] Darghouth, A. A. M. H. M.; Casida, M. E.; Zhu, X.; Natarajan, B.; Su, H.; Humeniuk, A.; Titov, E.; Miao, X.; Mitric, R. Effect of varying the td-lc-dftb range-separation parameter on charge and energy transfer in a model pentacene/buckminsterfullerene heterojunction. 2018.

[183] Nishimoto, Y. Time-dependent long-range-corrected density-functional tight-binding method combined with the polarizable continuum model. *J. Phys. Chem. A.* **2019**, *123*, 5649–5659.

[184] Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange-correlation functional using the coulomb-attenuating method (cam-b3lyp). *Chem. Phys. Lett.* **2004**, *393*, 51–57.

[185] Fihey, A.; Jacquemin, D. Performances of density functional tight-binding methods for describing ground and excited state geometries of organic molecules. *J. Chem. Theory Comput.* **2019**, *15*, 6267–6276.

[186] Kerr, J. N. D.; Denk, W. Imaging in vivo: watching the brain in action. *Nat. Rev. Neurosci.* **2008**, *9*, 195–205.

[187] Horton, N. G.; Wang, K.; Kobat, D.; Clark, C. G.; Wise, F. W.; Schaffer, C. B.; Xu, C. In vivo three-photon microscopy of subcortical structures within an intact mouse brain. *Nat. Photonics.* **2013**, *7*, 205–209.

[188] Maity, B.; Maiti, S. Label-free imaging of neurotransmitters in live brain tissue by multi-photon ultraviolet microscopy. *Neuronal Signaling.* **2018**, *2*.

[189] Weissleder, R.; Ntziachristos, V. Shedding light onto live molecular targets. *Nat. Med.* **2003**, *9*, 123–128.

[190] Shi, L.; Sordillo, L. A.; Rodríguez-Contreras, A.; Alfano, R. Transmission in

near-infrared optical windows for deep brain imaging. *J. Biophotonics.* **2016**, *9*, 38–43.

[191] Cheng, L.-C.; Horton, N. G.; Wang, K.; Chen, S.-J.; Xu, C. Measurements of multiphoton action cross sections for multiphoton microscopy. *Biomed. Opt. Express.* **2014**, *5*, 3427–3433.

[192] Kobat, D.; Horton, N. G.; Xu, C. In vivo two-photon microscopy to 1.6 mm depth in mouse cortex. *J. Biomed. Opt.* **2011**, *16*, 1–5.

[193] Wang, T.; Xu, C. Three-photon neuronal imaging in deep mouse brain. *Optica.* **2020**, *7*, 947–960.

[194] Liu, H.; Wang, J.; Peng, X.; Zhuang, Z.; Qiu, P.; Wang, K. Ex and in vivo characterization of the wavelength-dependent 3-photon action cross-sections of red fluorescent proteins covering the 1700-nm window. *J. Biophotonics.* **2018**, *11*, e201700351.

[195] Deng, X.; Zhuang, Z.; Liu, H.; Qiu, P.; Wang, K. Measurement of 3-photon excitation and emission spectra and verification of kasha's rule for selected fluorescent proteins excited at the 1700 nm window. *Opt. Express.* **2019**, *27*, 12723–12731.

[196] Xu, C.; Zipfel, W.; Shear, J. B.; Williams, R. M.; Webb, W. W. Multiphoton fluorescence excitation: new spectral windows for biological nonlinear microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10763–10768.

[197] Drobizhev, M.; Tillo, S.; Makarov, N. S.; Hughes, T. E.; Rebane, A. Absolute two-photon absorption spectra and two-photon brightness of orange and red fluorescent proteins. *J. Phys. Chem. B.* **2009**, *113*, 855–859.

[198] Molina, R. S.; Tran, T. M.; Campbell, R. E.; Lambert, G. G.; Salih, A.; Shaner, N. C.; Hughes, T. E.; Drobizhev, M. Blue-shifted green fluorescent

protein homologues are brighter than enhanced green fluorescent protein under two-photon excitation. *J. Phys. Chem. Lett.* **2017**, *8*, 2548–2554.

[199] Friese, D. H.; Beerepoot, M. T. P.; Ringholm, M.; Ruud, K. Open-ended recursive approach for the calculation of multiphoton absorption matrix elements. *J. Chem. Theory Comput.* **2015**, *11*, 1129–1144.

[200] Cronstrand, P.; Jansik, B.; Jonsson, D.; Luo, Y.; Ågren, H. Density functional response theory calculations of three-photon absorption. *J. Chem. Phys.* **2004**, *121*, 9239–9246.

[201] Xu, C.; Webb, W. W. In *Topics in Fluorescence Spectroscopy: Volume 5: Non-linear and Two-Photon-Induced Fluorescence*; Lakowicz, J. R., Ed.; Springer US: Boston, MA, 2002; pp 471–540.

[202] He, G. S.; Bhawalkar, J. D.; Prasad, P. N.; Reinhardt, B. A. Three-photon-absorption-induced fluorescence and optical limiting effects in an organic compound. *Opt. Lett.* **1995**, *20*, 1524–1526.

[203] Nobis, D.; Fisher, R. S.; Simmermacher, M.; Hopkins, P. A.; Tor, Y.; Jones, A. C.; Magennis, S. W. Single-molecule detection of a fluorescent nucleobase analogue via multiphoton excitation. *J. Phys. Chem. Lett.* **2019**, *10*, 5008–5012.

[204] Makarov, N. S.; Drobizhev, M.; Rebane, A. Two-photon absorption standards in the 550–1600 nm excitation wavelength range. *Opt. Express.* **2008**, *16*, 4029–4047.

[205] Albota, M. A.; Xu, C.; Webb, W. W. Two-photon fluorescence excitation cross sections of biomolecular probes from 690 to 960 nm. *Appl. Opt.* **1998**, *37*, 7352–7356.

[206] Xu, C.; Webb, W. W. Measurement of two-photon excitation cross sections of

molecular fluorophores with data from 690 to 1050 nm. *J. Opt. Soc. Am. B.* **1996**, *13*, 481–491.

[207] Juma, J. M.; Vuai, S. A. H.; Babu, N. S. Td-dft investigations on optoelectronic properties of fluorescein dye derivatives in dye-sensitized solar cells (dsscs). *Int. J. Photoenergy.* **2019**, *2019*, 4616198.

[208] Zhang, Z.; Zeng, S.; Liu, Y.; Zhou, W.; Chen, T.; Luo, Q. 5-ht spatial distribution imaging with multiphoton excitation of 5-ht correlative visible fluorescence in live cells. International Workshop on Photonics and Imaging in Biology and Medicine. 2002; pp 299–305.

[209] Szeitz, A.; Bandiera, S. M. Analysis and measurement of serotonin. *Biomed. Chromatogr.* **2018**, *32*, e4135.

[210] Jha, P. C.; Wang, Y.; Luo, Y.; Agren, H. A critical examination of two-photon absorption cross-sections of some reference dyes. 2006 International Symposium on Biophotonics, Nanophotonics and Metamaterials. 2006; pp 60–63.

[211] Batistela, V. R.; da Costa Cedran, J.; de Oliveira, H. P. M.; Scarminio, I. S.; Ueno, L. T.; da Hora Machado, A. E.; Hioka, N. Protolytic fluorescein species evaluated using chemometry and dft studies. *Dyes Pigm.* **2010**, *86*, 15–24.

[212] Nag, A.; Goswami, D. Solvent effect on two-photon absorption and fluorescence of rhodamine dyes. *J. Photochem. Photobiol. A.* **2009**, *206*, 188–197.

[213] Chattopadhyay, A.; Rukmini, R.; Mukherjee, S. Photophysics of a neurotransmitter: ionization and spectroscopic properties of serotonin. *Biophys. J.* **1996**, *71*, 1952–1960.

[214] Sjöback, R.; Nygren, J.; Kubista, M. Absorption and fluorescence properties of fluorescein. *Spectrochim. Acta A.* **1995**, *51*, L7–L21.

[215] Tamulis, A.; Tamuliene, J.; Balevicius, M.; Rinkevicius, Z.; Tamulis, V. Quantum mechanical studies of intensity in electronic spectra of fluorescein dianion and monoanion forms. *Struct. Chem.* **2003**, *14*, 643–648.

[216] Zehentbauer, F. M.; Moretto, C.; Stephen, R.; Thevar, T.; Gilchrist, J. R.; Pokrajac, D.; Richard, K. L.; Kiefer, J. Fluorescence spectroscopy of rhodamine 6g: concentration and solvent effects. *Spectrochim. Acta A.* **2014**, *121*, 147–151.

[217] Rudnick, G.; Clark, J. From synapse to vesicle: the reuptake and storage of biogenic amine neurotransmitters. *Biochim. Biophys. Acta - Bioenergetics.* **1993**, *1144*, 249–263.

[218] Pratuangdejkul, J.; Nosoongnoen, W.; Guérin, G.-A.; Loric, S.; Conti, M.; Launay, J.-M.; Manivet, P. Conformational dependence of serotonin theoretical pKa prediction. *Chem. Phys. Lett.* **2006**, *420*, 538–544.

[219] Knoth, J.; Isaacs, J. M.; Njus, D. Amine transport in chromaffin granule ghosts. pH dependence implies cationic form is translocated. *J. Biol. Chem.* **1981**, *256*, 6541–6543.

[220] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple [phys. rev. lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.

[221] Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. Xxiii. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665.

[222] Gerasimova, M. A.; Tomilin, F. N.; Malyar, E. Y.; Varganov, S. A.; Fedorov, D. G.; Ovchinnikov, S. G.; Slyusareva, E. A. Fluorescence and photoinduced proton transfer in the protolytic forms of fluorescein: experimental and computational study. *Dyes Pigm.* **2020**, *173*, 107851.

[223] Frisch, M. J. et al. Gaussian 16 revision c.01. 2016; Gaussian Inc. Wallingford CT.

[224] Dalton, a molecular electronic structure program, release v2018.2. 2019; See http://www.daltonprogram.org/.

[225] Papajak, E.; Zheng, J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on basis sets beautiful: seasonal plantings of diffuse basis functions. *J. Chem. Theory Comput.* **2011**, *7*, 3027–3034.

[226] Pritchard, D. D. B. G. T. D., Benjamin P. Altarawy; L., W. T. A new basis set exchange: an open, up-to-date resource for the molecular sciences community. *J. Chem. Inf. Model* **2019**, *59*, 4814–4820.

[227] Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

[228] Jensen, F. Unifying general and segmented contracted basis sets. segmented polarization consistent basis sets. *J. Chem. Theory Comput.* **2014**, *10*, 1074–1085.

[229] Jensen, F. Polarization consistent basis sets: principles. *J. Chem. Phys.* **2001**, *115*, 9113–9125.

[230] Jensen, F. How large is the elephant in the density functional theory room? *J. Phys. Chem. A.* **2017**, *121*, 6104–6107.

[231] Jensen, F. Polarization consistent basis sets. Iii. The importance of diffuse functions. *J. Chem. Phys.* **2002**, *117*, 9234–9240.

[232] Cronstrand, P.; Norman, P.; Luo, Y.; Ågren, H. Few-states models for three-photon absorption. *J. Chem. Phys.* **2004**, *121*, 2020–2029.

[233] Zhou, P.; Liu, J.; Yang, S.; Chen, J.; Han, K.; He, G. The invalidity of the photo-induced electron transfer mechanism for fluorescein derivatives. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15191–15198.

[234] Brackmann, U. *Lambdachrome laser dyes*, 3rd ed.; Lambda Physik AG, Goettin-gen, 2000.

[235] Rebane, A.; Mikhaylov, A. Improved reference standards for femtosecond three-photon excitation of fluorescence in the wavelength range 950–1750 nm. Multiphoton Microscopy in the Biomedical Sciences XVIII. 2018; pp 300–311.

[236] Lobayan, R. M.; Schmit, M. C. P. Conformational and nbo studies of serotonin as a radical scavenger. changes induced by the oh group. *J. Mol. Graph.* **2018**, *80*, 224–237.

[237] LeGreve, T. A.; Baquero, E. E.; Zwier, T. S. Infrared and ultraviolet spectral signatures and conformational preferences of jet-cooled serotonin. *J. Am. Chem. Soc.* **2007**, *129*, 4028–4038.

[238] van Mourik, T.; Emson, L. E. V. A theoretical study of the conformational land-scape of serotonin. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5863–5871.

[239] Pisterzi, L. F.; Almeida, D. R.; Chass, G. A.; Torday, L. L.; Papp, J. G.; Varro, A.; Csizmadia, I. G. Density functional molecular computations on protonated sero-tonin in the gas phase and various solvent media. *Chem. Phys. Lett.* **2002**, *365*, 542–551.

[240] Pratuangdejkul, J.; Jaudon, P.; Ducrocq, C.; Nosoongnoen, W.; Guerin, G.-A.; Conti, M.; Loric, S.; Launay, J.-M.; Manivet, P. Cation-π interactions in serotonin: conformational, electronic distribution, and energy decomposition analysis. *J. Chem. Theory Comput.* **2006**, *2*, 746–760.

[241] Bayarı, S.; Saglam, S.; Ustundag, H. F. Experimental and theoretical studies of the vibrational spectrum of 5-hydroxytryptamine. *J. Mol. Struct.* **2005**, *726*, 225–232.

[242] Omidyan, R.; Amanollahi, Z.; Azimi, G. Protonated serotonin: geometry, electronic structures and photophysical properties. *Spectrochim. Acta A.* **2017**, *182*, 8–16.

[243] Dian, B. C.; Longarte, A.; Zwier, T. S. Hydride stretch infrared spectra in the excited electronic states of indole and its derivatives: Direct evidence for the $1\pi\sigma^*$ state. *J. Chem. Phys.* **2003**, *118*, 2696–2706.

[244] Sobolewski, A. L.; Domcke, W. Ab initio investigations on the photophysics of indole. *Chem. Phys. Lett.* **1999**, *315*, 293–298.

[245] Arulmozhiraja, S.; Coote, M. L. 1la and 1lb states of indole and azaindole: is density functional theory inadequate? *J. Chem. Theory Comput.* **2012**, *8*, 575–584.

[246] Zhang, M.; Guo, Y.; Feng, X.; Yu, X.; Jin, X.; Qiu, L.; Zhao, G. Theoretical modeling of the hydrated serotonin in solution: insight into intermolecular hydrogen bonding dynamics and spectral shift in the electronic excited states. *J. Mol. Liq.* **2019**, *288*, 111093.

[247] Chen, R. F. Fluorescence of protonated excited-state forms of 5-hydroxytryptamine (serotonin) and related indoles. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *60*, 598–605.

[248] Maiti, S.; Shear, J. B.; Williams, R. M.; Zipfel, W. R.; Webb, W. W. Measuring serotonin distribution in live cells with three-photon excitation. *Science.* **1997**, *275*, 530–532.

[249] Shear, J. B.; Xu, C.; Webb, W. W. Multiphoton-excited visible emission by serotonin solutions. *Photochem. Photobiol.* **1997**, *65*, 931–936.

[250] Adhikesavalu, D. N.; Mastropaolo, D.; Camerman, A.; Camerman, N. Two rhodamine derivatives: 9-[2-(ethoxycarbonyl)phenyl]-3,6-bis(ethyl-amino)-2,7-di-methyl-xanthyl-ium chloride monohydrate and 3,6-di-amino-9-[2-(methoxy-carbonyl)-phenyl]xanthyl-ium chloride trihydrate. *Acta Crystallogr. C.* **2001**, *57*, 657–659.

[251] Watanabe, H.; Hayazawa, N.; Inouye, Y.; Kawata, S. Dft vibrational calculations of rhodamine 6g adsorbed on silver: analysis of tip-enhanced raman spectroscopy. *J. Phys. Chem. B.* **2005**, *109*, 5012–5020.

[252] Savarese, M.; Aliberti, A.; De Santo, I.; Battista, E.; Causa, F.; Netti, P. A.; Rega, N. Fluorescence lifetimes and quantum yields of rhodamine derivatives: new insights from theory and experiment. *J. Phys. Chem. A.* **2012**, *116*, 7491–7497.

[253] Zhou, P. Why the lowest electronic excitations of rhodamines are overestimated by time-dependent density functional theory. *Int. J. Quantum Chem.* **2018**, *118*, e25780.

[254] Kulesza, A. J.; Titov, E.; Daly, S.; Włodarczyk, R.; Megow, J.; Saalfrank, P.; Choi, C. M.; MacAleese, L.; Antoine, R.; Dugourd, P. Excited states of xan-thene analogues: photofragmentation and calculations by cc2 and time-dependent density functional theory. *ChemPhysChem.* **2016**, *17*, 3129–3138.

[255] Arden, J.; Deltau, G.; Huth, V.; Kringel, U.; Peros, D.; Drexhage, K. H. Fluores-cence and lasing properties of rhodamine dyes. *J. Lumin.* **1991**, *48-49*, 352–358.

[256] Xing, G.; Chakrabortty, S.; Ngiam, S. W.; Chan, Y.; Sum, T. C. Three-photon absorption in seeded cdse/cds nanorod heterostructures. *J. Phys. Chem. C.* **2011**, *115*, 17711–17716.

[257] Taniguchi, M.; Du, H.; Lindsey, J. S. Photochemcad 3: diverse modules for pho-

tophysical calculations with multiple spectral databases. *Photochem. Photobiol.* **2018**, *94*, 277–289.

[258] Lossau, H.; Kummer, A.; Heinecke, R.; Pöllinger-Dammer, F.; Kompa, C.; Bieser, G.; Jonsson, T.; Silva, C.; Yang, M.; Youvan, D.; Michel-Beyerle, M. Time-resolved spectroscopy of wild-type and mutant green fluorescent proteins reveals excited state deprotonation consistent with fluorophore-protein interactions. *Chem. Phys.* **1996**, *213*, 1 –16.

[259] Di Remigio, R.; Giovannini, T.; Ambrosetti, M.; Cappelli, C.; Frediani, L. Fully polarizable qm/fluctuating charge approach to two-photon absorption of aqueous solutions. *J. Chem. Theory Comput.* **2019**, *15*, 4056–4068.

[260] Fang, K.; Lieblich, S.; Tirrell, D. Incorporation of non-canonical amino acids into proteins by global reassignment of sense codons. *Methods Mol. Biol.* **2018**, *1798*, 173–186.

[261] Nediljko, B.; Prajna, P. P. Designing novel spectral classes of proteins with a tryptophan-expanded genetic code. *Biological Chemistry* **2004**, *385*, 893–904.

[262] Zhao, Y.; Campbell, R. E. In *New Techniques in Systems Neuroscience*; Douglass, A. D., Ed.; Springer International Publishing: Cham, 2015; pp 57–96.

[263] Voityuk, A. A.; Michel-Beyerle, M.-E.; Rösch, N. Quantum chemical modeling of structure and absorption spectra of the chromophore in green fluorescent proteins. *Chem. Phys.* **1998**, *231*, 13–25.

[264] List, N. H.; Olsen, J. M. H.; Rocha-Rinza, T.; Christiansen, O.; Kongsted, J. Performance of popular xc-functionals for the description of excitation energies in gfp-like chromophore models. *Int. J. Quantum Chem.* **2012**, *112*, 789–800.

[265] Sanchez-Garcia, E.; Doerr, M.; Hsiao, Y.-W.; Thiel, W. Qm/mm study of the monomeric red fluorescent protein dsred.m1. *J. Phys. Chem. B.* **2009**, *113*, 16622–16631.

[266] Compute Canada. `https://www.computecanada.ca/`, (accessed 2021-07-24).

[267] Merzlyak, E. M.; Goedhart, J.; Shcherbo, D.; Bulina, M. E.; Shcheglov, A. S.; Fradkov, A. F.; Gaintzeva, A.; Lukyanov, K. A.; Lukyanov, S.; Gadella, T. W. J.; Chudakov, D. M. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat. Methods.* **2007**, *4*, 555–557.

[268] Case, D. et al. Amber 2018. 2018; University of California, San Francisco.

[269] Dapprich, S.; Komáromi, I.; Byun, K.; Morokuma, K.; Frisch, M. J. A new oniom implementation in gaussian98. part i. the calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Mol. Struct.: THEOCHEM.* **1999**, *461-462*, 1–21.

[270] Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. Combining quantum mechanics methods with molecular mechanics methods in oniom. *J. Chem. Theory Comput.* **2006**, *2*, 815–826.

[271] Clemente, F. R.; Vreven, T.; Frisch, M. J. *Quantum Biochemistry*; John Wiley & Sons, Ltd, 2010; Chapter 2, pp 61–83.

[272] Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y. R.e.d. server: a web service for deriving resp and esp charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **2011**, *39*, W511–W517.

[273] Wang, F.; Becker, J. P.; Cieplak, P.; Dupradeau, F.-Y. R.e.d. python: object oriented programming for amber force fields. 2013.

[274] Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.e.d. tools: advances in resp and esp charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821–7839.

[275] Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

[276] Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

[277] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[278] The dalton quantum chemistry program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 269–284.

[279] Olsen, J. M. H.; List, N. H.; Steinmann, C.; Steindal, A. H.; Nørby, M. S.; Reinholdt, P. Pelib: the polarizable embedding library. 2018.

[280] Olsen, J. M. H. Pyframe: python framework for fragment-based multi-scale embedding (version 0.2.0). 2018; Available at https://gitlab.com/FraME-projects/PyFraME.

[281] Vahtras, O. Loprop for dalton. 2014.

[282] Gagliardi, L.; Lindh, R.; Karlström, G. Local properties of quantum chemical systems: the loprop approach. *J. Chem. Phys.* **2004**, *121*, 4494–4500.

[283] Pakhomov, A. A.; Martynov, V. I. Gfp family: structural insights into spectral tuning. *Chem. Biol.* **2008**, *15*, 755–764.

[284] Shu, X.; Shaner, N. C.; Yarbrough, C. A.; Tsien, R. Y.; Remington, S. J. Novel chromophores and buried charges control color in mfruits. *Biochemistry.* **2006**, *45*, 9639–9647.

[285] Subach, F. V.; Verkhusha, V. V. Chromophore transformations in red fluorescent proteins. *Chem. Rev.* **2012**, *112*, 4308–4327.

[286] Zhang, M.-Y.; Wang, J.-Y.; Lin, C. S.; Cheng, W.-D. First-principles study of one-and two-photon absorption of the h-bonding complexes from monomeric red fluorescent proteins with large stokes shifts. *J. Phys. Chem. B.* **2011**, *115*, 10750–10757.

[287] Konold, P.; Regmi, C. K.; Chapagain, P. P.; Gerstman, B. S.; Jimenez, R. Hydrogen bond flexibility correlates with stokes shift in mplum variants. *J. Phys. Chem. B.* **2014**, *118*, 2940–2948.

[288] Konold, P. E.; Yoon, E.; Lee, J.; Allen, S. L.; Chapagain, P. P.; Gerstman, B. S.; Regmi, C. K.; Piatkevich, K. D.; Verkhusha, V. V.; Joo, T.; Jimenez, R. Fluorescence from multiple chromophore hydrogen-bonding states in the far-red protein tagrfp675. *J. Phys. Chem. Lett.* **2016**, *7*, 3046–3051.

[289] Pletnev, V. Z.; Pletneva, N. V.; Sarkisyan, K. S.; Mishin, A. S.; Lukyanov, K. A.; Goryacheva, E. A.; Ziganshin, R. H.; Dauter, Z.; Pletnev, S. Structure of the green fluorescent protein nowgfp with an anionic tryptophan-based chromophore. *Acta Crystallogr. D.* **2015**, *71*, 1699–1707.

[290] Drobizhev, M.; Molina, R. S.; Callis, P. R.; Scott, J. N.; Lambert, G. G.; Salih, A.; Shaner, N. C.; Hughes, T. E. Local electric field controls fluorescence quantum yield of red and far-red fluorescent proteins. *Front. Mol. Biosci.* **2021**, *8*, 7.

[291] Shu, X.; Wang, L.; Colip, L.; Kallio, K.; Remington, S. Rcsb pdb–2qlg: mplum. `https://www.rcsb.org/structure/2QLG`, (accessed 2021-07-17).

[292] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

[293] Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. Pdb2pqr: an automated pipeline for the setup of poisson-boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667.

[294] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. Ucsf chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

[295] Dalton, a molecular electronic structure program, release vdalton2020.0. 2020; See http://www.daltonprogram.org/.

[296] Shapovalov, M. V.; Dunbrack Jr., R. L. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure.* **2011**, *19*, 844–858.

[297] Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph.* **2006**, *25*, 247–260.

[298] Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics.* **2012**, *4*, 17.

[299] Wieduwilt, E. K.; Boisson, J.-C.; Terraneo, G.; Hénon, E.; Genoni, A. A step toward the quantification of noncovalent interactions in large biological systems: the independent gradient model-extremely localized molecular orbital approach. *J. Chem. Inf. Model.* **2021**, *61*, 795–809.

[300] Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W. Revealing noncovalent interactions. *J. Am. Chem. Soc.* **2010**, *132*, 6498–6506.

[301] Schaftenaar, G.; Vlieg, E.; Vriend, G. Molden 2.0: quantum chemistry meets proteins. *J. Comput. Aided Mol. Des.* **2017**, *31*, 789–800.

[302] Schaftenaar, G.; Noordik, J. H. Molden: a pre-and post-processing program for molecular and electronic structures*. *J. Comput. Aided Mol. Des.* **2000**, *14*, 123–134.

[303] Xu, C.; Zipfel, W. R. Multiphoton excitation of fluorescent probes. *Cold Spring Harb. Protoc.* **2015**, *2015*, 473–475.

[304] Xu, C.; Williams, R. M.; Zipfel, W.; Webb, W. W. Multiphoton excitation cross-sections of molecular fluorophores. *Bioimaging.* **1996**, *4*, 198–207.

[305] Lesniak, W. G.; Jyoti, A.; Mishra, M. K.; Louissaint, N.; Romero, R.; Chugani, D. C.; Kannan, S.; Kannan, R. M. Concurrent quantification of tryptophan and its major metabolites. *Anal. Biochem.* **2013**, *443*, 222–231.

[306] Balaji, J.; Desai, R.; Maiti, S. Live cell ultraviolet microscopy: a comparison between two-and three-photon excitation. *Microsc. Res. Tech.* **2004**, *63*, 67–71.

[307] Kaatz, P.; Shelton, D. P. Two-photon fluorescence cross-section measurements calibrated with hyper-rayleigh scattering. *J. Opt. Soc. Am. B.* **1999**, *16*, 998–1006.

[308] Tan, W.; Parpura, V.; Haydon, P. G.; Yeung, E. S. Neurotransmitter imaging in living cells based on native fluorescence detection. *Anal. Chem.* **1995**, *67*, 2575–2579.

[309] Bindhu, C. V.; Harilal, S. S.; Nampoori, V. P. N.; Vallabhan, C. P. G. Studies of

147

nonlinear absorption and aggregation in aqueous solutions of rhodamine 6g using a transient thermal lens technique. *J. Phys. D: Appl. Phys.* **1999**, *32*, 407–411.

[310] Bindhu, C. V.; Harilal, S. S.; Bindu, V.; Nampoori, V. P. N.; Vallabhan, C. P. G. Multiphoton absorption studies in aqueous solutions of rhodamine 6g laser dye using transient thermal lens technique. Selected Papers from International Conference on Optics and Optoelectronics '98. 1999; pp 312–317.

Appendix A

Determination of Two-photon Absorption Cross-sections in Fluorescent Protein Chromophores Using Tight-binding Time-dependent Density Functional Theory

Figure A.1. Selection of transitions for RFP-like models. The comparison of 2PA cross-sections between methods was preceded by an evaluation of the molecular orbitals (MOs) involved in the transitions taken into account by TD-B3LYP/6-31+G(d,p),[12,14] for which case correspond to the S_1 in all the cases. From the examples shown in the figure can be observed that in the case of TD-DFTB2 the transitions that match (based on the MOs) with TD-B3LYP are not necessarily S_1, and also, the orbitals involved not always correspond to HOMO→LUMO.

Figure A.2. For RFP-like models, comparison of MOs obtained at TD-DFTB2, TD-PBE/6-31+G(d,p), and TD-B3LYP/6-31+G(d,p). [14]

Table A.1. Oscillator Strengths and Excitation Energies for Selected Models Using the PBE Functional and the 6-31+G(d,p) Basis Set. In Between Brackets, the Percent of Error With Respect to the Correspondent Values Obtained at TD-B3LYP/6-31+G(d,p) by Salem et al. [12,14] Notice (as Described Later) That the Data we Used as a Reference for the GFP-Like Models Was Obtained Using the Polarizable Continuum Model (PCM) and not at Gas-Phase

Model	S_n	OS	Energy (eV)
RFP-like			
16a	1	0.25 (−53)	2.367 (−16)
18	1	0.54 (−26)	2.443 (−12)
20	1	0.22 (−44)	1.891 (−20)
22	3	0.35 (−41)	2.456 (−15)
GFP-like			
19	3	0.57 (74)	3.204 (−1)
20	1	0.14 (−51)	2.325 (−18)

Table A.2. Excitation Energies, Oscillator Strengths (OS), 2PA Transition Moments (δ^{2PA}), and 2PA Cross-Sections (σ^{2PA}) of the HOMO→LUMO Transitions for the RFP-like Models at the TD-DFTB2 and TD-B3LYP/6-31+G(d,p) Levels of Theory [14]

Model	TD-DFTB2				TD-B3LYP/6-31+G(d,p)			
	Energy (eV)	OS	δ^{2PA}	σ^{2PA} (GM)	Energy (eV)	OS	δ^{2PA}	σ^{2PA} (GM)
1a	2.874	0.45	1808	5.5	3.032	0.53	505	1.7
1b	2.893	0.44	1053	3.2	3.016	0.55	341	1.2
2a	2.851	0.45	1844	5.5	3.050	0.51	761	2.5
2b	2.893	0.43	766	2.3	3.059	0.51	401	1.4
6	2.839	0.50	1420	4.2	2.949	0.68	1240	3.9
7	2.773	0.57	765	2.2	2.890	0.76	711	2.2
8	2.816	0.54	1614	4.7	2.959	0.64	423	1.4
9	2.791	0.51	6782	19.4	2.962	0.66	3503	11.3
11	2.79	0.55	1141	3.3	2.931	0.66	580	1.8
12	2.632	0.56	22205	56.4	2.857	0.76	6187	18.5
13	2.629	0.51	17851	45.2	2.872	0.67	7489	22.6
14	2.635	0.54	20188	51.4	2.874	0.7	7697	23.3
15	2.784	0.57	741	2.1	2.915	0.68	1000	3.1
16a	2.395	0.22	25863	54.4	2.832	0.52	14293	42.0
16b	2.443	0.26	22503	49.2	2.901	0.62	7873	24.3
17	2.611	0.53	20724	51.8	2.854	0.7	8669	25.9
18	2.574	0.51	18469	44.9	2.777	0.73	11678	33.0
19	2.631	0.38	34392	87.3	2.841	0.69	22183	65.6
20	2.191	0.23	47890	84.3	2.366	0.39	48027	98.5
21	2.197	0.04	15599	27.6	2.718	0.29	38869	105.2
22	2.804	0.49	1973	5.7	2.881	0.6	6372	19.4

Table A.3. Vectors Corresponding to the Difference Between the Excited and the Ground State Permanent Dipole Moments ($\|\Delta\mu\| = \left(\sum \Delta\mu_\alpha^2\right)^{1/2}$), The Transition Dipole Moments ($\|\mu_{01}\|$), and the Alignment Between Them Given by $\cos^2\theta$ for RFP-Like Models at Both the TD-DFTB2 and TD-B3LYP/6-31+G(d,p) Levels of Theory.[14] All Values Given in Atomic Units (au)

Model	DFTB2			TD-B3LYP/6-31+G(d,p)		
	$\|\Delta\mu\|$ (au)	$\|\mu_{01}\|$ (au)	$\cos^2\theta$	$\|\Delta\mu\|$ (au)	$\|\mu_{01}\|$ (au)	$\cos^2\theta$
1a	1.33	2.53	0.33	0.88	2.67	0.03
1b	1.13	2.48	0.21	0.77	2.60	0.00
2a	1.36	2.54	0.30	0.99	2.73	0.10
2b	1.06	2.47	0.09	0.82	2.61	0.03
6	1.17	2.68	0.24	0.88	3.06	0.45
7	0.91	2.89	0.03	0.70	3.27	0.21
8	1.15	2.81	0.28	0.73	2.96	0.00
9	1.96	2.74	0.65	1.41	3.02	0.58
11	1.06	2.83	0.12	0.81	3.03	0.03
12	2.82	2.96	0.90	1.54	3.29	0.75
13	2.70	2.81	0.86	1.81	3.09	0.76
14	2.79	2.89	0.87	1.80	3.15	0.75
15	0.93	2.89	0.00	0.86	3.09	0.26
16a	4.27	1.92	0.90	2.66	2.72	0.88
16b	3.64	2.07	1.00	1.87	2.96	0.87
17	2.80	2.88	0.87	1.87	3.17	0.77
18	2.66	2.85	0.85	2.02	3.27	0.81
19	4.12	2.44	1.00	2.78	3.15	0.98
20	4.89	2.09	0.90	4.26	2.59	0.90
21	6.49	0.90	0.90	5.35	2.08	0.97
22	1.14	2.67	0.56	1.70	2.92	0.85

Table A.4. Ground and Excited State Permanent Dipole Moment Components (μ_α) for Each of the RFP-like Chromophores of Figure 2.2 at TD-DFTB2 and TD-B3LYP/6-31+G(d,p),[14] Both Obtained Within the 2LM. All Values Are Given in Atomic Units (au)

Model	TD-DFTB2						TD-B3LYP/6-31+G(d,p)					
	μ_{00}			μ_{11}			μ_{00}			μ_{11}		
	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x
1a	1.51	−0.76	0.79	0.46	0.07	0.78	1.68	−1.25	0.85	2.07	−2.04	0.86
1b	−0.15	−0.74	0.85	0.58	−1.62	0.85	0.34	−1.03	0.95	0.14	−0.29	0.96
2a	0.54	−1.36	0.82	−0.49	−0.46	0.81	0.14	−1.90	0.89	0.67	−2.74	0.90
2b	−0.80	0.04	0.74	−1.24	1.01	0.74	−1.54	−0.08	0.81	−1.33	−0.86	0.81
6	1.32	−1.12	1.31	0.81	−0.07	1.39	0.65	−1.15	1.24	1.21	−1.83	1.21
7	0.49	−0.67	0.81	0.34	0.23	0.80	−0.19	−0.98	0.86	0.13	−1.61	0.87
8	1.28	0.24	0.81	0.48	1.07	0.81	1.36	0.31	0.89	1.47	−0.41	0.90
9	−0.47	−0.63	0.79	−2.17	0.35	0.80	−1.13	−0.92	0.87	0.02	−1.74	0.86
11	1.07	−0.83	0.80	0.56	0.09	0.80	0.77	−1.15	0.88	0.75	−1.95	0.88
12	−0.24	−0.43	0.80	−3.01	0.03	0.84	−0.30	−0.44	0.88	1.13	−1.00	0.87
13	−0.83	−1.23	0.80	−3.44	−0.53	0.84	−1.16	−1.45	0.87	0.51	−2.15	0.85
14	−0.64	−1.00	0.79	−3.24	−0.01	0.81	−1.03	−1.16	0.86	0.53	−2.06	0.86
15	1.50	−0.45	1.63	1.54	0.48	1.61	1.26	−0.59	1.82	0.81	−1.32	1.84
16a	−0.77	−1.42	0.82	−5.04	−1.29	0.87	−0.77	−1.88	0.88	1.87	−2.22	0.86
16b	−1.42	−0.63	0.86	−5.04	−1.08	0.90	−1.93	−0.45	0.95	−0.12	−0.92	0.93
17	−0.95	−1.05	0.70	−3.56	−0.02	0.73	−1.37	−1.23	0.81	0.26	−2.15	0.79
18	−1.41	−1.12	0.68	−4.00	−0.53	0.70	−1.99	−1.38	0.65	−0.04	−1.92	0.63
19	−0.19	−0.49	0.81	−4.31	−0.55	0.84	−0.94	−0.78	0.90	1.83	−0.95	0.88
20	−0.61	−0.97	0.57	−5.46	−0.40	0.51	−1.09	−1.41	0.52	3.14	−1.87	0.53
21	−0.60	−0.52	0.85	−7.00	−1.59	0.91	−1.36	−0.16	0.99	3.94	0.62	0.94
22	2.11	0.55	0.80	2.81	1.45	0.79	1.60	0.29	0.88	0.17	−0.64	0.90

Table A.5. Percent Error (%) Associated With the Excitation Energies, Oscillator Strengths (OS), the Vector Corresponding to the Difference Between the Excited and the Ground State Permanent Dipole Moments ($\|\Delta\mu\| = \left(\sum \Delta\mu_\alpha^2\right)^{1/2}$), and the Transition Dipole Moments Obtained for the RFP-Like Models Using TD-DFTB2 With Respect to the Values Reported Previously Using TD-B3LYP/6-31+G(d,p)[14]

Model	Energy (eV)	OS	$\|\Delta\mu\|$ (au)	$\|\mu_{01}\|$ (au)	δ^{2PA} (GM)
1a	−5	−15	52	−5	222
1b	−4	−21	47	−5	178
2a	−7	−12	38	−7	117
2b	−5	−16	30	−5	71
6	−4	−26	33	−12	6
7	−4	−25	30	−12	−1
8	−5	−15	57	−5	246
9	−6	−22	40	−9	72
11	−5	−17	31	−7	78
12	−8	−26	83	−10	205
13	−8	−24	50	−9	100
14	−8	−23	55	−8	121
15	−5	−16	8	−7	−32
16a	−15	−59	61	−30	30
16b	−16	−59	95	−30	103
17	−9	−24	50	−9	100
18	−7	−30	31	−13	36
19	−7	−45	48	−23	33
20	−7	−40	15	−19	−14
21	−19	−85	21	−57	−74
22	−3	−18	−33	−9	−71
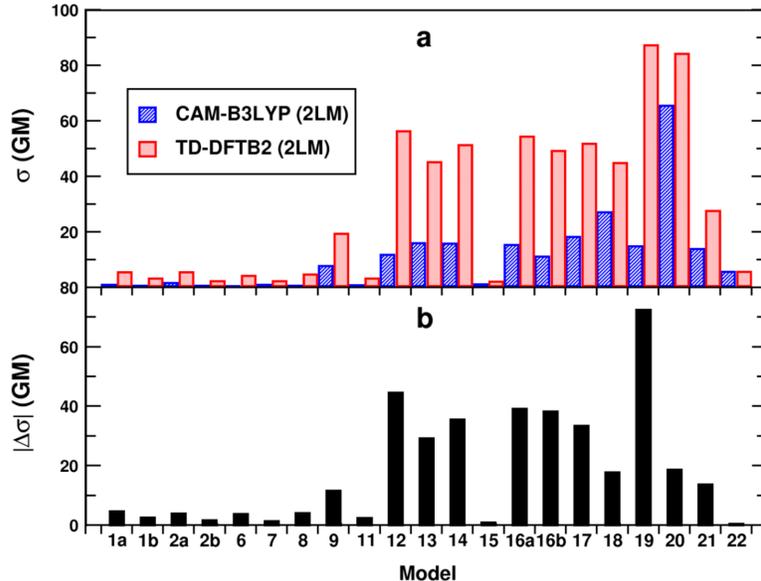
Figure A.3. For the RFP-like chromophores, (a) comparison of the magnitude of the 2PA cross-sections (σ^{2PA}) obtained using TD-DFTB2 with respect to those reported previously using TD-CAM-B3LYP/6-31+G(d,p) via the 2LM[14] and (b) the corresponding differences between the σ^{2PA} values obtained using TD-DFTB2 and TD-CAM-B3LYP/6-31+G(d,p) ($\left|\Delta\sigma^{2PA}\right| = \left|\sigma^{2PA}_{DFTB2} - \sigma^{2PA}_{B3LYP}\right|$). Negative differences are shaded in grey whereas the positive in black.
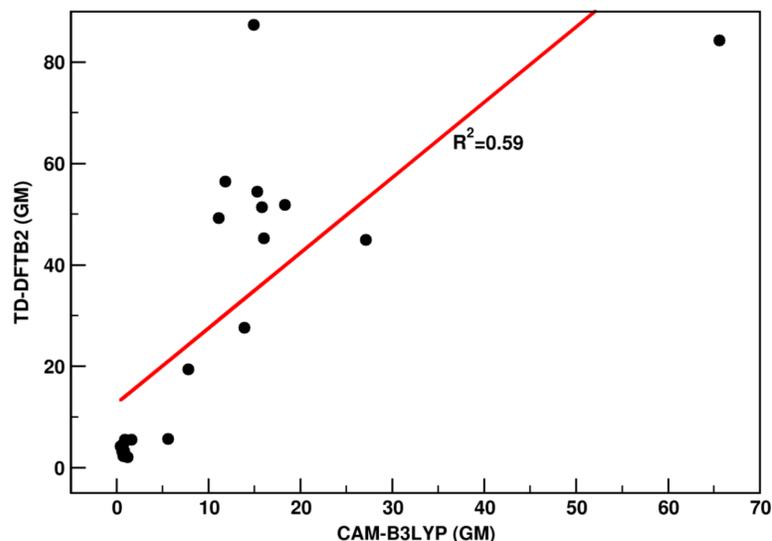
Figure A.4. Evaluation of the correlation, considering all the models in Figure 2.2, between TD-DFTB2 and CAM-B3LYP[14] 2PA cross-sections for the RFP-like chromophores, obtained via 2LM in both cases.
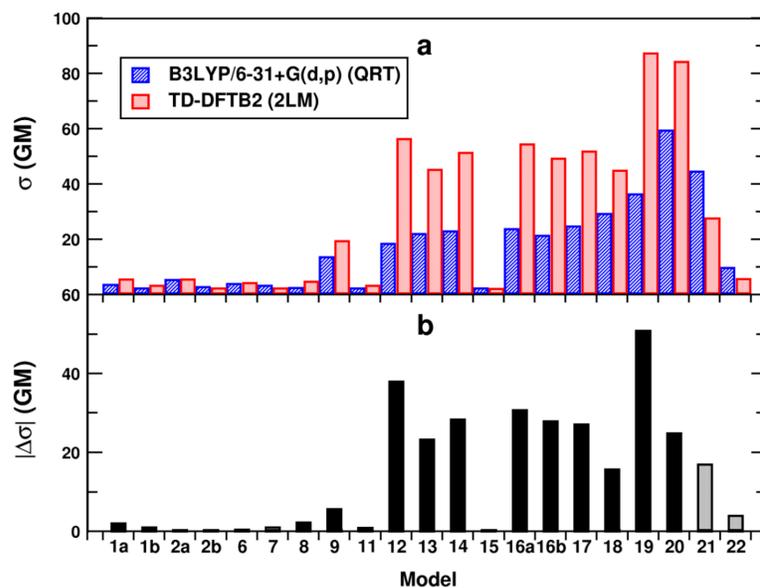
Figure A.5. For the RFP-like chromophores, (a) comparison of the magnitude of TD-DFTB2 2PA cross-sections (σ^{2PA}) with respect those obtained within QRT at TD-B3LYP/6-31+G(d,p)[14], and (b) The corresponding differences between the 2LM TD-DFTB2 and QRT TD-B3LYP/6-31+G(d,p) σ^{2PA} values ($\left|\Delta\sigma^{2PA}\right| = \left|\sigma^{2PA}_{DFTB2} - \sigma^{2PA}_{B3LYP}\right|$). Negative differences are shaded in grey whereas the positive in black.
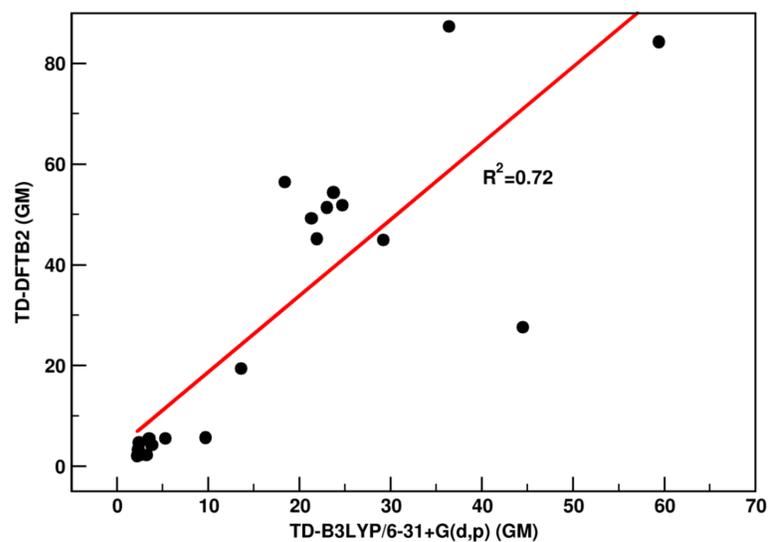
Figure A.6. Evaluation of the correlation, considering all the models in Figure 2.2, between TD-DFTB2 and TD-B3LYP/6-31+G(d,p)[14] 2PA cross-sections for the RFP-like models, obtained via 2LM and QRT, respectively.
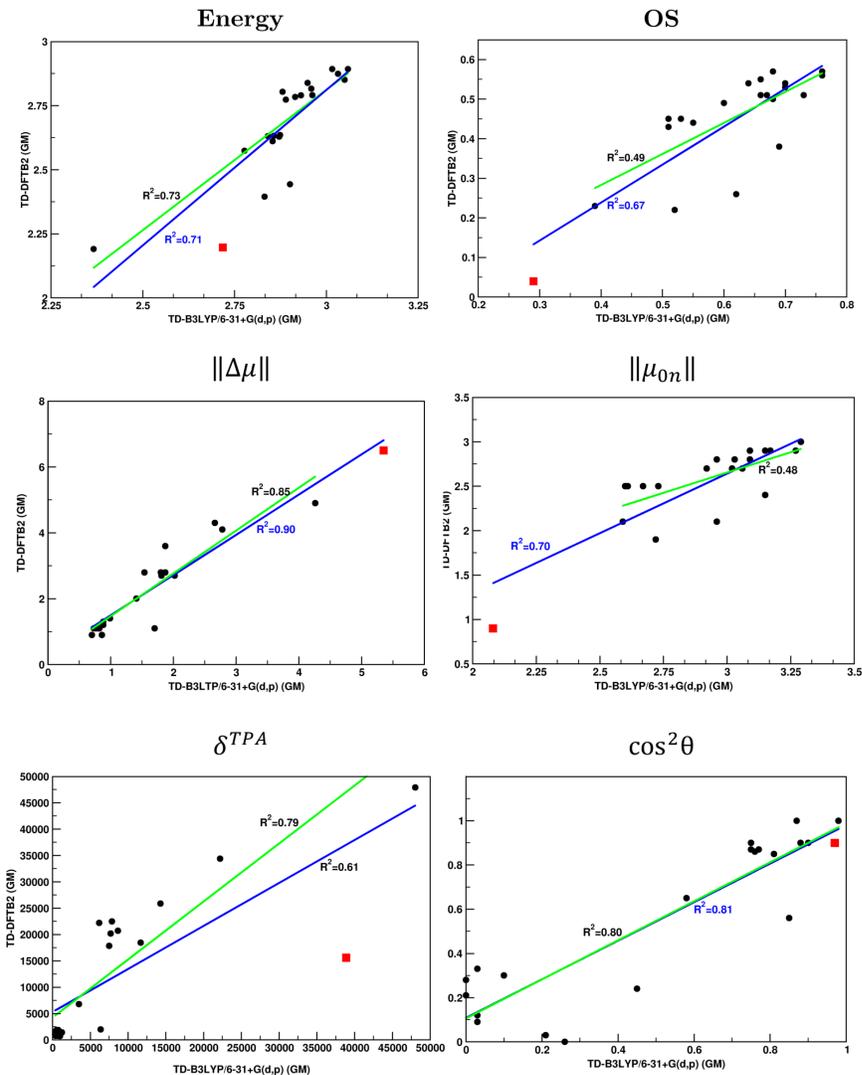
Figure A.7. Evaluation of the correlation i) in blue, between excitation energies, oscillator strengths (OS), $\|\Delta\mu\|$, 2PA transition moments (δ^{2PA}), transition dipole moments ($\|\mu_{01}\|$) and $\cos^2\theta$ obtained at TD-DFTB2 and those at TD-B3LYP/6-31+G(d,p)[14] for all the RFP-like models in Figure 2.2 and ii) in green, correlation between the same excited state properties but discarding model 21 (in red).

Table A.6. Considering All the RFP-Like Chromophores, Evaluation of the Linear Correlation (R^2) Between 2PA Cross-Sections (σ^{2PA}), Excitation Energy, δ^{2PA}, $\|\Delta\mu\|$, Ground and Excited State Permanent Dipole Moments (μ_{00} and μ_{11}, Respectively), as well as $\cos^2\theta$ Obtained via 2LM Using the TD-DFTB and the TD-B3LYP/6-31+G(d,p)[14] Methods

	σ^{2PA}	**Energy**	**OS**	δ^{2PA}	$\|\Delta\mu\|$	μ_{00}	μ_{11}	$\|\mu_{01}\|$	$\cos^2\theta$
R^2	0.72	0.17	0.22	0.61	0.05	0.55	$3.0E-3$	0.14	0.65

Table A.7. Excitation Energies, the Corresponding Excited State S_n, Oscillator Strengths (OS), 2PA Transition Moments (δ^{2PA}) and 2PA Cross-Sections (σ^{2PA}) of the HOMO→LUMO Transitions for the GFP-Like Models at the TD-DFTB2 and TD-B3LYP/6-31+G(d,p) Levels of Theory.[12] Salem and Brown[12] Only Reported Those Dipole Moments Where the First Gas-Phase Excited State Corresponded to the Results Obtained Within the Polarizable Continuum Model (PCM)

Model	S_n	TD-DFTB2 Energy (eV)	OS	δ^{2PA}	σ^{2PA} (GM)	TD-B3LYP/6-31+G(d,p) Energy (eV)	OS	δ^{2PA}	σ^{2PA} (GM)
1a	2	3.483	0.52	628	2.8				
1b	2	3.486	0.52	750	3.3				
2a	1	2.505	0.00	0	0.0				
2b	1	2.560	0.00	0	0.0				
6	4	3.229	0.60	8502	32.5	3.336	0.81	135	1
7	4	3.234	0.62	8306	31.9	3.331	0.81	275	1
8	2	3.361	0.65	1006	4.2				
9	2	3.445	0.66	221	1.0				
11	1	2.379	0.00	0	0.0				
12	2	3.189	0.68	8219	30.6	3.351	0.86	1240	5
13	2	3.270	0.62	4907	19.2	3.444	0.74	1468	6
14	2	3.278	0.68	6098	24.0	3.451	0.79	1430	6
15	1	2.227	0.00	221	0.4				
16a	2	2.896	0.18	6409	19.7	3.348	0.42	5434	22
16b	2	2.923	0.19	5739	18.0				
17	2	3.255	0.66	6366	24.7				
18	2	3.235	0.63	4120	15.8	3.352	0.79	2386	10
19	3	3.250	0.49	2874	11.1	3.250	0.33	7929	31
20	1	2.707	0.16	9221	24.8	2.826	0.28	11705	34
21	1	2.552	0.03	3485	8.3				
22	4	3.236	0.50	8256	31.7	3.257	0.61	11847	46

Table A.8. For the GFP-like Models, at Both TD-DFTB2 and TD-B3LYP/6-31+G(d,p) Levels of Theory, The Vectors Corresponding to the Difference Between the Excited State and the Ground State Permanent Dipole Moments ($\|\Delta\mu\| = \left(\sum \Delta\mu_\alpha^2\right)^{1/2}$), the Transition Dipole Moments ($\|\mu_{01}\|$), and the Alignment Between Them.[12] All Values are Given in Atomic Units. Salem and Brown[12] Only Reported Those Dipole Moments Where the First Gas-Phase Excited State Corresponded to the Results Obtained Within the Polarizable Continuum Model (PCM)

	TD-DFTB2			TD-B3LYP/6-31+G(d,p)		
Model	$\|\Delta\mu\|$ **(au)**	$\|\mu_{01}\|$ **(au)**	$\cos^2\theta$	$\|\Delta\mu\|$ **(au)**	$\|\mu_{01}\|$ **(au)**	$\cos^2\theta$
1a	0.75	2.47	0.90			
1b	0.81	2.46	0.94			
2a	5.89	0.00	0.00			
2b	6.05	0.00	0.00			
6	2.21	2.77	1.00	0.36	3.14	0.26
7	2.16	2.80	1.00	0.44	3.16	0.50
8	0.79	2.81	0.95			
9	0.40	2.80	0.86			
11	6.02	0.01	0.00			
12	2.01	2.96	1.00	0.75	3.24	1.00
13	1.69	2.78	1.00	0.92	2.96	1.00
14	1.82	2.90	1.00	0.88	3.06	1.00
15	10.07	0.07	0.98			
16a	3.14	1.58	0.88	2.27	2.26	0.96
16b	2.97	1.61	0.85			
17	1.85	2.88	1.00			
18	1.51	2.82	1.00	1.09	3.10	1.00
19	1.45	2.47	1.00	2.98	2.04	0.94
20	3.72	1.55	0.79	3.36	1.99	0.82
21	5.08	0.70	0.65			
22	2.42	2.50	1.00	2.65	2.75	1.00
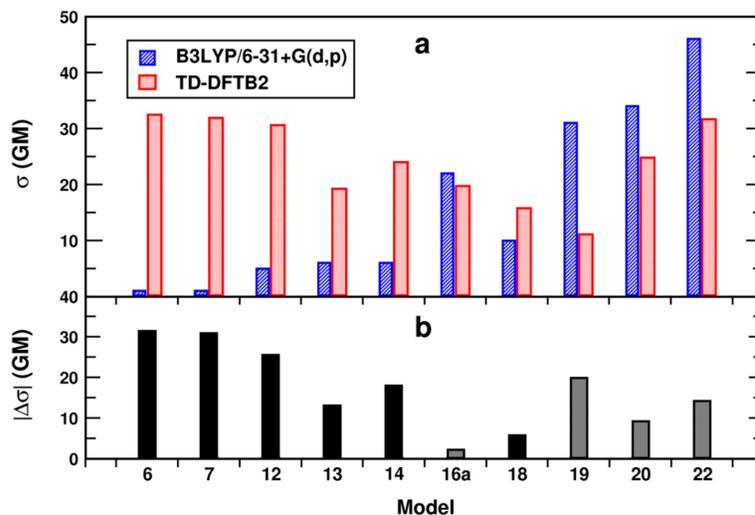
Figure A.8. For the GFP-like chromophores, (a) comparison of the magnitude of the 2PA cross-sections (σ^{2PA}) obtained using TD-DFTB2 with respect to those reported previously using TD-B3LYP/6-31+G(d,p) via the 2LM[14] and (b) the corresponding differences between the σ^{2PA} values obtained using TD-DFTB2 and TD-B3LYP/6-31+G(d,p) ($\left|\Delta\sigma^{2PA}\right| = \left|\sigma^{2PA}_{DFTB2} - \sigma^{2PA}_{B3LYP}\right|$). Negative differences are shaded in grey whereas the positive in black.

Table A.9. Mean, Median, and Standard Deviation of the Excitation Energy, Oscillator Strength (OS), 2PA Transition Moment δ^{2PA}, 2PA Cross-Sections (σ^{2PA}), $\|\Delta\mu\|$, Transition Dipole Moment ($\|\mu_{01}\|$) and $\cos^2\theta$ of the GFP-Like Models Obtained at TD-DFTB2 Level of Theory Within the 2LM Model Considering All 21 Models

	Energy (eV)	OS	δ^{2PA}	σ^{2PA}(GM)	$\|\Delta\mu\|$	$\|\mu_{01}\|$	$\cos^2\theta$
Mean	3.033	0.40	4063	14.5	2.99	1.95	0.80
Median	3.234	0.52	4120	15.8	2.16	2.47	0.95
Mean Deviation	0.335	0.25	2995	10.8	1.81	0.96	0.24
Standard Deviation	0.380	0.27	3313	11.9	2.34	1.09	0.34

159

Table A.10. Ground and Excited State Permanent Dipole Moment Components (μ_α) for Each of the GFP-like Chromophores of Figure 2.2 at TD-DFTB2 and TD-B3LYP/6-31+G(d,p),[12] Both Obtained Within the 2LM. All Values Are Given in Atomic Units

Model	TD-DFTB2						TD-B3LYP/6-31+G(d,p)					
	μ_{00}			μ_{11}			μ_{00}			μ_{11}		
	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x	μ_x
1a	1.87	1.51	0.00	2.62	1.60	0.00						
1b	1.06	−0.01	0.00	1.80	0.33	0.00						
2a	1.25	−0.97	0.00	−4.34	−2.84	0.00						
2b	0.87	1.79	0.05	−5.07	0.66	0.01						
3	0.00	0.00	0.00	0.00	0.00	0.00						
5	0.00	0.00	0.00	0.00	0.00	0.00						
6	0.89	−1.40	0.00	3.07	−1.76	0.00	0.38	−1.48	0.00	0.59	−1.19	0.00
7	0.82	−1.37	0.15	2.95	−1.71	0.20	0.36	−1.43	0.12	0.70	−1.15	0.11
8	−1.87	−0.75	0.00	−2.64	−0.94	0.00						
9	−0.12	−0.98	0.00	0.24	−1.12	0.00						
10	0.00	0.00	0.00	0.00	0.00	0.00						
11	1.36	−1.64	0.02	−4.65	−1.98	0.00						
12	−0.32	−0.91	0.00	1.69	−0.98	0.00	−0.49	−0.83	0.00	−1.24	−0.80	0.00
13	0.46	−1.45	0.00	2.15	−1.49	0.00	0.61	−1.46	0.00	−0.30	−1.49	0.00
14	−0.24	−1.43	0.00	−2.06	−1.41	0.00	−0.43	−1.41	0.00	0.44	−1.47	0.00
15	1.83	−1.30	0.95	−8.17	−2.43	0.82				0.00	0.00	0.00
16a	−0.40	−1.61	0.00	−3.53	−1.32	0.00	−0.36	−1.93	0.00	1.91	−1.99	0.00
16b	0.78	0.59	0.00	3.70	1.17	0.00						
17	−0.54	−1.39	−0.14	−2.39	−1.33	−0.16						
18	0.87	−1.13	0.18	2.39	−1.22	0.18	1.23	−1.08	0.30	0.14	−1.10	0.31
19	0.47	−0.89	0.00	1.92	−0.95	0.00	−0.04	−0.87	0.00	2.93	−0.59	0.00
20	−0.19	−1.25	0.33	3.41	−0.31	0.42	0.21	−1.40	0.43	−3.05	−2.22	0.43
21	−0.05	−0.79	0.00	5.02	−1.13	0.00						
22	2.72	−0.67	0.00	5.13	−0.82	0.00	2.40	−0.63	0.00	−0.24	−0.58	0.00

Table A.11. For the GFP-Like Chromophores, Evaluation of the Linear Correlation (R^2) Between 2PA Cross-Sections (σ^{2PA}), Excitation Energy, δ^{2PA}, $\|\Delta\mu\|$, Ground and Excited State Permanent Dipoles (μ_{00} and μ_{11}, Respectively), and $\cos^2\theta$ Obtained Both via the 2LM at the TD-DFTB and TD-B3LYP/6-31+G(d,p)[12] Levels of Theory

	σ^{2PA}	**Energy**	**OS**	δ^{2PA}	$\|\Delta\mu\|$	μ_{00}	μ_{11}	$\|\mu_{01}\|$	$\cos^2\theta$
R^2	0.03	0.61	0.76	0.01	0.25	0.86	0.04	0.71	$3.0E-3$
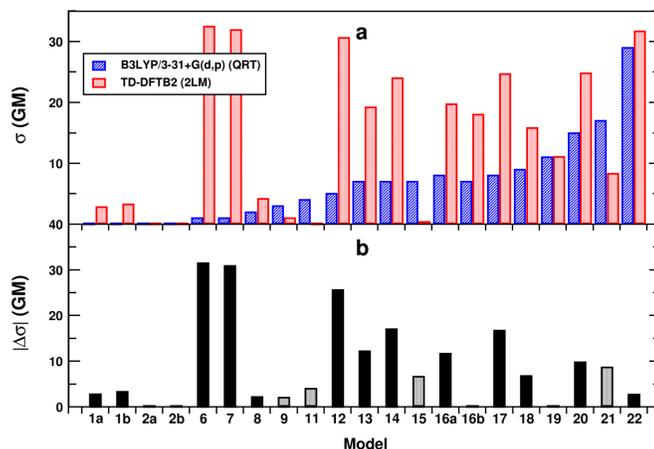
Figure A.9. For the GFP-like chromophores, (a) comparison of the magnitude of TD-DFTB2 2PA cross-sections (σ^{2PA}) with respect to B3LYP/6-31+G(d,p) ones,[12] obtained within the 2LM and QRT, respectively (b) The differences between the TD-DFTB2 and B3LYP/6-31+G(d,p) σ^{2PA} values ($\left|\Delta\sigma^{2PA}\right| = \left|\sigma^{2PA}_{DFTB2} - \sigma^{2PA}_{B3LYP}\right|$).
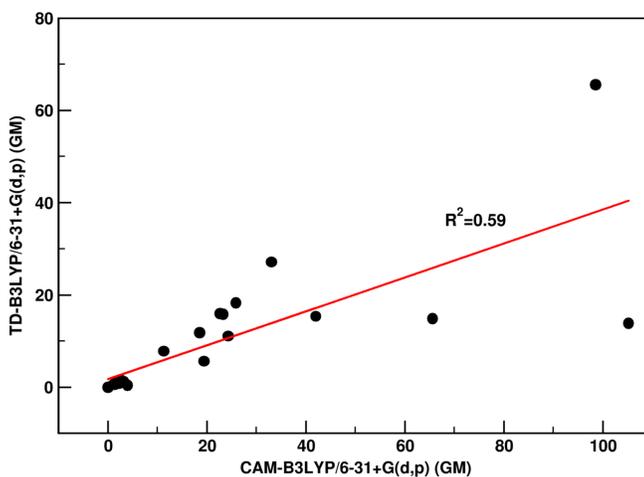
Figure A.10. For the RFP-like chromophores, evaluation of the linear correlation (R^2) between 2PA cross-sections (σ^{2PA}) at TD-B3LYP/6-31+G(d,p) and CAM-B3LYP/6-31+G(d,p)[12,14] levels of theory, both obtained via 2LM.

161

Appendix B

Determination of Two-Photon-Absorption Cross Sections of Fluorescent Protein Chromophores Using Long-Range Corrected Time-Dependent Density Functional Theory Tight Binding

Table B.1. For the Canonical Models Shown in Figure 3.1, the Vertical Excitation Energies (VEE), Transition Nature (MOs), 2PA Cross-section (σ^{2PA}), and Two-level Model (2LM) Elements Corresponding to the First Three Lowest Excited States (S_{1-3}) Obtained Using LC-TD-DFTB

Chro	S_n	VEE (eV)	OS	MOs	$\Delta\mu$ (au)	μ_{trans}^{0n} (au)	$\cos(\theta)^2$	δ_{2PA}	σ_n^{2PA} (GM)
BFP_1	1	3.194	0.001	HOMO-1 → LUMO	2.49	0.04	0	1	0
BFP_1	2	3.879	0.451	HOMO → LUMO	0.56	0.86	1	36	0
BFP_1	3	4.592	0.000	HOMO-5 → LUMO	2.47	0.02	0	0	0
BLB	1	2.812	0.001	HOMO-1 → LUMO	6.36	0.04	0	8	0
BLB	2	3.131	0.000	HOMO-2 → LUMO	8.44	0.02	0	2	0
BLB	3	3.392	0.479	HOMO → LUMO	8.50	0.95	1	13300	56
CFP	1	3.217	0.001	HOMO-2 → LUMO	4.29	0.04	0	2	0
CFP	2	3.799	0.475	HOMO → LUMO	1.27	0.89	1	202	1
CFP	3	4.548	0.000	HOMO-4 → LUMO	3.70	0.02	0	0	0
GFP_n	1	3.125	0.001	HOMO-1 → LUMO	0.92	0.04	0	0	0
GFP_n	2	3.799	0.486	HOMO → LUMO	2.73	0.90	1	990	5
GFP_n	3	4.462	0.000	HOMO-3 → LUMO	1.81	0.01	0	0	0

Table B.2. For the Non-Canonical Models Shown in Figure 3.2, the Vertical Excitation Energies (VEE), Transition Nature (MOs), 2PA Cross-Sections (σ^{2PA}s), and Two-Level Model (2LM) Components Corresponding to the First Three Lowest Excited States (S_{1-3})

Chro	S_n	VEE (eV)	OS	MOs	$\Delta\mu$ (au)	μ^{0n}_{trans} (au)	$\cos(\theta)^2$	δ_{2PA}	σ_n^{2PA} (GM)
1a	1	2.862	0.001	HOMO–2 → LUMO	16.3	0.0	0.0	42	0
1a	2	3.400	0.384	HOMO → LUMO	15.4	0.8	0.6	26285	111
1a	3	3.811	0.001	HOMO–2 → LUMO	15.2	0.0	0.1	29	0
1b	1	2.863	0.001	HOMO–2 → LUMO	13.2	0.0	0.0	29	0
1b	2	3.400	0.378	HOMO → LUMO	11.9	0.8	0.7	16869	72
1b	3	3.831	0.001	HOMO–1 → LUMO	12.0	0.0	0.0	18	0
2a	1	2.866	0.001	HOMO–2 → LUMO	4.4	0.0	0.0	3	0
2a	2	3.395	0.388	HOMO → LUMO	4.3	0.9	0.1	1156	5
2a	3	3.517	0.000	HOMO–1 → LUMO+1	3.4	0.0	0.0	0	0
2b	1	2.871	0.001	HOMO–3 → LUMO	9.5	0.0	0.0	15	0
2b	2	3.404	0.375	HOMO → LUMO	9.7	0.8	0.1	5567	24
2b	3	3.541	0.000	HOMO–2 → LUMO+1	8.6	0.0	0.0	2	0
8	1	2.852	0.001	HOMO–2 → LUMO	15.9	0.0	0.0	42	0
8	2	3.344	0.446	HOMO → LUMO	14.6	0.9	0.7	31087	127
8	3	3.822	0.001	HOMO–1 → LUMO	14.7	0.0	0.0	28	0
9	1	2.897	0.001	HOMO–1 → LUMO	3.8	0.0	0.1	3	0
9	2	3.388	0.429	HOMO → LUMO	6.1	0.9	0.3	3434	15
9	3	3.855	0.001	HOMO–2 → LUMO	4.2	0.0	0.1	3	0
11	1	2.850	0.001	HOMO–2 → LUMO	7.4	0.0	0.0	9	0
11	2	3.333	0.455	HOMO → LUMO	7.9	0.9	0.1	4272	17
11	3	3.394	0.000	HOMO–1 → LUMO	6.7	0.0	0.0	2	0
12	1	2.931	0.001	HOMO–1 → LUMO	15.2	0.0	0.0	36	0
12	2	3.344	0.523	HOMO → LUMO	13.8	1.0	0.7	32255	132
12	3	3.888	0.002	HOMO–2 → LUMO	14.7	0.1	0.0	29	0
13	1	2.972	0.001	HOMO–1 → LUMO	6.8	0.0	0.0	7	0
13	2	3.369	0.517	HOMO → LUMO	7.2	1.0	0.1	3954	17
13	3	3.904	0.002	HOMO–2 → LUMO	6.7	0.1	0.2	8	0
14	1	2.965	0.001	HOMO–1 → LUMO	12.8	0.0	0.0	25	0
14	2	3.364	0.526	HOMO → LUMO	11.7	1.0	0.7	21627	90
14	3	3.871	0.002	HOMO–2 → LUMO	12.5	0.0	0.1	23	0
16a	1	2.941	0.001	HOMO–1 → LUMO	4.9	0.0	0.0	4	0
16a	2	3.362	0.487	HOMO → LUMO	5.4	1.0	0.0	1968	8
16a	3	3.864	0.001	HOMO–3 → LUMO	4.8	0.0	0.3	5	0
16b	1	2.942	0.001	HOMO–1 → LUMO	6.5	0.0	0.0	7	0
16b	2	3.370	0.495	HOMO → LUMO	7.2	1.0	0.0	3570	15
16b	3	3.898	0.002	HOMO–3 → LUMO	6.4	0.1	0.2	8	0
17	1	2.969	0.001	HOMO–1 → LUMO	7.1	0.0	0.0	8	0
17	2	3.357	0.531	HOMO → LUMO	7.8	1.0	0.0	4514	19
17	3	3.877	0.002	HOMO–2 → LUMO	7.1	0.1	0.3	11	0
18	1	2.972	0.001	HOMO–1 → LUMO	8.8	0.0	0.0	12	0
18	2	3.351	0.521	HOMO → LUMO	8.3	1.0	0.3	7980	33
18	3	3.881	0.002	HOMO–2 → LUMO	8.6	0.1	0.2	14	0
19	1	2.862	0.001	HOMO–1 → LUMO	15.7	0.0	0.0	39	0
19	2	3.333	0.497	HOMO → LUMO	14.5	1.0	0.7	35224	143
19	3	3.808	0.001	HOMO–2 → LUMO	14.6	0.0	0.0	28	0
20	1	3.040	0.001	HOMO–1 → LUMO	11.9	0.1	0.3	52	0
20	2	3.313	0.504	HOMO → LUMO	10.7	1.0	0.7	18501	75
20	3	3.870	0.002	HOMO–3 → LUMO	12.1	0.0	0.1	23	0
21	1	2.870	0.001	HOMO–2 → LUMO	13.0	0.0	0.0	28	0
21	2	3.359	0.479	HOMO → LUMO	12.0	1.0	0.6	19870	82
21	3	3.860	0.002	HOMO–3 → LUMO	11.9	0.1	0.0	20	0
22	1	2.777	0.001	HOMO–2 → LUMO	6.1	0.0	0.0	6	0
22	2	3.302	0.403	HOMO → LUMO	7.6	0.9	0.0	3362	13
22	3	3.757	0.001	HOMO–1 → LUMO	5.7	0.0	0.2	5	0
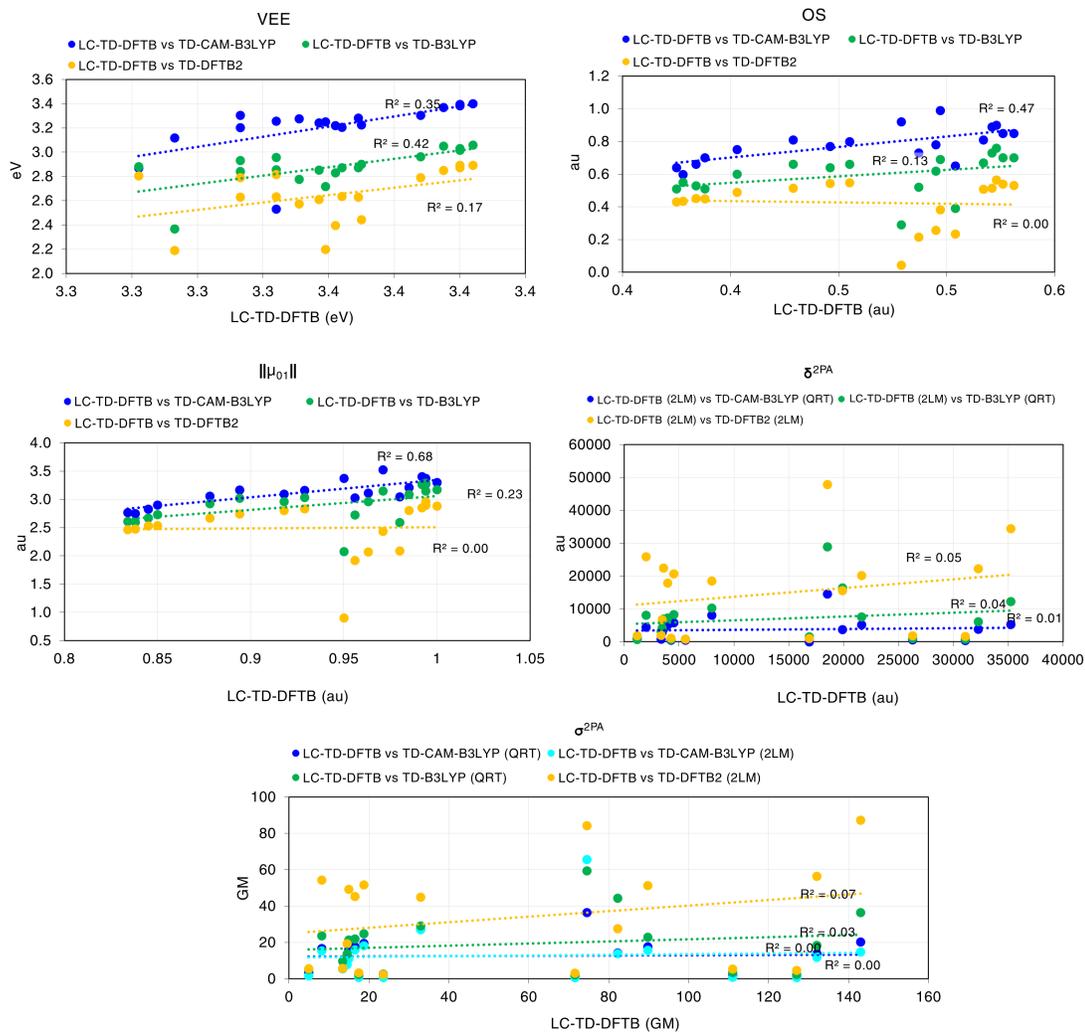
Figure B.1. For all models in Figures 3.1 and 3.2, the correlation between the 2PA cross-sections and the 2LM elements including vertical excitation energies (VEE), transition dipole moments ($\|\mu_{02}\|$), 2PA transition moment (δ^{2PA}), and oscillator strengths (OS) obtained using LC-TD-DFTB2 and TD-B3LYP/6-31+G(d,p),[14] as well as TD-CAM-B3LYP[14] and TD-DFTB2[61]
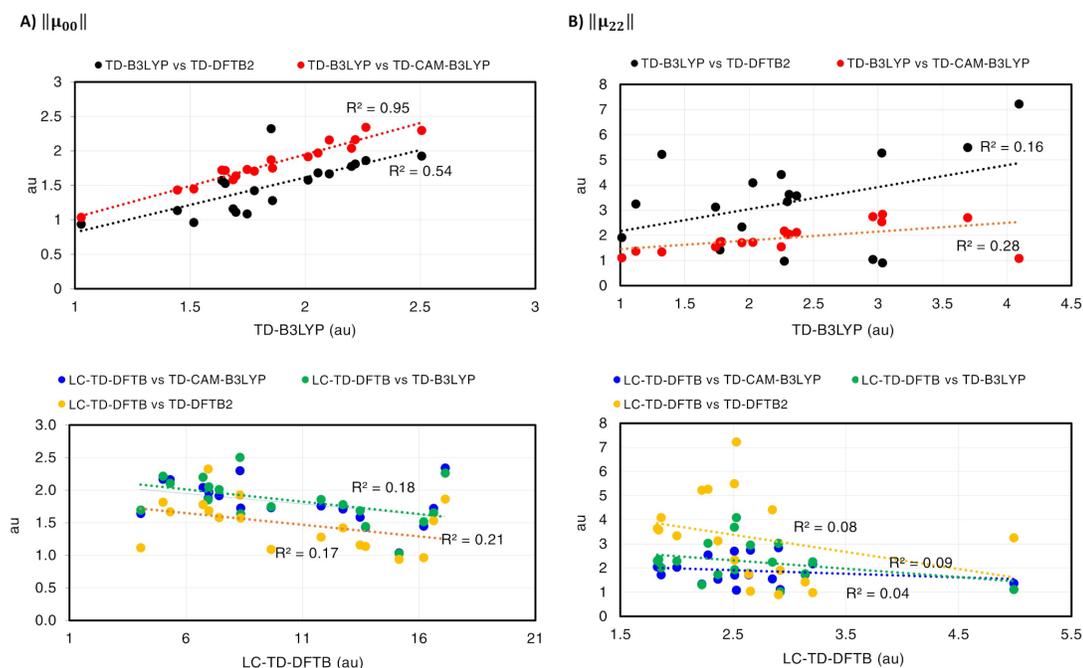
Figure B.2. For all models in Figures 3.1 and 3.2, correlation between the A) ground- and B) excited-state (S_2) permanent dipole moments results obtained in this work using LC-TD-DFTB and those previously obtained using TD-B3LYP/6-31+G(d,p),[14] TD-CAM-B3LYP,[14] and TD-DFTB2.[61]

Figure B.3. For selected models and their lowest energy transitions (S_n) with oscillator strength different from zero (refer to main text for the corresponding discussion on this), the molecular orbitals with the most significant contribution to the transition.

Appendix C

Three-photon Absorption Cross-Sections of Serotonin, Fluorescein, Rhodamine 6G, and Fluorescent Protein Chromophores

C.1   Molecular Orbitals
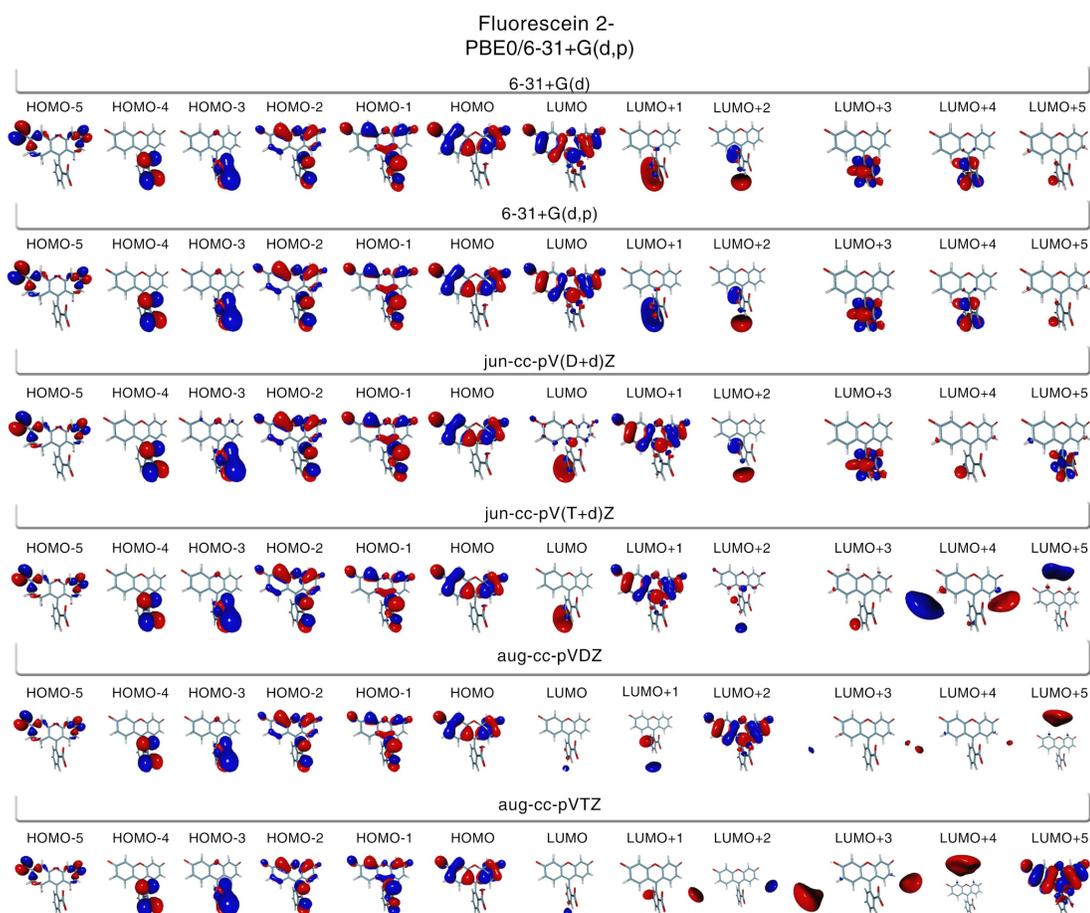
Figure C.1. Part A. Fluorescein^{2-} molecular orbitals computed in gas-phase using the CAM-B3LYP functional and different basis sets. All plots were obtained using gmolden. [301,302]

Fluorescein 2-
PBE0/6-31+G(d,p)

pcseg-1

HOMO-5 HOMO-4 HOMO-3 HOMO-2 HOMO-1 HOMO LUMO LUMO+1 LUMO+2 LUMO+3 LUMO+4 LUMO+5

pcseg-2

HOMO-5 HOMO-4 HOMO-3 HOMO-2 HOMO-1 HOMO LUMO LUMO+1 LUMO+2 LUMO+3 LUMO+4 LUMO+5

aug-pcseg-1

HOMO-5 HOMO-4 HOMO-3 HOMO-2 HOMO-1 HOMO LUMO LUMO+1 LUMO+2 LUMO+3 LUMO+4 LUMO+5

aug-pcseg-2

HOMO-5 HOMO-4 HOMO-3 HOMO-2 HOMO-1 HOMO LUMO LUMO+1 LUMO+2 LUMO+3 LUMO+4 LUMO+5
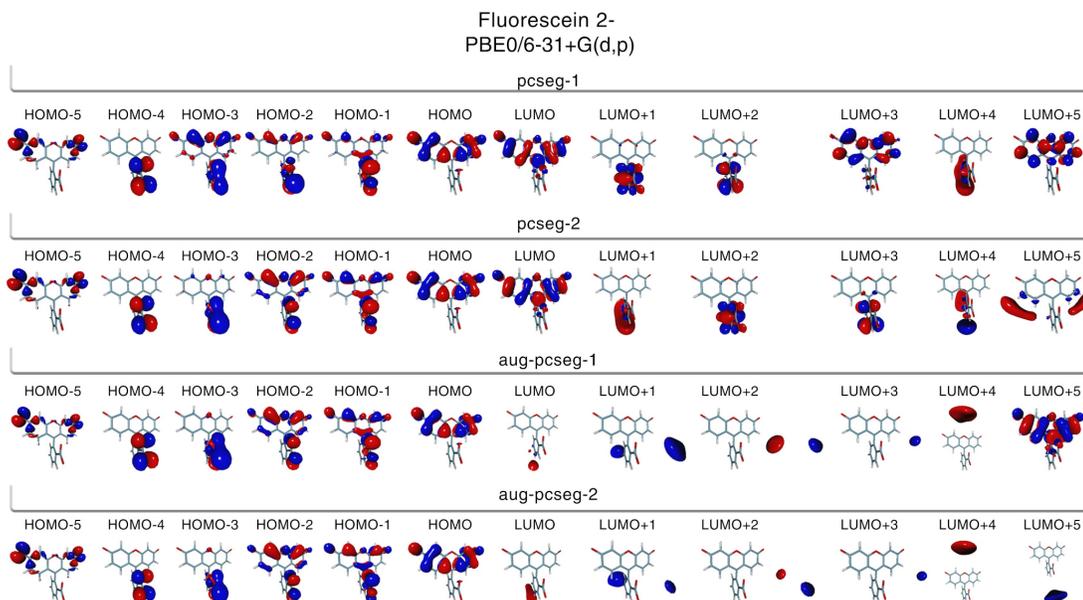
Figure C.2. Part B. Fluorescein^{2-} molecular orbitals computed in gas-phase using the CAM-B3LYP functional and different basis sets. All plots were obtained using gmolden.[301,302]

Fluorescein 2-
Structure obtained by Gerasimova et al.

6-31+G(d,p)

HOMO-5 HOMO-4 HOMO-3 HOMO-2 HOMO-1 HOMO LUMO LUMO+1 LUMO+2 LUMO+3 LUMO+4 LUMO+5

jun-cc-pV(T+d)Z

HOMO-5 HOMO-4 HOMO-3 HOMO-2 HOMO-1 HOMO LUMO LUMO+1 LUMO+2 LUMO+3 LUMO+4 LUMO+5

aug-cc-pVTZ

HOMO-5 HOMO-4 HOMO-3 HOMO-2 HOMO-1 HOMO LUMO LUMO+1 LUMO+2 LUMO+3 LUMO+4 LUMO+5

pcseg-2

HOMO-5 HOMO-4 HOMO-3 HOMO-2 HOMO-1 HOMO LUMO LUMO+1 LUMO+2 LUMO+3 LUMO+4 LUMO+5

aug-pcseg-2

HOMO-5 HOMO-4 HOMO-3 HOMO-2 HOMO-1 HOMO LUMO LUMO+1 LUMO+2 LUMO+3 LUMO+4 LUMO+5
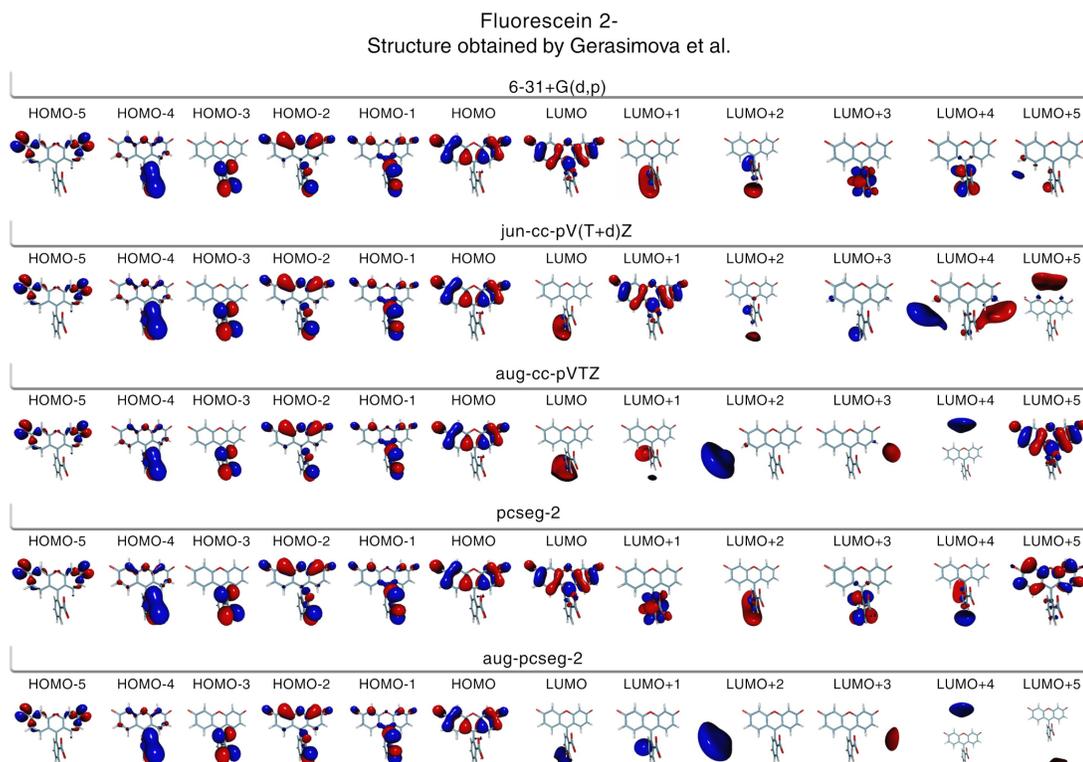
Figure C.3. For the fluorescein^{2-} structure obtained from the work published by Gerasimova et al.,[222] molecular orbitals computed in gas-phase using the CAM-B3LYP functional and different basis sets. All plots were obtained using gmolden.[301,302]

167
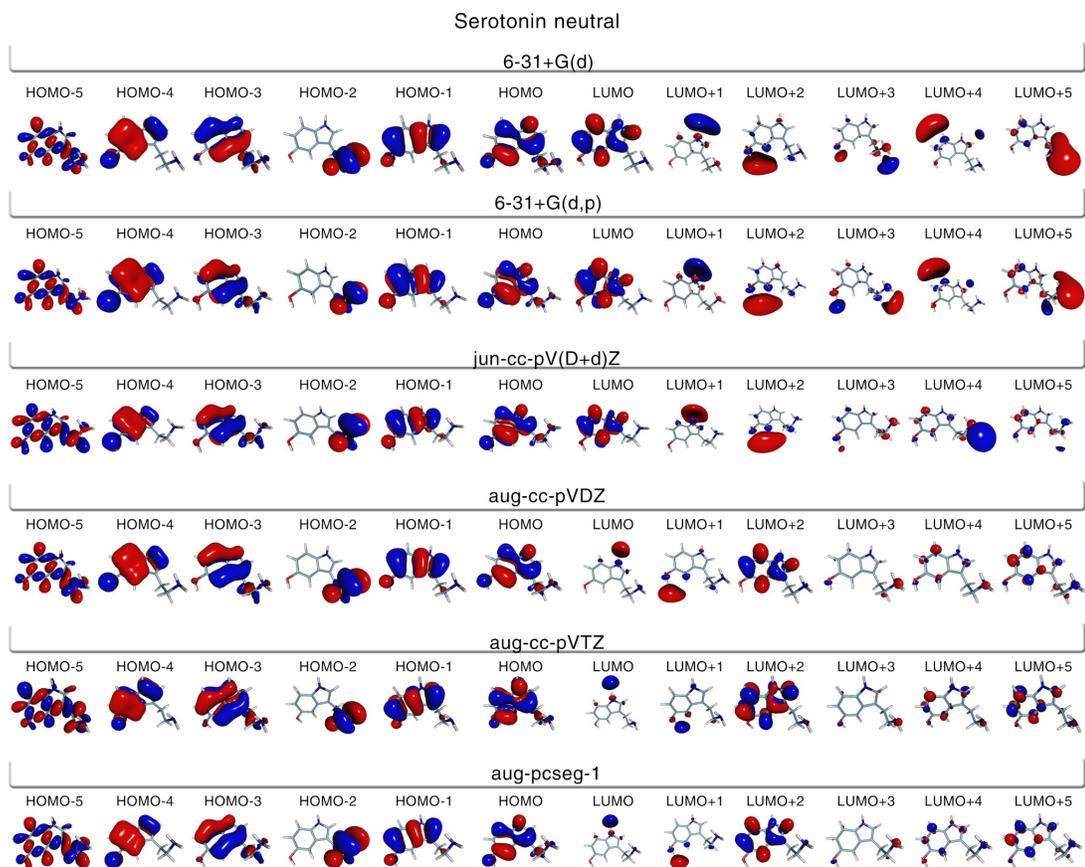
Figure C.4. Serotonin (neutral) molecular orbitals computed in gas-phase using the CAM-B3LYP functional and different basis sets. All plots were obtained using gmolden.[301,302]
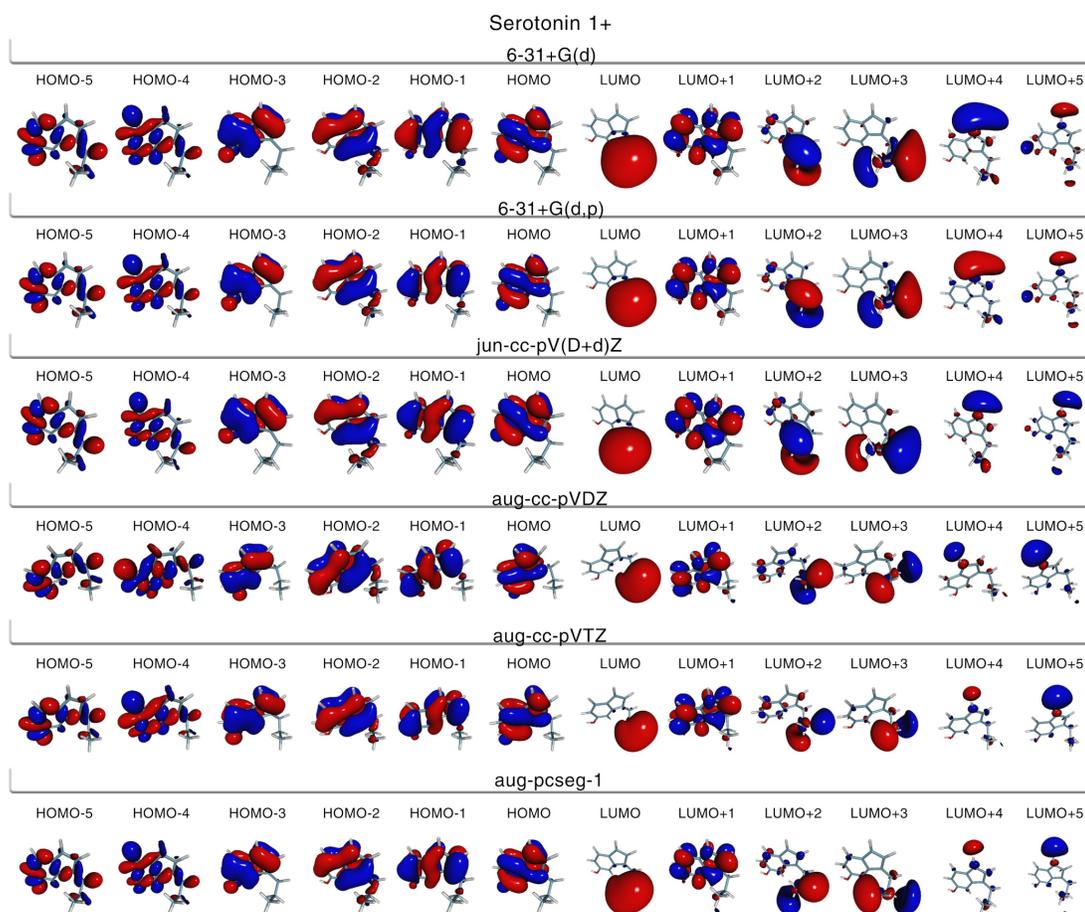
Figure C.5. Serotonin$^+$ molecular orbitals computed in gas-phase using the CAM-B3LYP functional and different basis sets. All plots were obtained using gmolden.[301,302]

Rhodamine 6G 1+

B3LYP/6-31+G(d,p)

aug-cc-pVDZ

| HOMO-5 | HOMO-4 | HOMO-3 | HOMO-2 | HOMO-1 | HOMO | LUMO | LUMO+1 | LUMO+2 | LUMO+3 | LUMO+4 | LUMO+5 |

CAM-B3LYP/6-31+G(d,p)

aug-cc-pVDZ

| HOMO-5 | HOMO-4 | HOMO-3 | HOMO-2 | HOMO-1 | HOMO | LUMO | LUMO+1 | LUMO+2 | LUMO+3 | LUMO+4 | LUMO+5 |
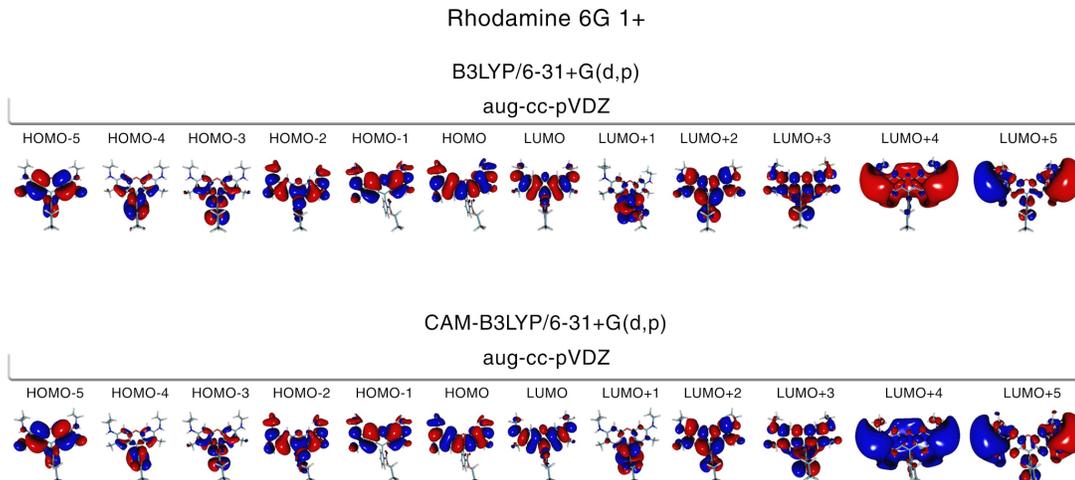
Figure C.6. Rhodamine $6G^+$ molecular orbitals computed in gas-phase using the CAM-B3LYP/aug-cc-pVDZ method. Top: results for the structure optimized using the B3LYP/6-31+G(d,p) method, bottom: results for the structure optimized using the CAM-B3LYP/6-31+G(d,p) method. All plots were obtained using gmolden. [301,302]

C.2 Obtaining Two-photon Absorption Cross-Sections

Two-photon absorption cross-sections are obtained according to the following equation

$$\sigma^{2PA} = \frac{N\pi^2 a_0^5 \alpha}{c} \frac{\omega^2}{\Gamma} \delta^{2PA} \tag{C.1}$$

where ω is half the vertical excitation energies, N is set as 4 and Γ_f is the broadening factor assumed to be half width at half maximum (HWHM) and set to 0.1 eV to ease comparison with experiment. [60]

C.3 One-photon Absorption Data

Table C.1. For Fluorescein^{2-}, the Vertical Excitation Energies (VEE), Oscillator Strengths (OS), and Molecular Orbitals (MOs) Involved in the First Five Electronic Transitions (S_n) Computed in Gas-phase Using the CAM-B3LYP Functional and Different Basis Sets

S_n	VEE (eV)	VEE (nm)	OS	Transition MOs
		Fluorescein 2-		
		6-31+G(d) (Cartesian)		
1	3.091	401	0.755	HOMO $\rightarrow$ LUMO
2	3.937	315	0.075	HOMO–1 $\rightarrow$ LUMO
3	4.003	310	0.006	HOMO $\rightarrow$ LUMO+1
4	4.047	306	0.006	HOMO–2 $\rightarrow$ LUMO
5	4.219	294	0.047	HOMO $\rightarrow$ LUMO+3
		6-31+G(d) (Spherical)		
1	3.090	401	0.755	HOMO $\rightarrow$ LUMO
2	3.936	315	0.075	HOMO–1 $\rightarrow$ LUMO
3	4.002	310	0.006	HOMO $\rightarrow$ LUMO+1
4	4.046	306	0.006	HOMO–2 $\rightarrow$ LUMO
5	4.218	294	0.047	HOMO $\rightarrow$ LUMO+3

Table C.1 Continued

	6-31+G(d,p) (Cartesian)			
1	3.089	401	0.754	HOMO → LUMO
2	3.935	315	0.076	HOMO–1 → LUMO
3	3.999	310	0.006	HOMO → LUMO+1
4	4.043	307	0.005	HOMO–2 → LUMO
5	4.211	294	0.048	HOMO → LUMO+3
	6-31+G(d,p) (Spherical)			
1	3.088	402	0.755	HOMO → LUMO
2	3.934	315	0.076	HOMO–1 → LUMO
3	3.999	310	0.006	HOMO → LUMO+1
4	4.042	307	0.005	HOMO–2 → LUMO
5	4.211	294	0.048	HOMO → LUMO+3
	jun-cc-pV(D+d)Z			
1	3.084	402	0.752	HOMO → LUMO+1
2	3.768	329	0.005	HOMO → LUMO
3	3.928	316	0.078	HOMO–1 → LUMO+1
4	4.029	308	0.003	HOMO–2 → LUMO+1
5	4.084	304	0.000	HOMO → LUMO+4
	jun-cc-pV(T+d)Z			
1	3.083	402	0.734	HOMO → LUMO+1
2	3.510	353	0.008	HOMO → LUMO
3	3.809	326	0.000	HOMO → LUMO+3
4	3.934	315	0.006	HOMO → LUMO+4
5	3.955	313	0.066	HOMO–1 → LUMO+1
	aug-cc-pVDZ			
1	3.080	403	0.716	HOMO → LUMO+2
2	3.259	380	0.024	HOMO → LUMO
3	3.562	348	0.000	HOMO → LUMO+4
4	3.683	337	0.003	HOMO → LUMO+3
5	3.750	331	0.002	HOMO → LUMO+1
	aug-cc-pVTZ			
1	3.063	405	0.409	HOMO → LUMO+5
2	3.111	399	0.330	HOMO → LUMO
3	3.389	366	0.001	HOMO → LUMO+3
4	3.504	354	0.001	HOMO → LUMO+2
5	3.540	350	0.002	HOMO → LUMO+1
	pcseg-1			
1	3.143	394	0.708	HOMO → LUMO
2	3.808	326	0.084	HOMO–1 → LUMO
3	4.069	305	0.002	HOMO–3 → LUMO
4	4.173	297	0.000	HOMO–6 → LUMO
5	4.279	290	0.050	HOMO → LUMO+1
	pcseg-2			
1	3.104	399	0.728	HOMO → LUMO
2	3.945	314	0.073	HOMO–1 → LUMO
3	4.047	306	0.003	HOMO–2 → LUMO
4	4.226	293	0.002	HOMO–6 → LUMO
5	4.243	292	0.068	HOMO → LUMO+2
	aug-pcseg-1			
1	3.064	405	0.410	HOMO → LUMO+5
2	3.111	399	0.336	HOMO → LUMO

Table C.1 Continued

3	3.395	365	0.001	HOMO → LUMO+3
4	3.501	354	0.002	HOMO → LUMO+2
5	3.543	350	0.002	HOMO → LUMO+1
aug-pcseg-2				
1	2.912	426	0.006	HOMO → LUMO
2	3.083	402	0.725	HOMO → LUMO+9
3	3.185	389	0.009	HOMO → LUMO+1
4	3.285	377	0.001	HOMO → LUMO+2
5	3.297	376	0.002	HOMO → LUMO+4

Table C.2. For Fluorescein^{2-}, Obtained From the Work by Gerasimova et al.,[222] the Vertical Excitation energies (VEE), Oscillator Strengths (OS), and Molecular Orbitals (MOs) Involved in the First Five Electronic Transitions (S_n) Computed in Gas-phase Using the CAM-B3LYP Functional and Different Basis Sets.

S_n	VEE (eV)	VEE (nm)	OS	Transition MOs
Fluorescein 2-[222]				
6-31+G(d) (Cartesian)				
1	3.026	410	0.755	HOMO → LUMO
2	3.750	331	0.068	HOMO–1 → LUMO
3	3.929	316	0.003	HOMO–1 → LUMO+3
4	3.947	314	0.002	HOMO–2 → LUMO
5	3.994	310	0.004	HOMO → LUMO+1
6-31+G(d) (Spherical)				
1	3.025	410	0.755	HOMO → LUMO
2	3.748	331	0.068	HOMO–1 → LUMO
3	3.927	316	0.003	HOMO–1 → LUMO+3
4	3.946	314	0.002	HOMO–2 → LUMO
5	3.993	311	0.004	HOMO → LUMO+1
6-31+G(d,p) (Cartesian)				
1	3.024	410	0.754	HOMO → LUMO
2	3.747	331	0.067	HOMO–1 → LUMO
3	3.924	316	0.003	HOMO–1 → LUMO+3
4	3.944	314	0.002	HOMO–2 → LUMO
5	3.989	311	0.004	HOMO → LUMO+1
6-31+G(d,p) (Spherical)				
1	3.024	410	0.755	HOMO → LUMO
2	3.745	331	0.067	HOMO–1 → LUMO
3	3.922	316	0.003	HOMO–1 → LUMO+3
4	3.944	314	0.002	HOMO–2 → LUMO
5	3.989	311	0.004	HOMO → LUMO+1
jun-cc-pV(D+d)Z				
1	3.020	411	0.752	HOMO → LUMO
2	3.736	332	0.066	HOMO–1 → LUMO
3	3.775	328	0.004	HOMO → LUMO+1
4	3.906	317	0.003	HOMO–1 → LUMO+3
5	3.936	315	0.003	HOMO–2 → LUMO
jun-cc-pV(T+d)Z				
1	3.019	411	0.735	HOMO → LUMO+1
2	3.519	352	0.005	HOMO → LUMO

Table C.2 Continued

3	3.775	328	0.064	HOMO–1 → LUMO+1
4	3.839	323	0.000	HOMO → LUMO+3
5	3.931	315	0.006	HOMO–2 → LUMO+1
aug-cc-pVDZ				
1	3.018	411	0.728	HOMO → LUMO+2
2	3.265	380	0.011	HOMO → LUMO
3	3.591	345	0.000	HOMO → LUMO+4
4	3.707	334	0.002	HOMO → LUMO+3
5	3.766	329	0.002	HOMO → LUMO+1
aug-cc-pVTZ				
1	3.014	411	0.692	HOMO → LUMO+5
2	3.108	399	0.047	HOMO → LUMO
3	3.418	363	0.001	HOMO → LUMO+3
4	3.525	352	0.001	HOMO → LUMO+2
5	3.562	348	0.001	HOMO → LUMO+1
pcseg-1				
1	3.074	403	0.704	HOMO → LUMO
2	3.551	349	0.059	HOMO–1 → LUMO
3	3.786	327	0.002	HOMO–1 → LUMO+1
4	3.984	311	0.004	HOMO–4 → LUMO
5	3.998	310	0.002	HOMO–5 → LUMO
pcseg-2				
1	3.039	408	0.727	HOMO → LUMO
2	3.750	331	0.062	HOMO–1 → LUMO
3	3.927	316	0.003	HOMO–1 → LUMO+1
4	3.953	314	0.003	HOMO–2 → LUMO
5	4.056	306	0.000	HOMO–5 → LUMO
aug-pcseg-1				
1	3.014	411	0.690	HOMO → LUMO+5
2	3.100	400	0.054	HOMO → LUMO
3	3.421	362	0.001	HOMO → LUMO+1
4	3.527	352	0.002	HOMO → LUMO+2
5	3.564	348	0.001	HOMO → LUMO+1
aug-pcseg-2				
1	2.924	424	0.020	HOMO → LUMO
2	3.020	411	0.718	HOMO → LUMO+9
3	3.206	387	0.003	HOMO → LUMO+1
4	3.316	374	0.001	HOMO → LUMO+2
5	3.321	373	0.001	HOMO → LUMO+4

Table C.3. For Serotonin (Neutral) the Vertical Excitation Energies (VEE), Oscillator Strengths (OS), and Molecular Orbitals (MOs) Involved in the First Five Electronic Transitions (S_n) Computed in Gas-phase Using the CAM-B3LYP Functional and the aug-cc-pVDZ Basis Set

Serotonin neutral				
S_n	**VEE (eV)**	**VEE (nm)**	**OS**	**Transition MOs**
6-31+G(d) (Cartesian)				
1	4.668	266	0.057	HOMO → LUMO
2	4.970	249	0.084	HOMO–1 → LUMO
3	5.108	243	0.001	HOMO → LUMO+1

Table C.3 Continued

4	5.167	240	0.001	HOMO → LUMO+2
5	5.419	229	0.002	HOMO–1 → LUMO+1

6-31+G(d) (Spherical)				
1	4.668	266	0.057	HOMO → LUMO
2	4.970	249	0.084	HOMO–1 → LUMO
3	5.108	243	0.001	HOMO → LUMO+1
4	5.167	240	0.001	HOMO → LUMO+2
5	5.420	229	0.002	HOMO–1 → LUMO+1

6-31+G(d,p) (Cartesian)				
1	4.657	266	0.058	HOMO → LUMO
2	4.958	250	0.085	HOMO–1 → LUMO
3	5.115	242	0.001	HOMO → LUMO+1
4	5.174	240	0.001	HOMO → LUMO+2
5	5.427	228	0.002	HOMO–1 → LUMO+1

6-31+G(d,p) (Spherical)				
1	4.657	266	0.058	HOMO → LUMO
2	4.959	250	0.085	HOMO–1 → LUMO
3	5.116	242	0.001	HOMO → LUMO+1
4	5.175	240	0.001	HOMO → LUMO+2
5	5.428	228	0.002	HOMO–1 → LUMO+1

jun-cc-pV(D+d)Z				
1	4.626	268	0.055	HOMO → LUMO
2	4.926	252	0.082	HOMO–1 → LUMO
3	5.102	243	0.001	HOMO → LUMO+1
4	5.130	242	0.001	HOMO → LUMO+1
5	5.377	231	0.002	HOMO–1 → LUMO+1

aug-cc-pVDZ				
1	4.580	271	0.051	HOMO → LUMO+2
2	4.888	254	0.078	HOMO–1 → LUMO+2
3	4.954	250	0.002	HOMO → LUMO
4	5.016	247	0.001	HOMO → LUMO+1
5	5.238	237	0.002	HOMO–1 → LUMO

aug-cc-pVTZ				
1	4.578	271	0.051	HOMO → LUMO+2
2	4.880	254	0.077	HOMO–1 → LUMO+2
3	4.961	250	0.002	HOMO → LUMO
4	5.028	247	0.001	HOMO → LUMO+1
5	5.236	237	0.002	HOMO–1 → LUMO

aug-pcseg-1				
1	4.605	269	0.052	HOMO → LUMO+2
2	4.913	252	0.076	HOMO–1 → LUMO+2
3	4.961	250	0.004	HOMO → LUMO
4	5.020	247	0.001	HOMO → LUMO+1
5	5.244	236	0.002	HOMO–1 → LUMO

Table C.4. For Serotonin$^+$, the Vertical Excitation Energies (VEE), Oscillator Strengths (OS), and Molecular Orbitals (MOs) Involved in the First Five Electronic Transitions (S$_n$) Computed in Gas-phase Using the CAM-B3LYP Functional and Different Basis Sets

S_n	VEE (eV)	VEE (nm)	OS	Transition MOs
		Serotonin 1+		
		6-31+G(d) (Cartesian)		
1	4.461	278	0.007	HOMO → LUMO
2	4.664	266	0.083	HOMO → LUMO+1
3	4.849	256	0.015	HOMO−1 → LUMO
4	5.155	241	0.066	HOMO−1 → LUMO+1
5	5.681	218	0.011	HOMO → LUMO+2
		6-31+G(d) (Spherical)		
1	4.460	278	0.007	HOMO → LUMO
2	4.665	266	0.083	HOMO → LUMO+1
3	4.849	256	0.015	HOMO−1 → LUMO
4	5.155	241	0.066	HOMO−1 → LUMO+1
5	5.681	218	0.011	HOMO → LUMO+2
		6-31+G(d,p) (Cartesian)		
1	4.458	278	0.008	HOMO → LUMO
2	4.656	266	0.084	HOMO → LUMO+1
3	4.843	256	0.015	HOMO−1 → LUMO
4	5.145	241	0.067	HOMO−1 → LUMO+1
5	5.679	218	0.011	HOMO → LUMO+2
		6-31+G(d,p) (Spherical)		
1	4.460	278	0.008	HOMO → LUMO
2	4.657	266	0.084	HOMO → LUMO+1
3	4.845	256	0.015	HOMO−1 → LUMO
4	5.146	241	0.067	HOMO−1 → LUMO+1
5	5.680	218	0.011	HOMO → LUMO+2
		jun-cc-pV(D+d)Z		
1	4.445	279	0.010	HOMO → LUMO
2	4.632	268	0.080	HOMO → LUMO+1
3	4.822	257	0.015	HOMO−1 → LUMO
4	5.115	242	0.064	HOMO−1 → LUMO+1
5	5.636	220	0.008	HOMO → LUMO+2
		aug-cc-pVDZ		
1	4.407	281	0.010	HOMO → LUMO
2	4.592	270	0.074	HOMO → LUMO+1
3	4.778	259	0.014	HOMO−1 → LUMO
4	5.076	244	0.062	HOMO−1 → LUMO+1
5	5.585	222	0.007	HOMO → LUMO+2
		aug-cc-pVTZ		
1	4.428	280	0.014	HOMO → LUMO
2	4.597	270	0.071	HOMO → LUMO+1
3	4.795	259	0.015	HOMO−1 → LUMO
4	5.072	244	0.060	HOMO−1 → LUMO+1
5	5.605	221	0.006	HOMO → LUMO+2
		aug-pcseg-1		
1	4.431	280	0.010	HOMO → LUMO
2	4.618	268	0.075	HOMO → LUMO+1
3	4.805	258	0.014	HOMO−1 → LUMO
4	5.104	243	0.063	HOMO−1 → LUMO+1

Table C.4 Continued

5	5.607	221	0.007	HOMO $\rightarrow$ LUMO+2

Table C.5. For Rhodamine 6G^{+}, Optimized Using B3LYP/6-31+G(d,p) in Vacuum, the Vertical Excitation Energies (VEE), Oscillator Strengths (OS), and Molecular Orbitals (MOs) Involved in the Five Lowest Energy Electronic Transitions (S$_n$) Computed in Gas-phase Using the CAM-B3LYP Functional and the 6-31+G(d,p) and aug-cc-pVDZ Basis Sets

Rhodamine 6G 1+ (B3LYP/6-31+G(d,p))				
S$_n$	**VEE (eV)**	**VEE (nm)**	**OS**	**Transition MOs**
6-31+G(d,p) (Spherical)				
1	3.060	405	0.863	HOMO $\rightarrow$ LUMO
2	3.668	338	0.000	HOMO–1 $\rightarrow$ LUMO
3	4.407	281	0.245	HOMO–2 $\rightarrow$ LUMO
4	4.456	278	0.001	HOMO–3 $\rightarrow$ LUMO
5	4.530	274	0.001	HOMO–5 $\rightarrow$ LUMO
aug-cc-pVDZ				
1	3.049	407	0.859	HOMO $\rightarrow$ LUMO
2	3.661	339	0.000	HOMO–1 $\rightarrow$ LUMO
3	4.398	282	0.229	HOMO–2 $\rightarrow$ LUMO
4	4.457	278	0.001	HOMO–3 $\rightarrow$ LUMO
5	4.523	274	0.001	HOMO–5 $\rightarrow$ LUMO

Table C.6. For Rhodamine 6G^{+}, Optimized Using the CAM-B3LYP/6-31+G(d,p) Method, the Vertical Excitation Energies (VEE), Oscillator Strengths (OS), and Molecular Orbitals (MOs) Involved in the Five Lowest Energy Electronic Transitions (S$_n$) Computed in Gas-phase Using the CAM-B3LYP Functional and the 6-31+G(d,p) and aug-cc-pVDZ Basis Sets

Rhodamine 6G 1+ (CAM-B3LYP/6-31+G(d,p))				
S$_n$	**VEE (eV)**	**VEE (nm)**	**OS**	**Transition MOs**
6-31+G(d,p) (Spherical)				
1	3.103	400	0.861	HOMO $\rightarrow$ LUMO
2	3.720	333	0.000	HOMO–1 $\rightarrow$ LUMO
3	4.454	278	0.236	HOMO–2 $\rightarrow$ LUMO
4	4.520	274	0.001	HOMO–3 $\rightarrow$ LUMO
5	4.593	270	0.001	HOMO–5 $\rightarrow$ LUMO
aug-cc-pVDZ				
1	3.092	401	0.858	HOMO $\rightarrow$ LUMO
2	3.714	334	0.000	HOMO–1 $\rightarrow$ LUMO
3	4.444	279	0.219	HOMO–2 $\rightarrow$ LUMO
4	4.521	274	0.001	HOMO–3 $\rightarrow$ LUMO
5	4.587	270	0.001	HOMO–5 $\rightarrow$ LUMO

C.4 Three-photon Absorption Data

C.4.1 Dyes

Table C.7. For Fluorescein^{2-}, the Excitation Energies, 3PA Components δ_f and δ_g, 3PA Transition Moment δ^{3PA}, and σ^{3PA}. Results in This Table Were Obtained Using the CAM-B3LYP Functional an Different Basis Sets

S_n	ω (eV)	ω (nm)	δ_f (au)	δ_g (au)	δ^{3PA} (au)	σ^{3PA} (cm^6 s^2 photon^{-2})
			Fluorescein 2-			
			6-31+G(d) (Cartesian)			
1	3.09	401	7.09E+05	6.50E+06	4.32E+05	1.33E-84
2	3.94	315	1.60E+08	1.41E+08	2.18E+07	1.38E-82
3	4.00	310	1.01E+07	3.21E+07	2.70E+06	1.78E-83
4	4.05	306	2.04E+07	1.29E+08	9.14E+06	6.29E-83
5	4.22	294	1.09E+08	1.88E+08	2.00E+07	1.55E-82
			6-31+G(d) (Spherical)			
1	3.09	401	7.16E+05	6.51E+06	4.34E+05	1.34E-84
2	3.94	315	1.61E+08	1.41E+08	2.18E+07	1.38E-82
3	4.00	310	1.04E+07	3.31E+07	2.78E+06	1.84E-83
4	4.05	306	2.05E+07	1.30E+08	9.18E+06	6.32E-83
5	4.22	294	1.09E+08	1.89E+08	2.01E+07	1.56E-82
			6-31+G(d,p) (Cartesian)			
1	3.09	401	6.61E+05	6.54E+06	4.30E+05	1.33E-84
2	3.94	315	1.61E+08	1.42E+08	2.19E+07	1.39E-82
3	4.00	310	9.95E+06	3.19E+07	2.68E+06	1.77E-83
4	4.04	307	2.19E+07	1.27E+08	9.13E+06	6.16E-83
5	4.21	294	1.11E+08	1.93E+08	2.05E+07	1.59E-82
			6-31+G(d,p) (Spherical)			
1	3.09	401	6.83E+05	6.58E+06	4.34E+05	1.34E-84
2	3.93	315	1.61E+08	1.42E+08	2.19E+07	1.36E-82
3	4.00	310	1.00E+07	3.21E+07	2.70E+06	1.78E-83
4	4.04	307	2.21E+07	1.28E+08	9.19E+06	6.20E-83
5	4.21	294	1.12E+08	1.94E+08	2.07E+07	1.60E-82
			jun-ccpV(D+d)Z			
1	3.08	403	4.45E+05	5.83E+06	3.72E+05	1.12E-84
2	3.77	329	2.58E+06	1.35E+07	9.90E+05	5.53E-84
3	3.93	315	1.58E+08	1.40E+08	2.16E+07	1.34E-82
4	4.03	308	2.46E+07	1.15E+08	8.66E+06	5.84E-83
5	4.08	304	9.63E+06	1.69E+07	1.79E+06	1.26E-83
			jun-ccpV(T+d)Z			
1	3.08	403	2.82E+04	4.80E+06	2.77E+05	8.31E-85
2	3.51	353	1.48E+05	7.37E+06	4.34E+05	1.94E-84
3	3.81	325	7.53E+06	1.33E+07	1.40E+06	7.99E-84
4	3.93	315	6.39E+07	5.39E+07	8.56E+06	5.32E-83
5	3.95	314	1.50E+08	1.25E+08	2.00E+07	1.27E-82
			aug-cc-pVDZ			
1	3.08	403	7.62E+04	5.69E+06	3.32E+05	9.97E-85
2	3.26	380	2.75E+05	6.12E+06	3.74E+05	1.34E-84
3	3.56	348	1.21E+07	9.85E+06	1.60E+06	7.48E-84
4	3.68	337	9.01E+07	1.18E+08	1.45E+07	7.42E-83
5	3.75	331	1.87E+05	4.27E+06	2.60E+05	1.42E-84
			pcseg-1			
1	3.14	395	2.07E+06	5.82E+06	5.10E+05	1.61E-84

177

Table C.7 Continued

2	3.81	325	1.63E+08	1.87E+08	2.46E+07	1.40E-82
3	4.07	305	2.59E+07	1.18E+08	8.96E+06	6.29E-83
4	4.17	297	1.67E+05	4.89E+05	4.22E+04	3.14E-85
5	4.28	290	8.83E+07	1.43E+08	1.58E+07	1.27E-82
			pcseg-2			
1	3.1	400	1.34E+06	5.96E+06	4.55E+05	1.40E-84
2	3.94	315	1.42E+08	1.28E+08	1.94E+07	1.23E-82
3	4.05	306	1.81E+07	1.11E+08	7.87E+06	5.42E-83
4	4.23	293	4.01E+06	5.65E+06	6.66E+05	5.16E-84
5	4.24	292	1.05E+08	1.91E+08	1.99E+07	1.57E-82
			aug-pcseg-1			
1	3.06	405	1.22E+06	8.36E+06	5.82E+05	1.70E-84
2	3.11	399	1.13E+06	3.13E+06	2.75E+05	8.48E-85
3	3.40	365	2.11E+07	1.88E+07	2.88E+06	1.17E-83
4	3.50	354	1.81E+08	2.23E+08	2.82E+07	1.26E-82
5	3.54	350	7.10E+03	6.53E+06	3.74E+05	1.71E-84

Table C.8. For Fluorescein^{2-},[222] the Excitation Energies, 3PA Components δ_f and δ_g, 3PA Transition Moments δ^{3PA}, and σ^{3PA} Obtained Using the CAM-B3LYP Functional an Different Basis Sets

			Fluorescein 2-[222]			
S_n	ω (eV)	ω (nm)	δ_f (au)	δ_g (au)	δ^{3PA} (au)	σ^{3PA} (cm^6 s^2 photon^{-2})
			6-31+G(d) (Cartesian)			
1	3.03	409	1.80E+06	9.16E+06	6.78E+05	1.93E-84
2	3.75	331	1.86E+08	2.07E+08	2.78E+07	1.52E-82
3	3.93	315	2.15E+07	4.81E+07	4.59E+06	2.85E-83
4	3.95	314	2.51E+07	1.15E+08	8.72E+06	5.53E-83
5	3.99	311	1.30E+07	4.40E+07	3.63E+06	2.40E-83
			6-31+G(d) (Spherical)			
1	3.03	409	1.84E+06	9.22E+06	6.85E+05	1.95E-84
2	3.75	331	1.87E+08	2.08E+08	2.79E+07	1.53E-82
3	3.93	315	2.01E+07	4.20E+07	4.12E+06	2.56E-83
4	3.95	314	2.68E+07	1.21E+08	9.23E+06	5.85E-83
5	3.99	311	1.32E+07	4.45E+07	3.67E+06	2.43E-83
			6-31+G(d,p) (Cartesian)			
1	3.02	411	1.77E+06	9.25E+06	6.80E+05	1.93E-84
2	3.74	332	1.85E+08	2.07E+08	2.76E+07	1.48E-82
3	3.92	316	1.69E+07	2.72E+07	3.00E+06	1.86E-83
4	3.94	315	3.44E+07	1.37E+08	1.08E+07	6.85E-83
5	3.99	311	1.30E+07	4.46E+07	3.66E+06	2.42E-83
			6-31+G(d,p) (Spherical)			
1	3.02	411	1.77E+06	9.25E+06	6.80E+05	1.93E-84
2	3.74	332	1.85E+08	2.07E+08	2.76E+07	1.48E-82
3	3.92	316	1.69E+07	2.72E+07	3.00E+06	1.86E-83
4	3.94	315	3.44E+07	1.37E+08	1.08E+07	6.85E-83
5	3.99	311	1.30E+07	4.46E+07	3.66E+06	2.42E-83
			jun-cc-pV(D+d)Z			
1	3.02	411	1.40E+06	8.31E+06	5.95E+05	1.69E-84
2	3.74	332	1.81E+08	2.07E+08	2.73E+07	1.46E-82
3	3.78	328	3.91E+06	1.75E+07	1.34E+06	7.49E-84

Table C.8 Continued

4	3.91	317	1.31E+07	1.37E+07	1.91E+06	1.19E-83
5	3.94	315	4.25E+07	1.37E+08	1.15E+07	7.29E-83
aug-cc-pVDZ						
1	3.02	411	6.03E+05	7.37E+06	4.73E+05	1.35E-84
2	3.26	380	1.62E+06	7.94E+06	5.93E+05	2.13E-84
3	3.59	345	1.36E+07	1.23E+07	1.87E+06	8.95E-84
4	3.71	334	7.17E+07	9.71E+07	1.17E+07	6.12E-83
5	3.77	329	2.84E+06	1.04E+07	8.37E+05	4.68E-84
pcseg-1						
1	3.07	404	3.37E+06	8.03E+06	7.48E+05	2.25E-84
2	3.55	349	2.10E+08	2.69E+08	3.34E+07	1.53E-82
3	3.79	327	9.25E+06	1.10E+07	1.42E+06	7.93E-84
4	3.98	312	2.57E+07	1.01E+08	7.94E+06	5.14E-83
5	4.00	310	7.35E+06	2.56E+07	2.09E+06	1.38E-83
pcseg-2						
1	3.04	408	2.48E+06	8.31E+06	6.88E+05	2.01E-84
2	3.75	331	1.57E+08	1.80E+08	2.38E+07	1.30E-82
3	3.93	315	1.28E+07	1.47E+07	1.94E+06	1.21E-83
4	3.95	314	3.53E+07	1.29E+08	1.04E+07	6.60E-83
5	4.06	305	3.66E+05	4.31E+05	5.60E+04	3.85E-85
aug-pcseg-1						
1	3.01	412	9.23E+05	8.43E+06	5.61E+05	1.60E-84
2	3.1	400	4.01E+06	6.70E+06	7.27E+05	2.24E-84
3	3.42	363	1.97E+07	2.02E+07	2.85E+06	1.19E-83
4	3.53	351	1.48E+08	1.79E+08	2.29E+07	1.05E-82
5	3.56	348	3.11E+06	1.02E+07	8.49E+05	3.97E-84

Table C.9. For Serotonin (Neutral) the Excitation Energies, 3PA Components δ_f and δ_g, 3PA Transition Moments δ^{3PA}, and σ^{3PA} Obtained Using the CAM-B3LYP Functional and Different Basis Sets

Serotonin neutral						
S_n	ω (eV)	ω (nm)	δ_f (au)	δ_g (au)	δ^{3PA} (au)	σ^{3PA} (cm^6 s^2 photon^{-2})
6-31+G(d) (Cartesian)						
1	4.67	265	4.35E+05	4.24E+05	6.15E+04	6.51E-85
2	4.97	249	4.57E+06	4.20E+06	6.31E+05	8.04E-84
3	5.11	243	4.18E+06	3.19E+06	5.40E+05	7.46E-84
4	5.17	240	2.41E+06	1.83E+06	3.11E+05	4.44E-84
5	5.42	229	5.94E+06	1.17E+07	1.18E+06	1.93E-83
6-31+G(d) (Spherical)						
1	4.67	265	4.36E+05	4.25E+05	6.17E+04	6.53E-85
2	4.97	249	4.57E+06	4.20E+06	6.32E+05	8.06E-84
3	5.11	243	4.16E+06	3.17E+06	5.38E+05	7.44E-84
4	5.17	240	2.44E+06	1.84E+06	3.14E+05	4.48E-84
5	5.42	229	5.94E+06	1.17E+07	1.18E+06	1.93E-83
6-31+G(d,p) (Cartesian)						
1	4.66	266	4.38E+05	4.23E+05	6.17E+04	6.42E-85
2	4.96	250	4.63E+06	4.25E+06	6.40E+05	8.03E-84
3	5.12	242	4.29E+06	3.26E+06	5.54E+05	7.66E-84
4	5.17	240	2.44E+06	1.84E+06	3.15E+05	4.49E-84
5	5.43	228	5.98E+06	1.18E+07	1.19E+06	1.98E-83

Table C.9 Continued

			6-31+G(d,p) (Spherical)			
1	4.66	266	4.39E+05	4.24E+05	6.18E+04	6.43E-85
2	4.96	250	4.64E+06	4.25E+06	6.40E+05	8.03E-84
3	5.12	242	4.29E+06	3.26E+06	5.54E+05	7.66E-84
4	5.18	239	2.47E+06	1.86E+06	3.18E+05	4.54E-84
5	5.43	228	5.99E+06	1.18E+07	1.19E+06	1.98E-83
			jun-cc-pV(D+d)Z			
1	4.63	268	3.77E+05	3.89E+05	5.46E+04	5.58E-85
2	4.93	251	4.59E+06	4.18E+06	6.33E+05	7.81E-84
3	5.10	243	3.53E+06	2.75E+06	4.60E+05	6.26E-84
4	5.13	242	3.30E+06	2.42E+06	4.22E+05	5.93E-84
5	5.38	230	7.93E+06	1.50E+07	1.54E+06	2.49E-83
			aug-cc-pVDZ			
1	4.58	271	4.05E+05	3.93E+05	5.72E+04	5.64E-85
2	4.89	254	4.47E+06	4.10E+06	6.18E+05	7.50E-84
3	4.95	250	5.50E+06	4.27E+06	7.16E+05	8.98E-84
4	5.02	247	3.92E+06	3.19E+06	5.18E+05	6.71E-84
5	5.24	237	1.01E+07	1.52E+07	1.74E+06	2.60E-83
			aug-cc-pVTZ			
1	4.58	271	3.61E+05	3.78E+05	5.26E+04	5.19E-85
2	4.88	254	4.36E+06	4.00E+06	6.02E+05	7.18E-84
3	4.96	250	5.59E+06	4.65E+06	7.44E+05	9.33E-84
4	5.03	246	3.65E+06	3.19E+06	4.95E+05	6.52E-84
5	5.24	237	1.13E+07	1.60E+07	1.88E+06	2.81E-83
			aug-pcseg-1			
1	4.61	269	4.20E+05	4.23E+05	6.02E+04	6.04E-85
2	4.91	253	4.97E+06	4.54E+06	6.85E+05	8.31E-84
3	4.96	250	5.20E+06	4.16E+06	6.83E+05	8.57E-84
4	5.02	247	4.63E+06	3.68E+06	6.07E+05	7.87E-84
5	5.24	237	1.22E+07	1.76E+07	2.06E+06	3.08E-83

Table C.10. For Serotonin$^+$, the Excitation Energies, 3PA Components δ_f and δ_g, 3PA Transition Moments δ^{3PA}, and σ^{3PA} Obtained Using the CAM-B3LYP Functional and Different Basis Sets

			Serotonin 1+			
S_n	ω (eV)	ω (nm)	δ_f (au)	δ_g (au)	δ^{3PA} (au)	σ^{3PA} (cm^6 s^2 photon^{-2})
			6-31+G(d) (Cartesian)			
1	4.46	278	1.62E+06	2.92E+06	3.05E+05	2.80E-84
2	4.66	266	7.50E+06	5.38E+06	9.50E+05	9.88E-84
3	4.85	256	1.54E+07	1.32E+07	2.07E+06	2.43E-83
4	5.16	240	4.12E+06	4.36E+06	6.03E+05	8.60E-84
5	5.68	218	2.36E+07	1.65E+07	2.96E+06	5.62E-83
			6-31+G(d) (Spherical)			
1	4.46	278	1.62E+06	2.92E+06	3.06E+05	2.81E-84
2	4.66	266	7.48E+06	5.36E+06	9.47E+05	9.85E-84
3	4.85	256	1.54E+07	1.32E+07	2.07E+06	2.43E-83
4	5.16	240	4.13E+06	4.36E+06	6.03E+05	8.60E-84
5	5.68	218	2.36E+07	1.65E+07	2.97E+06	5.64E-83
			6-31+G(d,p) (Cartesian)			
1	4.46	278	1.49E+06	2.79E+06	2.87E+05	2.63E-84

Table C.10 Continued

2	4.66	266	7.39E+06	5.30E+06	9.37E+05	9.75E-84
3	4.85	256	1.52E+07	1.31E+07	2.05E+06	2.41E-83
4	5.15	241	4.12E+06	4.37E+06	6.03E+05	8.47E-84
5	5.68	218	2.37E+07	1.65E+07	2.98E+06	5.66E-83
6-31+G(d,p) (Spherical)						
1	4.46	278	1.50E+06	2.79E+06	2.88E+05	2.64E-84
2	4.66	266	7.41E+06	5.31E+06	9.39E+05	9.77E-84
3	4.84	256	1.51E+07	1.30E+07	2.03E+06	2.38E-83
4	5.15	241	4.07E+06	4.33E+06	5.96E+05	8.37E-84
5	5.68	218	2.38E+07	1.66E+07	2.99E+06	5.68E-83
jun-cc-pV(D+d)Z						
1	4.45	279	1.49E+06	2.70E+06	2.82E+05	2.59E-84
2	4.63	268	7.53E+06	5.40E+06	9.55E+05	9.76E-84
3	4.82	257	1.71E+07	1.44E+07	2.29E+06	2.64E-83
4	5.11	243	4.17E+06	4.37E+06	6.07E+05	8.39E-84
5	5.64	220	2.15E+07	1.49E+07	2.69E+06	4.96E-83
aug-cc-pVDZ						
1	4.41	281	9.44E+05	2.06E+06	1.98E+05	1.75E-84
2	4.59	270	6.39E+06	4.50E+06	8.05E+05	8.08E-84
3	4.78	259	1.83E+07	1.52E+07	2.44E+06	2.77E-83
4	5.08	244	3.91E+06	4.10E+06	5.69E+05	7.74E-84
5	5.59	222	2.04E+07	1.42E+07	2.56E+06	4.59E-83
aug-cc-pVTZ						
1	4.43	280	7.24E+05	1.75E+06	1.62E+05	1.46E-84
2	4.60	270	5.86E+06	4.13E+06	7.38E+05	7.41E-84
3	4.80	258	1.83E+07	1.53E+07	2.44E+06	2.77E-83
4	5.07	245	4.07E+06	4.22E+06	5.89E+05	7.88E-84
5	5.61	221	1.95E+07	1.35E+07	2.44E+06	4.44E-83
aug-pcseg-1						
1	4.43	280	1.04E+06	2.19E+06	2.14E+05	1.93E-84
2	4.62	268	6.61E+06	4.66E+06	8.32E+05	8.50E-84
3	4.81	258	1.79E+07	1.49E+07	2.38E+06	2.75E-83
4	5.10	243	4.00E+06	4.16E+06	5.80E+05	7.89E-84
5	5.61	221	2.07E+07	1.44E+07	2.60E+06	4.73E-83

Table C.11. For Rhodamine 6G$^+$, Optimized Using the B3LYP/6-31+G(d,p) Method, the Vertical Excitation Energies, 3PA Components δ_f and δ_g, 3PA Transition Moments δ^{3PA}, and σ^{3PA} Obtained Using the CAM-B3LYP Functional and Different Basis Sets

S_n	ω (eV)	ω (nm)	δ_f (au)	δ_g (au)	δ^{3PA} (au)	σ^{3PA} (cm^6 s^2 photon^{-2})
			Rhodamine 6G 1+ [B3LYP/6-31+G(d,p)]			
			6-31+G(d,p) (Spherical)			
1	3.06	405	3.04E+07	3.09E+07	4.37E+06	1.28E-83
2	3.67	338	6.81E+07	1.66E+08	1.53E+07	7.83E-83
3	4.41	281	3.09E+09	7.56E+09	6.97E+08	6.16E-81
4	4.46	278	1.60E+07	4.75E+07	4.09E+06	3.75E-83
5	4.53	274	4.74E+08	1.08E+09	1.02E+08	9.71E-82
			aug-cc-pVDZ			
1	3.05	407	2.98E+07	2.86E+07	4.19E+06	1.22E-83
2	3.66	339	5.87E+07	1.45E+08	1.33E+07	6.81E-83
3	4.40	282	3.06E+09	7.49E+09	6.90E+08	6.10E-81
4	4.46	278	1.29E+07	3.97E+07	3.38E+06	3.10E-83
5	4.52	274	6.30E+08	1.54E+09	1.42E+08	1.35E-81

Table C.12. For Rhodamine 6G$^+$ Optimized Using the CAM-B3LYP/6-31+G(d,p), the Excitation Energies, 3PA Components δ_f and δ_g, 3PA Transition Moments δ^{3PA}, and σ^{3PA} Obtained Using the CAM-B3LYP Functional and Different Basis Sets

S_n	ω (eV)	ω (nm)	δ_f (au)	δ_g (au)	δ^{3PA} (au)	σ^{3PA} (cm^6 s^2 photon^{-2})
			Rhodamine 6G 1+ [CAM-B3LYP/6-31+G(d,p)]			
			6-31+G(d,p) (Spherical)			
1	3.10	400	3.38E+07	3.19E+07	4.72E+06	1.45E-83
2	3.72	333	6.84E+07	1.68E+08	1.54E+07	8.24E-83
3	4.45	279	2.49E+09	5.99E+09	5.56E+08	5.10E-81
4	4.52	274	1.38E+07	4.39E+07	3.69E+06	3.51E-83
5	4.59	270	4.56E+08	1.02E+09	9.71E+07	9.75E-82
			aug-cc-pVDZ			
1	3.09	401	3.32E+07	2.98E+07	4.55E+06	1.40E-83
2	3.71	334	5.91E+07	1.46E+08	1.34E+07	7.01E-83
3	4.44	279	2.41E+09	5.81E+09	5.39E+08	4.86E-81
4	4.52	274	9.74E+06	3.29E+07	2.71E+06	2.58E-83
5	4.59	270	6.07E+08	1.47E+09	1.36E+08	1.37E-81

C.4.2 Fluorescent Protein Chromophores

Table C.13. Part 1/2. For the Chromophores Previously Studied by Salem and Brown [13] (Figure 4.1 in the Main Text), the Vertical Excitation Energies (ω), δ_f, δ_g, δ^{3PA}, and σ^{3PA} (cm^6 s^2 photon^{-2}). Data was Computed for the Five Lowest Energy Transitions Using the CAM-B3LYP functional and Two Different Basis Sets

	6-31+G(d,p)					aug-cc-pVDZ				
S_n	ω (eV)	δ_f (au)	δ_g (au)	δ^{3PA} (au)	σ^{3PA}	ω (eV)	δ_f (au)	δ_g (au)	δ^{3PA} (au)	σ^{3PA}
				BFP$_1$						
1	3.70	1.4E+07	1.0E+07	1.80E+06	9.42E-84	3.65	1.5E+07	1.1E+07	1.86E+06	9.31E-84
2	4.13	1.2E+03	4.8E+03	3.82E+02	2.79E-87	4.07	9.2E+02	4.6E+03	3.43E+02	2.41E-87
3	5.02	7.2E+07	7.4E+07	1.04E+07	1.35E-82	4.96	8.3E+07	8.3E+07	1.18E+07	1.48E-82
4	5.26	1.2E+05	4.2E+05	3.48E+04	5.20E-85	5.25	1.2E+05	4.1E+05	3.41E+04	5.10E-85
5	5.60	8.9E+07	3.3E+08	2.66E+07	4.84E-82	5.35	4.6E+07	1.3E+08	1.12E+07	1.78E-82
				BFP$_2$						
1	3.96	3.1E+03	1.2E+04	9.72E+02	6.29E-87	3.90	2.6E+07	2.1E+07	3.39E+06	2.06E-83
2	3.96	2.6E+07	2.0E+07	3.33E+06	2.16E-83	3.91	2.7E+03	1.1E+04	8.73E+02	5.42E-87
3	5.15	6.1E+07	6.4E+07	8.83E+06	1.24E-82	5.09	6.8E+07	6.9E+07	9.80E+06	1.33E-82
4	5.35	1.8E+04	6.4E+04	5.19E+03	8.25E-86	5.16	1.5E+05	1.4E+07	7.92E+05	1.13E-83
5	5.38	2.3E+04	9.8E+06	5.62E+05	9.07E-84	5.31	8.8E+03	8.1E+04	5.39E+03	8.31E-86
				BLB						
1	3.15	1.3E+09	1.2E+09	1.83E+08	5.94E-82	3.11	1.2E+09	1.1E+09	1.66E+08	5.12E-82
2	3.55	1.4E+04	5.8E+04	4.51E+03	2.06E-86	3.50	1.8E+04	7.1E+04	5.57E+03	2.49E-86
3	3.60	2.8E+02	2.5E+04	1.44E+03	6.89E-87	3.54	2.3E+02	1.8E+04	1.06E+03	4.84E-87
4	4.22	1.3E+08	1.9E+08	2.22E+07	1.72E-82	4.21	2.1E+08	2.7E+08	3.28E+07	2.54E-82
5	4.53	2.4E+10	2.6E+10	3.50E+09	3.33E-80	4.49	2.4E+10	2.7E+10	3.60E+09	3.36E-80
				CFP						
1	3.54	1.3E+08	1.2E+08	1.74E+07	7.95E-83	3.50	1.3E+08	1.2E+08	1.75E+07	7.81E-83
2	4.10	1.1E+03	4.8E+03	3.66E+02	2.62E-87	4.04	6.8E+02	4.2E+03	3.00E+02	2.02E-87
3	4.58	1.2E+08	1.6E+08	1.93E+07	1.90E-82	4.54	1.7E+08	2.1E+08	2.66E+07	2.58E-82
4	4.79	5.0E+08	5.6E+08	7.46E+07	8.46E-82	4.72	3.7E+08	4.1E+08	5.54E+07	5.97E-82
5	4.96	1.3E+09	1.3E+09	1.89E+08	2.37E-81	4.90	1.4E+09	1.3E+09	1.93E+08	2.34E-81
				GFP$_A$						
1	3.13	7.7E+06	2.7E+07	2.21E+06	6.99E-84	3.10	5.9E+06	2.5E+07	1.94E+06	5.98E-84
2	3.82	1.8E+08	4.2E+08	3.96E+07	2.26E-82	3.32	3.2E+08	7.6E+08	7.12E+07	2.69E-82
3	3.87	1.2E+03	1.9E+03	2.09E+02	1.24E-87	3.78	4.0E+07	1.3E+08	1.06E+07	5.92E-83
4	4.33	8.3E+06	4.8E+07	3.45E+06	2.88E-83	3.84	6.7E+03	5.4E+04	3.67E+03	2.14E-86
5	4.51	2.8E+06	1.0E+08	6.13E+06	5.83E-83	3.92	3.9E+06	2.6E+08	1.50E+07	9.32E-83
				GFP$_N$						
1	3.72	1.1E+08	9.8E+07	1.48E+07	7.92E-83	3.67	1.0E+08	9.3E+07	1.41E+07	7.22E-83
2	4.03	2.5E+03	1.1E+04	8.18E+02	5.52E-87	3.97	1.8E+03	9.2E+03	6.77E+02	4.38E-87
3	4.72	1.8E+07	2.0E+07	2.71E+06	2.92E-83	4.67	1.7E+07	1.8E+07	2.47E+06	2.61E-83
4	4.92	4.3E+08	4.5E+08	6.22E+07	7.67E-82	4.86	4.3E+08	4.4E+08	6.15E+07	7.34E-82
5	5.52	1.8E+06	9.6E+06	7.00E+05	1.22E-83	5.32	5.5E+06	1.7E+07	1.43E+06	2.24E-83

Table C.14. Part 2/2. For the chromophores previously studied by Salem and Brown [13] (Figure 4.1 in the Main Text), the Vertical Excitation Energies (ω), δ_f, δ_g, δ^{3PA}, and σ^{3PA} (cm^6 s^2 $photon^{-2}$). Data was Computed for the Five Lowest Energy Transitions Using the CAM-B3LYP and Two Different Basis Sets

	6-31+G(d,p)					aug-cc-pVDZ				
					KO_A					
1	2.61	2.6E+08	4.3E+08	4.61E+07	8.48E-83	2.58	2.9E+08	4.4E+08	4.99E+07	8.90E-83
2	3.60	5.8E+04	5.9E+04	8.34E+03	3.99E-86	3.58	1.1E+05	1.2E+05	1.64E+04	7.85E-86
3	3.97	2.5E+11	2.8E+11	3.79E+10	2.45E-79	3.80	3.3E+11	3.4E+11	4.76E+10	2.72E-79
4	4.22	3.3E+06	3.4E+06	4.74E+05	3.67E-84	3.92	1.4E+11	1.5E+11	2.03E+10	1.26E-79
5	4.23	2.9E+08	2.8E+08	4.07E+07	3.15E-82	4.13	5.8E+09	5.7E+09	8.21E+08	6.00E-81
					OR_A					
1	2.64	1.3E+08	2.7E+08	2.65E+07	5.03E-83	2.61	1.8E+08	3.1E+08	3.27E+07	6.02E-83
2	3.61	5.8E+03	9.0E+03	1.01E+03	4.94E-87	3.59	7.8E+03	1.8E+04	1.71E+03	8.18E-87
3	4.04	7.0E+10	8.2E+10	1.07E+10	7.22E-80	3.81	9.3E+10	9.5E+10	1.34E+10	7.65E-80
4	4.22	1.1E+06	1.2E+06	1.65E+05	1.28E-84	4.02	1.2E+10	1.5E+10	1.90E+09	1.28E-80
5	4.34	6.4E+08	6.8E+08	9.35E+07	7.82E-82	4.15	5.6E+09	5.5E+09	7.91E+08	5.89E-81
					RFP_A					
1	2.48	1.8E+08	2.4E+08	2.96E+07	4.64E-83	2.45	1.7E+08	2.3E+08	2.75E+07	4.17E-83
2	3.41	1.4E+03	6.8E+03	5.04E+02	2.05E-87	3.39	8.2E+02	5.9E+03	4.06E+02	1.65E-87
3	3.88	2.8E+05	2.0E+06	1.37E+05	8.33E-85	3.84	1.4E+05	8.4E+05	6.06E+04	3.53E-85
4	3.91	2.4E+04	5.7E+05	3.44E+04	2.14E-85	3.87	1.1E+05	1.2E+06	7.70E+04	4.59E-85
5	3.99	3.0E+08	3.6E+08	4.61E+07	3.05E-82	3.98	2.8E+08	3.5E+08	4.37E+07	2.83E-82

C.5 Geometry Optimization and Coordinates

Structure	θ; $i-j-k-l$ (°)
■ PBE0/6-31+G(d,p) in vacuum	96.528
■ B3LYP/aug-cc-pVDZ/PCM	90.256
■ B3LYP/TZVP/IEFPCM	109.986
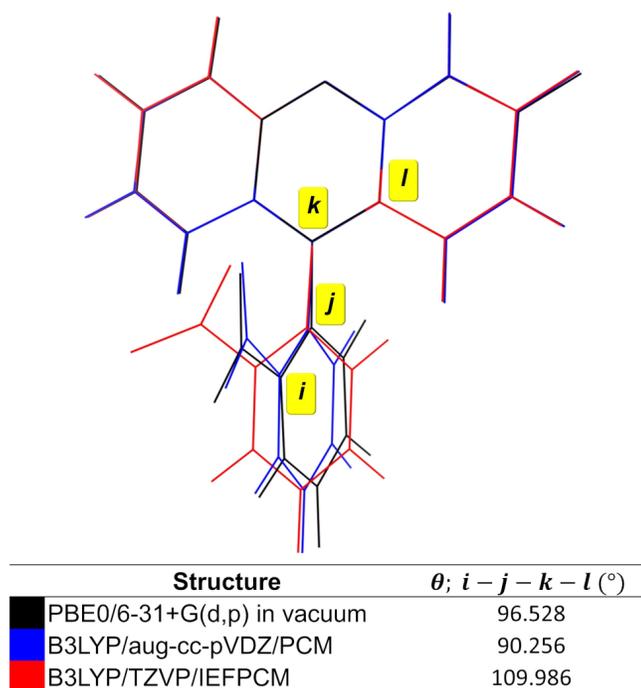
Figure C.7. Overlayed structures of fluorescein^{2-} optimized at different levels of theory i) in black, PBE/6-31+G(d,p) in vacuum as studied in this work, ii) in blue, B3LYP/aug-cc-pVDZ/PCM (using a non-equilibrium solvation model) by Gerasimova et al.,[222] and iii) in red, B3LYP/TZVP/IEFPCM by Zhou et al.[233]

a) Serotonin neutral
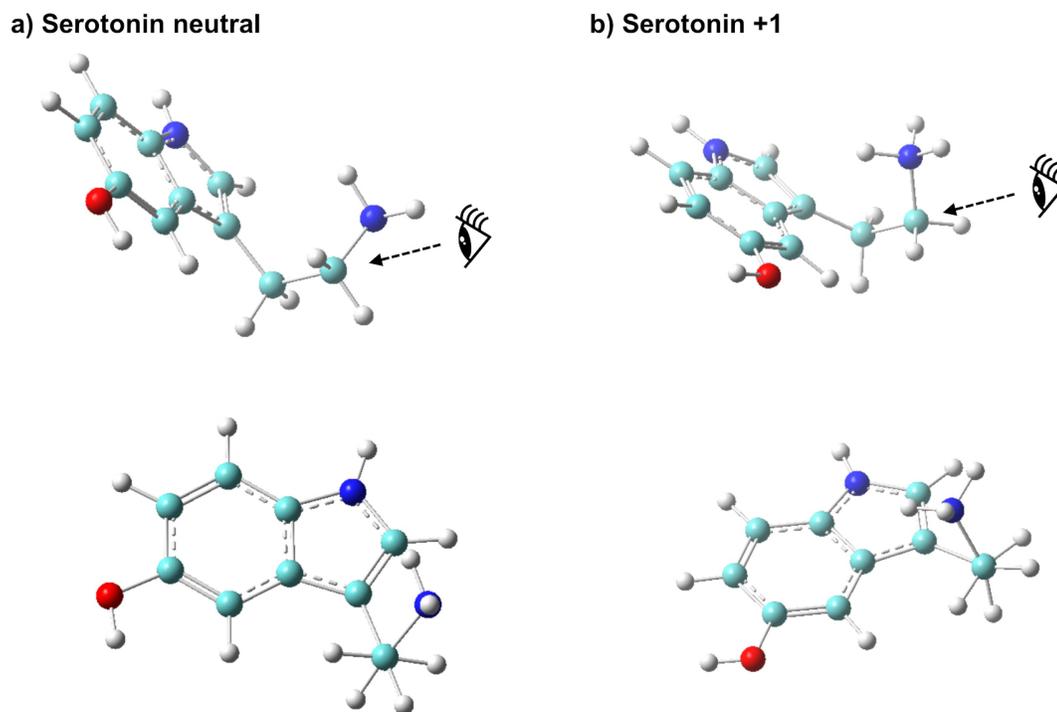
b) Serotonin +1

Figure C.8. Depiction of the gauche conformation of the optimized structures of serotonin, 5-HT and 5-HT+. This depiction is inspired on Lobayan and Schmit's contribution. [236]

	Structure	$\theta; i - j - k - l$ (°)
(blue)	B3LYP/6-31+G(d,p) in vacuum	92.598
(red)	CAM-B3LYP/6-31+G(d,p) in vacuum	92.544
(black)	PBE0/6-31+G(d,p) in vacuum	92.572
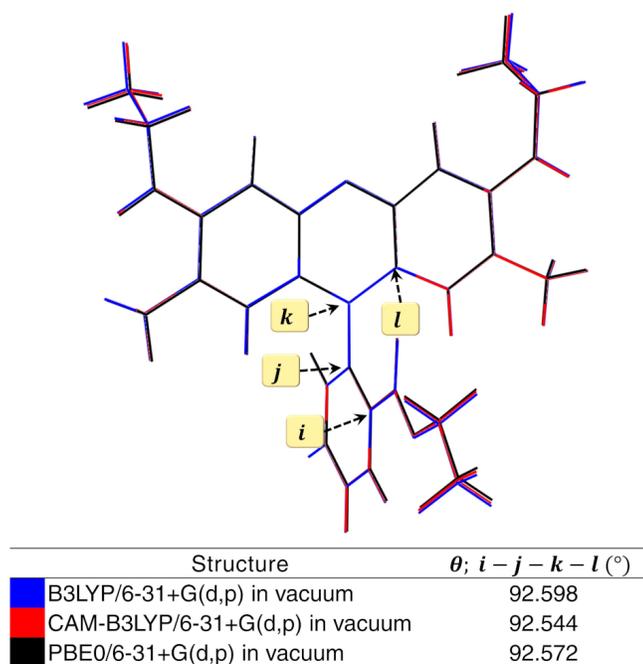
Figure C.9. Overlayed structures of rhodamine 6G optimized using the B3LYP/6-31+G(d,p) and CAM-B3LYP/6-31+G(d,p) methods in vacuum.

Fluorescein 2- PBE0/6-31+G(d,p) in Vacuum Coordinates

O –2.83871 –4.68872 –0.07389
O –2.84635 4.68424 –0.07388
C –2.75990 –2.32584 0.03953
C –2.76371 2.32149 0.03957
C –0.71717 –3.63940 –0.17620
C –0.72311 3.63835 –0.17617
C 0.02057 –2.49501 –0.19213
C 0.01651 2.49519 –0.19212
C –0.58041 –1.20905 –0.10626
C –0.58237 1.20823 –0.10624
C –1.99566 –1.18727 0.02369
C –1.99760 1.18417 0.02371
C –2.16667 –3.62899 –0.06716
C –2.17259 3.62560 –0.06712
O –2.65565 –0.00209 0.11095
C 0.12734 0.00017 –0.14530
C 1.57104 0.00127 –0.51543
C 1.84072 0.00154 –1.88926
C 3.14318 0.00247 –2.36775
C 4.20378 0.00315 –1.46569
C 3.94018 0.00291 –0.10393
C 2.63541 0.00196 0.39332
H –3.83547 –2.24878 0.12765
H –3.83915 2.24268 0.12768
H –0.24093 –4.61074 –0.23772
H –0.24848 4.61048 –0.23771
H 1.10064 –2.54149 –0.26290
H 1.09651 2.54344 –0.26291
H 1.00657 0.00099 –2.58136
H 3.32590 0.00266 –3.43638
H 5.22717 0.00390 –1.82554
H 4.72865 0.00345 0.63842
C 2.42338 0.00185 1.91676
O 3.47175 0.00244 2.60186
O 1.23160 0.00121 2.27945

Serotonin neutral PBE0/6-31+G(d,p) in Vacuum Coordinates

C –0.16693000 0.12525900 –0.29555700
C –0.75368600 1.35527400 0.06647500
C –2.10232000 1.45658400 0.40185100
C –2.86595100 0.30671700 0.37186700
C –2.29729600 –0.92643500 0.01000600

C –0.96074900 –1.03138000 –0.32349300
C 1.21477300 0.37933500 –0.58697700
C 1.40176500 1.72159600 –0.38990600
H –2.54844100 2.40359200 0.67792900
H –3.91714600 0.33096700 0.62337800
H –0.53557700 –1.98910500 –0.60325600
H 2.30129300 2.30455000 –0.50554600
H 0.10573000 3.28757200 0.18943000
N 0.22867600 2.31152300 0.00197400
O –3.14260900 –1.99661000 0.01016000
C 2.24889100 –0.62087100 –0.98380100
H –2.66366300 –2.78305900 –0.25328600
H 1.85332700 –1.27027000 –1.77029200
H 3.11669700 –0.10382000 –1.39776300
C 2.71896800 –1.48449300 0.18225100
H 1.86634700 –2.06305800 0.56498500
H 3.45608300 –2.20174500 –0.18493700
N 3.35548600 –0.65695600 1.18917800
H 3.70359000 –1.21569700 1.95454200
H 2.68871600 0.00147000 1.56916800

Serotonin 1+ PBE0/6-31+G(d,p) in Vacuum Coordinates

C –0.04093000 0.22560400 –0.30944700
C –0.68247600 1.41390300 0.10309700
C –2.04202000 1.44941700 0.40198100
C –2.75771100 0.27429000 0.29094000
C –2.13948000 –0.91904700 –0.12854600
C –0.79040200 –0.95582000 –0.43342100
C 1.34309800 0.55721700 –0.52232800
C 1.46867600 1.89737900 –0.24291800
H –2.53246900 2.36300300 0.71063700
H –3.81646100 0.27261800 0.51932400
H –0.37779600 –1.87950900 –0.81715400
H 2.34187100 2.52770900 –0.30375100
H 0.09048400 3.37281500 0.34606600
N 0.26351100 2.40933600 0.12393000
O –2.82831000 –2.07528000 –0.25607900
C 2.44691700 –0.36535300 –0.92420900
H 2.32953900 –0.71936900 –1.95136800
H 3.39964700 0.16632900 –0.88793000
C 2.51067600 –1.60502200 –0.03431900
H 1.74972500 –2.33407400 –0.29886500
H 3.48317200 –2.09114000 –0.05816500
N 2.22308900 –1.21792300 1.38338100

H 2.13196200 –2.02571800 1.99571600
H 1.34279200 –0.68385800 1.38579500
H 2.94486300 –0.60651400 1.76047800
H –3.76512300 –1.93534600 –0.10831100

Rhodamine 6G 1+ B3LYP/6-31+G(d,p) in Vacuum Coordinates

C 3.67327 –0.39142 –0.40488
C 2.50812 0.28011 –0.67859
C 1.21614 –0.27397 –0.43661
C 1.18408 –1.58440 0.11340
C 2.33164 –2.30013 0.40176
C 3.59497 –1.72306 0.15873
C 0.00009 0.39701 –0.68732
C –1.18382 –1.58460 0.11302
C –1.21592 –0.27418 –0.43700
C –2.50792 0.27967 –0.67941
H –2.56466 1.27910 –1.09741
C –3.67305 –0.39206 –0.40607
C –3.59470 –1.72369 0.15757
C –2.33134 –2.30054 0.40102
H 2.56482 1.27955 –1.09657
H 2.22932 –3.29693 0.80985
H –2.22898 –3.29732 0.80914
C 0.00007 1.75493 –1.31736
C 0.00032 1.81969 –2.71769
C –0.00017 2.95700 –0.57463
C 0.00031 3.04964 –3.37923
H 0.00050 0.89757 –3.29144
C –0.00017 4.18681 –1.25149
C 0.00007 4.23657 –2.64429
H 0.00050 3.07597 –4.46457
H –0.00036 5.10160 –0.67111
H 0.00007 5.19584 –3.15184
C –0.00043 2.90793 0.92153
O –0.00065 4.12682 1.48002
O –0.00044 1.87041 1.56395
O 0.00014 –2.19914 0.37773
N 4.74394 –2.38148 0.45060
H 5.60951 –1.92040 0.21511
N –4.74364 –2.38229 0.44909
H –5.60922 –1.92136 0.21332
C –0.00091 4.18189 2.93725
C –0.00113 5.64258 3.33977
H 0.88438 3.65063 3.29784

188

H –0.88623 3.65045 3.29753
H –0.00133 5.71727 4.43177
H 0.88749 6.15708 2.96260
H –0.88971 6.15690 2.96229
C 4.85436 –3.73290 1.00252
C 4.73859 –4.84230 –0.04904
H 4.09871 –3.85428 1.78555
H 5.82582 –3.78867 1.50186
H 4.83321 –5.82156 0.43103
H 3.77542 –4.80799 –0.56675
H 5.53064 –4.75401 –0.79892
C –4.85401 –3.73373 1.00099
C –4.73773 –4.84312 –0.05051
H –5.82561 –3.78965 1.50004
H –4.09858 –3.85497 1.78426
H –4.83233 –5.82239 0.42955
H –5.52956 –4.75498 –0.80064
H –3.77440 –4.80866 –0.56793
C –5.01642 0.23808 –0.67603
H –5.60722 –0.34665 –1.39329
H –5.61065 0.34146 0.24118
H –4.89468 1.23793 –1.09749
C 5.01662 0.23895 –0.67440
H 5.61053 0.34242 0.24300
H 5.60775 –0.34566 –1.39148
H 4.89484 1.23880 –1.09589

Rhodamine 6G 1+ CAM-B3LYP/6-31+G(d,p) in Vacuum Coordinates

C 3.65626 –0.38216 –0.40545
C 2.49889 0.28557 –0.67932
C 1.21079 –0.27139 –0.44020
C 1.17810 –1.57460 0.10620
C 2.32237 –2.28874 0.39497
C 3.57793 –1.71096 0.15490
C 0.00000 0.39475 –0.68895
C –1.17814 –1.57457 0.10620
C –1.21080 –0.27136 –0.44020
C –2.49889 0.28563 –0.67931
H –2.55426 1.28520 –1.09618
C –3.65627 –0.38208 –0.40545
C –3.57797 –1.71087 0.15490
C –2.32243 –2.28869 0.39498
H 2.55428 1.28514 –1.09618

H 2.21989 –3.28518 0.80187
H –2.21997 –3.28513 0.80188
C 0.00002 1.74972 –1.31553
C 0.00001 1.81623 –2.70907
C 0.00003 2.94165 –0.57275
C 0.00003 3.04438 –3.36233
H 0.00000 0.89642 –3.28539
C 0.00005 4.16935 –1.23838
C 0.00005 4.22385 –2.62590
H 0.00003 3.07574 –4.44675
H 0.00006 5.08012 –0.65249
H 0.00006 5.18411 –3.12989
C 0.00004 2.88389 0.91955
O 0.00006 4.08891 1.48561
O 0.00002 1.84503 1.54930
O –0.00003 –2.18427 0.36677
N 4.72340 –2.36377 0.44414
H 5.58880 –1.89968 0.21921
N –4.72346 –2.36367 0.44414
H –5.58885 –1.89956 0.21921
C 0.00006 4.12755 2.93359
C 0.00008 5.57887 3.34949
H 0.88420 3.59281 3.28863
H –0.88410 3.59284 3.28863
H 0.00008 5.64504 4.44057
H 0.88763 6.09517 2.97587
H –0.88745 6.09520 2.97588
C 4.82393 –3.70964 0.99579
C 4.67481 –4.81150 –0.04952
H 4.08080 –3.81928 1.79128
H 5.80115 –3.77987 1.47887
H 4.75799 –5.79270 0.42556
H 3.70696 –4.75907 –0.55450
H 5.45702 –4.73509 –0.80916
C –4.82401 –3.70953 0.99580
C –4.67491 –4.81140 –0.04952
H –5.80123 –3.77974 1.47887
H –4.08089 –3.81919 1.79129
H –4.75812 –5.79260 0.42557
H –5.45713 –4.73497 –0.80915
H –3.70706 –4.75898 –0.55449
C –4.99855 0.24205 –0.66862
H –5.58784 –0.34367 –1.38378
H –5.58719 0.34230 0.25064
H –4.88122 1.24172 –1.08882

C 4.99855 0.24194 –0.66862
H 5.58719 0.34217 0.25063
H 5.58783 –0.34380 –1.38379
H 4.88124 1.24161 –1.08882

Rhodamine 6G 1+ PBE0/6-31+G(d,p) in Vacuum Coordinates

The rhodamine 6g 1+ structure optimized using the B3LYP/6-31+G(d,p) method was
employed as initial structure. The absence of imaginary frequencies could not be verified
(refer to the main text).

C 3.64941900 –0.35835200 –0.40292500
C 2.48861300 0.30301900 –0.67929300
C 1.21014900 –0.26248100 –0.44022500
C 1.18080100 –1.56232100 0.11028600
C 2.32708000 –2.26804700 0.39993700
C 3.57702500 –1.68189000 0.15832000
C –0.00111700 0.39693300 –0.68811900
C –1.16428800 –1.57369700 0.10982700
C –1.20604000 –0.27421300 –0.44068000
C –2.48980200 0.27886000 –0.68034100
H –2.54592400 1.27572800 –1.09877000
C –3.64425700 –0.39377400 –0.40453600
C –3.55925700 –1.71657900 0.15668900
C –2.30378300 –2.29054400 0.39890400
H 2.53523800 1.30042200 –1.09762500
H 2.23213700 –3.26153200 0.81025300
H –2.19939100 –3.28308400 0.80920900
C –0.00755900 1.74310600 –1.31333000
C –0.00653000 1.81146200 –2.70512500
C –0.01477100 2.93039500 –0.56758600
C –0.01271900 3.04083700 –3.35257900
H –0.00095400 0.89452300 –3.28092500
C –0.02096600 4.15926000 –1.22717700
C –0.01998100 4.21678800 –2.61294800
H –0.01186600 3.07640000 –4.43388400
H –0.02649100 5.06578700 –0.63918400
H –0.02481600 5.17548300 –3.11338200
C –0.01550200 2.85959800 0.91747100
O –0.02338700 4.05687700 1.48731300
O –0.00901400 1.81682500 1.53572800
O 0.01115500 –2.17557800 0.37133300
N 4.72200700 –2.32322700 0.45041300
H 5.58051100 –1.85903600 0.21642100

N −4.69806300 −2.36910800 0.44815300
H −5.56095600 −1.91333500 0.21376600
C −0.02425700 4.07283900 2.92640700
C −0.03380500 5.51202800 3.36107600
H 0.86038500 3.54055900 3.27464100
H −0.90231700 3.52915800 3.27368800
H −0.03472100 5.56266700 4.44949100
H 0.84787700 6.03713000 2.99656600
H −0.92181800 6.02567400 2.99560700
C 4.82341400 −3.66599300 0.98471700
C 4.66307000 −4.74884500 −0.06877400
H 4.08768900 −3.78115000 1.78257300
H 5.80095200 −3.74384400 1.45887800
H 4.74073300 −5.73379100 0.39134700
H 3.69654100 −4.68168900 −0.56733100
H 5.44035800 −4.66785400 −0.82792500
C −4.78656400 −3.71284000 0.98232800
C −4.61509300 −4.79401800 −0.07113400
H −5.76352900 −3.80030300 1.45599400
H −4.05014400 −3.82083000 1.78054600
H −4.68334000 −5.77969900 0.38890900
H −5.39275800 −4.72060800 −0.83066900
H −3.64901900 −4.71739200 −0.56921000
C −4.98389500 0.21971600 −0.66572700
H −5.56783300 −0.36893300 −1.37817400
H −5.56906700 0.31717200 0.25233500
H −4.87168300 1.21655800 −1.08598100
C 4.98313600 0.26819300 −0.66342200
H 5.56681900 0.37142000 0.25495600
H 5.57319800 −0.31475500 −1.37551300
H 4.86140600 1.26386800 −1.08379300

C.6 Three-photon Absorption Experimental Benchmark Data

Table C.15. Experimentally Measured One-, Two-, and Three-photon Absorption Cross-sections for Serotonin, Fluorescein, and Rhodamine 6G

Dye	1PA	2PA	3PA
		Experimental σ^{XPA}	
	M^{-1} cm^{-1}	10^{-50} cm^4 s molecule^{-1} photon^{-1} **(GM)**	cm^6 s^2 photon^{-2}
Fluorescein	9E+04 [204,211]	32.8 [205]	1.63E−83 [191]
	8.5E+04 [235]	37 [303]	3.2E−84/4E-83 [235]
	7.7E+04 [214]	∼50 [304] (pH∼13)	3E−83/4E-82 [235]
		36 [204]	
Serotonin	3.25E+04 [305]	∼0.1 [249]	5E−84†
		∼0.08 [306]	∼7E−84 [248]†
Rhodamine 6G	1E+05 [204]	167 [205]	0.4/1.9E-81 [235]
	1.2E+05 [235]	2.0 [307]	6E-81 [256]
		130, 65, 10 [204]	
		λ for which σ^{XPA} is reported (nm)	
Fluorescein	490 [204,235]	800 [205]	1300 [191]
		780 [303]	1490 [235]
		800 [204]	
Serotonin	220 [305]	560 [306]	710 [248]
			740 [208]
Rhodamine 6G	530 [204]	700 [205]	1580 [235]
	527 [235]	1064 [307]	1300 [256]
		700, 800, 1060 [204]	
		λ at which σ^{XPA} is maximum (nm)	
Fluorescein	490 [211]	640 [204]	∼1450–1490 [235]
	490–500 [234]	780 [304]	
Serotonin	275 [247]†		831 nm† [249]
	275 (5-HT$^+$) [213,308]		
	307 [237]†		
Rhodamine 6G	530 [234,255]	700 [204,205]	1400 and ∼1580 [235]
	540 [235]	532 [309] (S$_{0\rightarrow3}$)	
	527, 346, 277, 246 (S$_{1\rightarrow4}$) [310]		

† The protolytic state for which the measurement was made is unclear in the original source.

Appendix D

Two-photon Absorption Cross-sections in Fluorescent Proteins Containing Non-canonical Chromophores Using Polarizable QM/MM
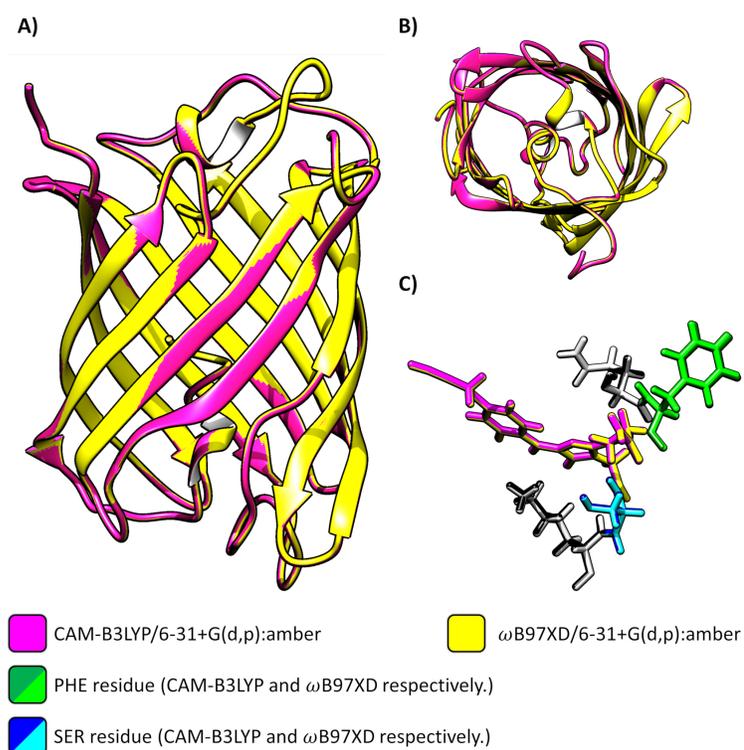
Figure D.1. For system nCC-14, A) side-view and B) top-view of overlapped structures of 1ZGO protein, and C) overlapped chromophore structures optimized using CAM-B3LYP/6-31+G(d,p) and ωB97XD methods within ONIOM scheme.
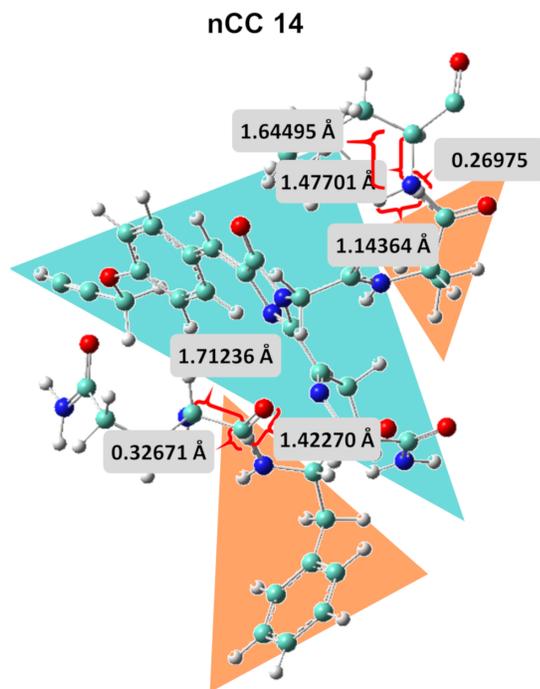
nCC 14

Figure D.2. Distances between dangling hydrogen atoms at the N-terminal and C-terminal side of the SER-nCC-PHE QM area and selected atoms in the MM zone.
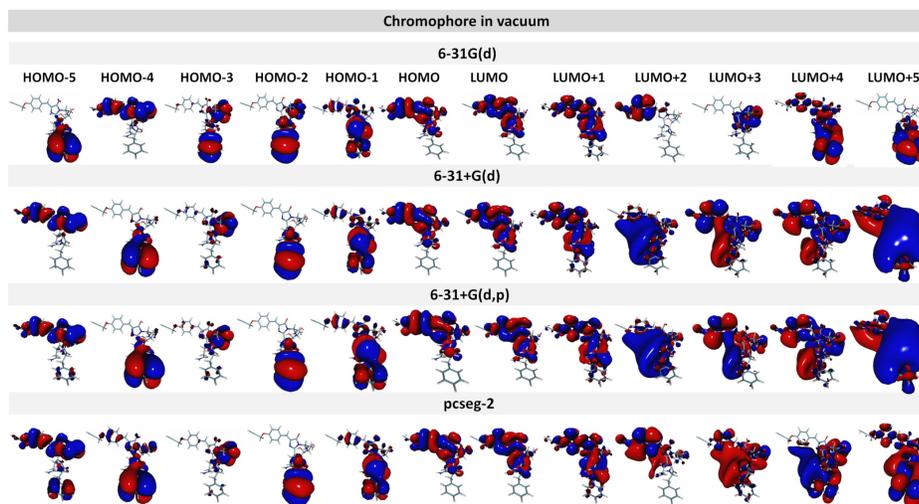
Figure D.3. Molecular orbitals computed for the isolated nCC 14 using CAM-B3LYP functional and different basis sets. Plots were obtained using Molden software. [301,302] Isovalue=0.01.
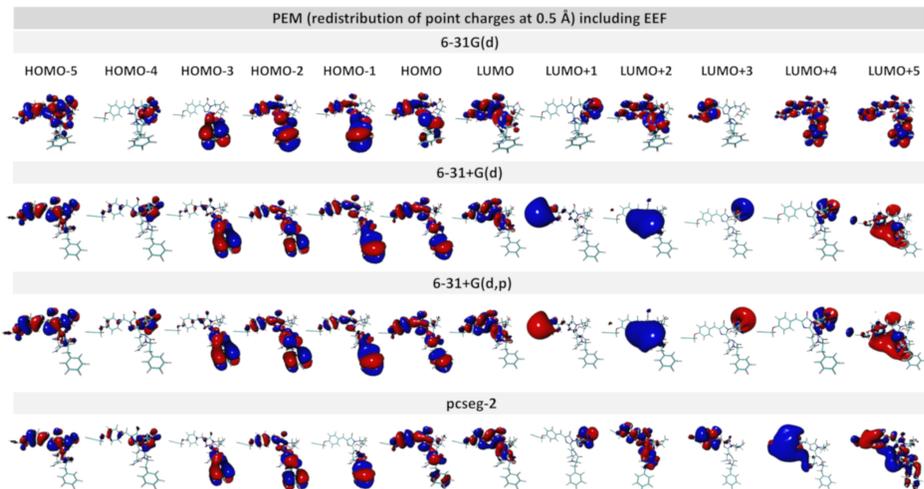
Figure D.4. Molecular orbitals computed for the 14-DsRed system using CAM-B3LYP functional and different basis sets. Charge redistribution at 0.5 Å was used. Plots were obtained using Molden software.[301,302] Isovalue=0.005.
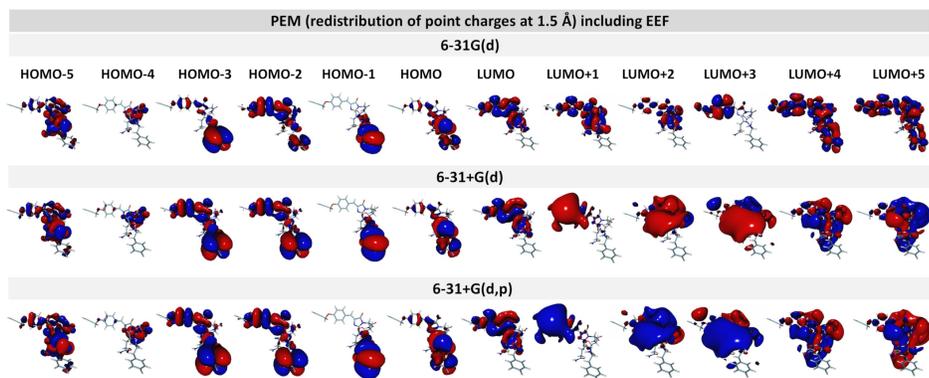
Figure D.5. Molecular orbitals computed for the 14-DsRed system using CAM-B3LYP functional and different basis sets. Charge redistribution at 1.5 Å was used. Plots were obtained using Molden software.[301,302] Isovalue=0.005.
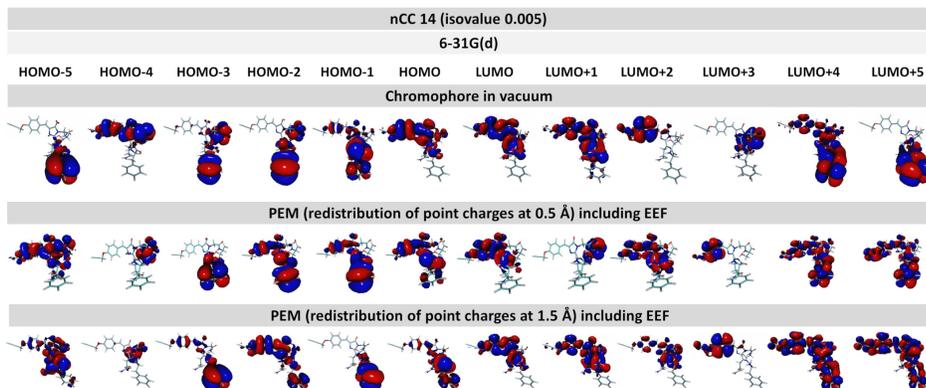
Figure D.6. Molecular orbitals computed using CAM-B3LYP/6-31G(d) method for the nCC 14, i) in vacuum and ii) in protein using charge redistribution at 0.5 Å and ii) in protein using charge redistribution at 1.5 Å. Plots were obtained using Molden software.[301,302] Isovalue=0.005.
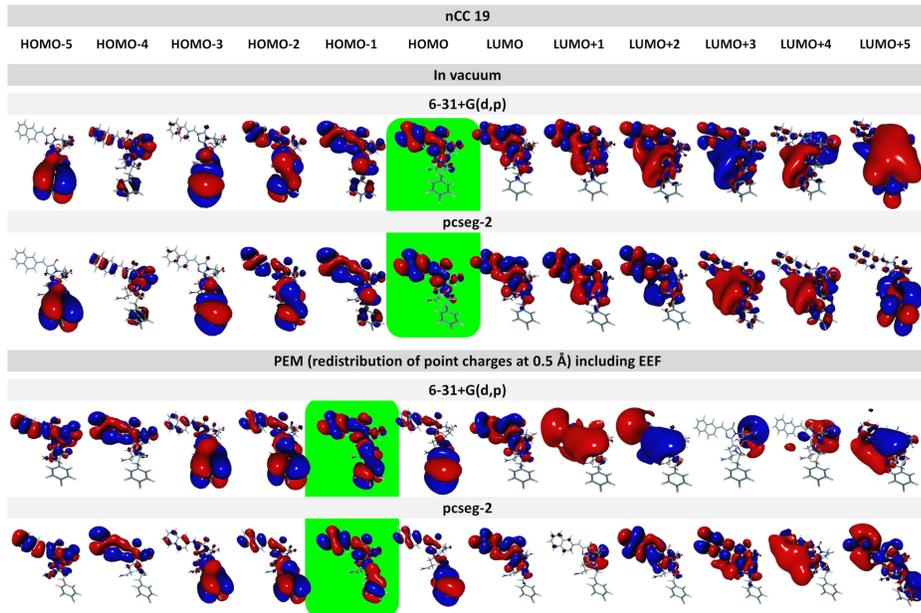
Figure D.7. Molecular orbitals computed for the nCC 19 and the nCC19-DsRed system using CAM-B3LYP and different basis sets. The shift between states of orbitals obtained in protein (through the PE model) and in vacuum is stressed in green. Plots were obtained using Molden software.[301,302] Isovalue=0.005.
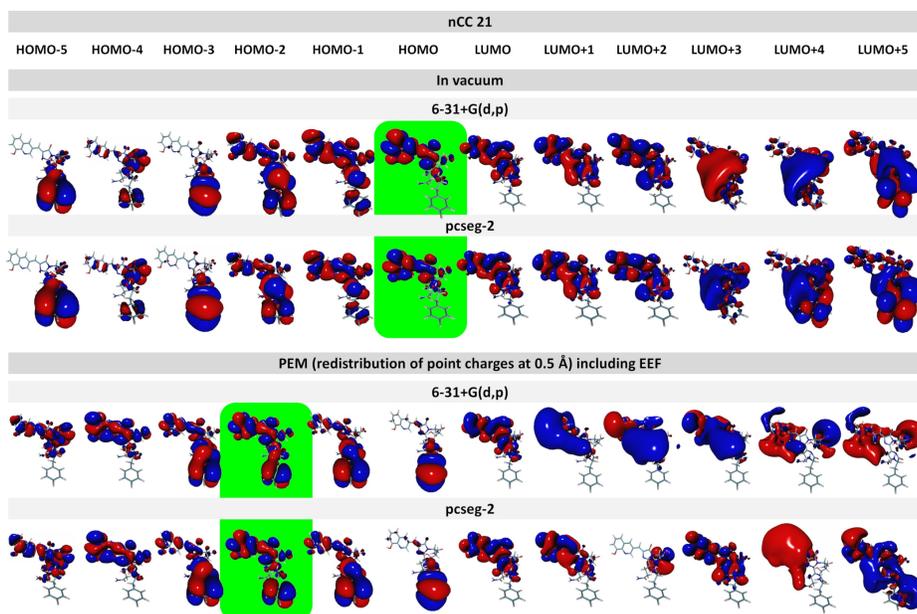
Figure D.8. Molecular orbitals computed for the nCC 21 and the nCC21-DsRed system using CAM-B3LYP and different basis sets. The shift between states of orbitals obtained in protein (through the PE model) and in vacuum is stressed in green. Plots were obtained using Molden software.[301,302] Isovalue=0.005.
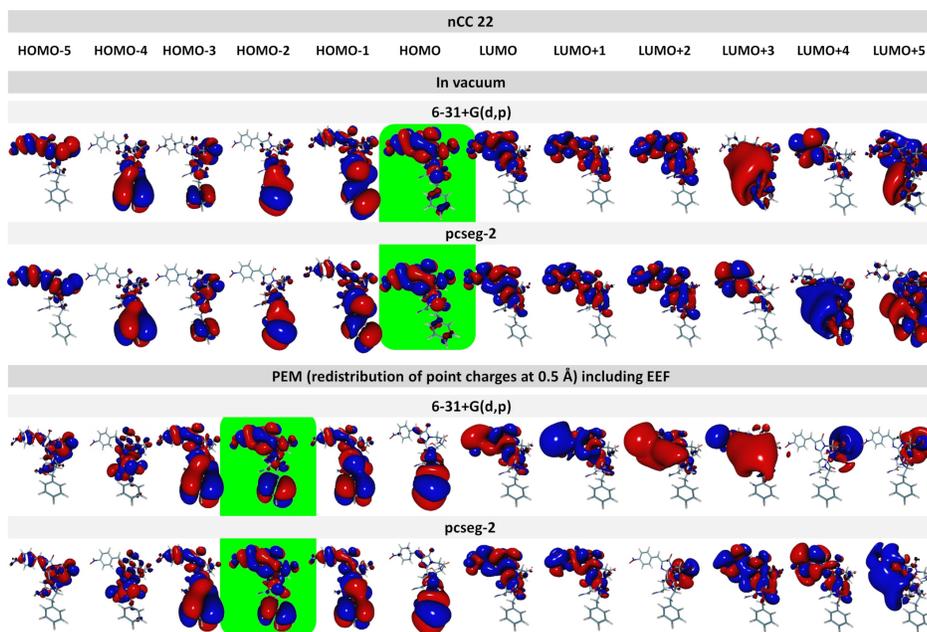
Figure D.9. Molecular orbitals computed for the nCC 22 and the nCC22-DsRed system using CAM-B3LYP and different basis sets. The shift between states of orbitals obtained in protein (through the PE model) and in vacuum is stressed in green. Plots were obtained using Molden software.[301,302] Isovalue=0.005.

Table D.1. Tilt and Twist Angles, and Angle Differences Between Structures Optimized Using ONIOM Scheme CAM-B3LYP/6-31+G(d,p):Amber, PBE0/6-31+G(d,p) in Vacuum,[14] and the Crystal Structure[7]

nCC	PBE0		CAM-B3LYP		CAM-B3LYP-PBE0		CAM-B3LYP-crystal		PBE0-crystal	
	Tilt (°)	Twist (°)	Tilt (°)	Twist (°)	Tilt (°)	Twist (°)	Tilt (°)	Twist (°)	Tilt (°)	Twist (°)
13	0.17	−0.01	−1.90	−0.67	−2.07	−0.66	−6.70	−3.98	−4.63	−3.32
14	−0.57	−1.12	0.84	−4.59	1.41	−3.46	−3.96	−7.89	−5.37	−4.43
16a	−0.36	−1.33	−1.55	−3.92	−1.19	−2.58	−6.35	−7.23	−5.16	−4.64
16b	0.21	−0.08	−3.46	−5.04	−3.67	−4.96	−8.26	−8.35	−4.59	−3.39
17	−0.53	−1.36	−5.89	−16.80	−5.36	−15.45	−10.69	−20.11	−5.34	−4.67
18	−0.40	−1.39	−3.93	−0.13	−3.53	1.26	−8.74	−3.44	−5.20	−4.70
19	−0.75	−3.71	−1.39	2.11	−0.64	5.82	−6.20	−1.20	−5.56	−7.02
20	−0.09	6.12	0.27	8.29	0.36	2.17	−4.53	4.98	−4.90	2.82
21	−0.15	−2.61	−2.94	−2.30	−2.78	0.30	−7.74	−5.61	−4.96	−5.92
22	0.23	0.11	−1.63	−1.76	−1.87	−1.86	−6.44	−5.06	−4.57	−3.20
20*			−1.59	11.8	−1.49	5.05	−6.39	7.87		

	Crystal structure	
	Tilt (°)	Twist (°)
CRQ	4.81	3.31

*Using an electrostatic embedding scheme.

nCC 20 from nCC 20-DsRed optimized using:
- Mechanical Embedding
- Electronic Embedding
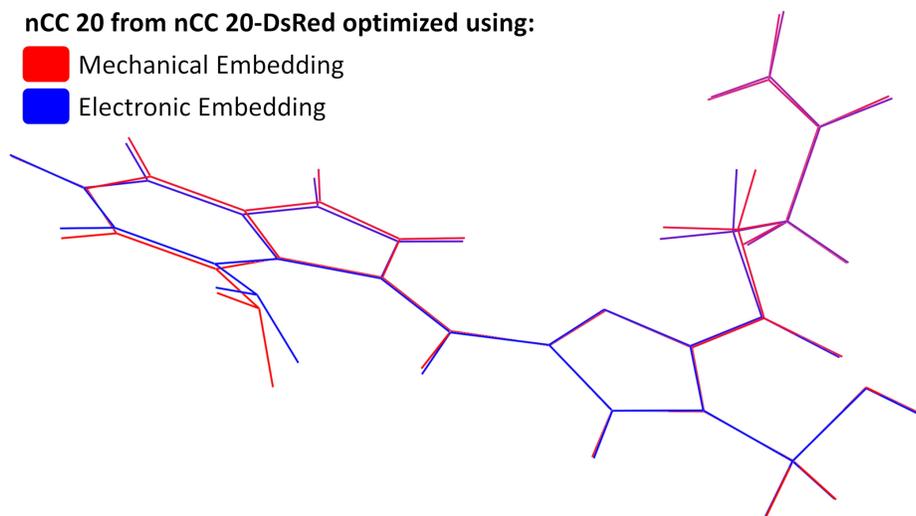
Figure D.10. Depiction of the structural differences in nCC 20 between optimizing nCC20-DsRed using mechanical embedding and electrostatic embedding.
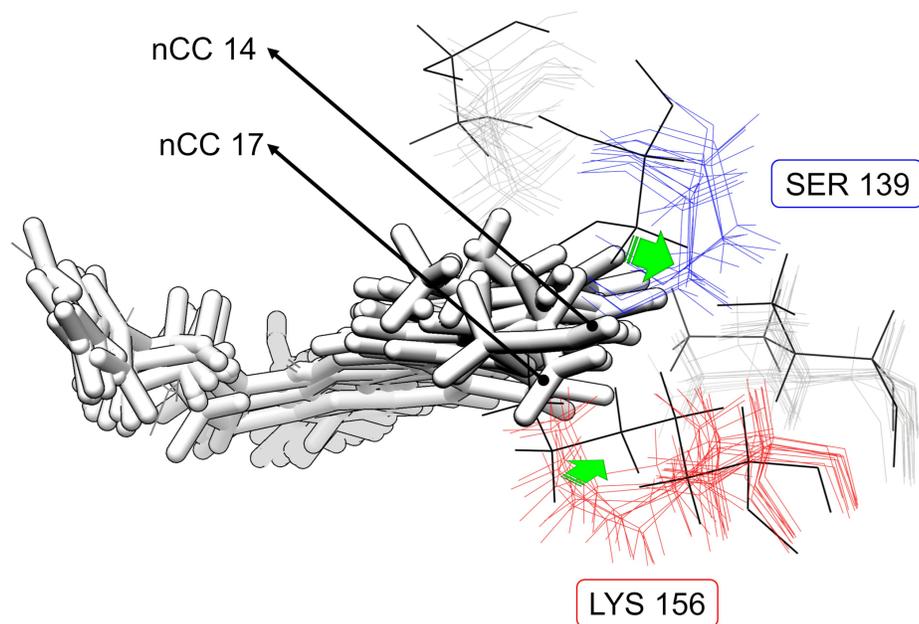
Figure D.11. Structural changes of selected residues, SER 139 (or 146 in the 1ZGO indexing) and LYS 156 (or 163 in the 1ZGO indexing), non-covalently bonded to the non-canonical chromophore. The green arrows indicate the contraction of serine and lysine towards the chromophore as the identity of the chromophore is changed. 1ZGO crystal structure is represented by the wire structure in black.

Table D.2. For System nCC14-DsRed, Excitation Energies, Oscillator Strengths, Two-Photon Absorption Cross-Sections, and Molecular Orbitals of the First Three Excited States Computed Using Different Basis Sets and Redistributing Point Charges that are Within 0.5 and 1.5 Å of a QM Atom (end Removing All Other Parameters)

	ω (eV)	OS	MOs	σ^{2PA} (GM)
nCC 14				
Redistributing point charges within 0.5 Å				
6-31G(d)				
S_1	3.433	0.700	HOMO→LUMO	4.0E+00
S_2	3.535	0.103	HOMO–2→LUMO	4.1E+00
S_3	3.969	0.002	HOMO–9→LUMO	7.0E−01
6-31+G(d)				
S_1	3.339	0.762	HOMO→LUMO	5.9E+00
S_2	3.568	0.012	HOMO–2→LUMO	2.8E+00
S_3	3.777	0.013	HOMO→LUMO+1	4.0E+00
6-31+G(d,p)				
S_1	3.336	0.760	HOMO→LUMO	5.9E+00
S_2	3.569	0.011	HOMO–2→LUMO	2.7E+00
S_3	3.774	0.014	HOMO→LUMO+1	4.0E+00
pcseg-2				
S_1	3.338	0.778	HOMO→LUMO	5.8E+00
S_2	3.565	0.013	HOMO–2→LUMO	2.5E+00
S_3	3.966	0.002	HOMO–6→LUMO	1.1E+00
Redistributing point charges within 1.5 Å				
6-31G(d)				
S_1	3.245	0.051	HOMO→LUMO	1.8E+00
S_2	3.474	0.789	HOMO–2→LUMO	9.0E+00
S_3	3.987	0.001	HOMO–10→LUMO	7.6E−01
6-31+G(d)				
S_1	3.260	0.186	HOMO→LUMO	1.6E+00
S_2	3.374	0.615	HOMO–3→LUMO	1.0E+01
S_3	3.861	0.013	HOMO–3→LUMO+1	4.4E+00
6-31+G(d,p)				
S_1	3.262	0.203	HOMO→LUMO	1.6E+00
S_2	3.372	0.596	HOMO–2→LUMO	1.0E+01
S_3	3.857	0.013	HOMO–2→LUMO+1	4.4E+00

Table D.3. For Transition S_1 of all Non-canonical Chromophores (nCC) Used in this Work, Excitation Energies (ω), Molecular Orbitals Involved in the Transition, and Oscillator Strengths Computed for Both, Chromophore in Vacuum and Protein-nCC Systems Using QM/MM Polarized Embedding (PE) Model

		PE			Chromophore (vacuum)		
nCC	Basis set	ω (eV)	Transition	OS	ω (eV)	Transition	OS
13	6-31+G(d,p)	3.273	HOMO→LUMO	0.654	3.242	HOMO→LUMO	0.612
13	pcseg-2	3.276	HOMO→LUMO	0.655	3.228	HOMO→LUMO	0.609
14	6-31G(d)	3.433	HOMO→LUMO	0.700	3.384	HOMO→LUMO	0.693
14	6-31+G(d)	3.339	HOMO→LUMO	0.762	3.320	HOMO→LUMO	0.724
14	6-31+G(d,p)	3.336	HOMO→LUMO	0.760	3.317	HOMO→LUMO	0.723
14	pcseg-2	3.338	HOMO→LUMO	0.778	3.302	HOMO→LUMO	0.724
16a	6-31+G(d,p)	3.369	HOMO→LUMO	0.581	3.384	HOMO→LUMO	0.539
16a	pcseg-2	3.367	HOMO→LUMO	0.599	3.363	HOMO→LUMO	0.543
16b	6-31+G(d,p)	3.401	HOMO→LUMO	0.658	3.375	HOMO→LUMO	0.662
16b	pcseg-2	3.396	HOMO→LUMO	0.676	3.353	HOMO→LUMO	0.665
17	6-31+G(d,p)	3.245	HOMO→LUMO	0.894	3.242	HOMO→LUMO	0.788
17	pcseg-2	3.244	HOMO→LUMO	0.920	3.223	HOMO→LUMO	0.783
18	6-31+G(d,p)	3.172	HOMO→LUMO	0.778	3.179	HOMO→LUMO	0.737
18	pcseg-2	3.198	HOMO→LUMO	0.830	3.167	HOMO→LUMO	0.733
19	6-31+G(d,p)	3.296	HOMO−1→LUMO	0.890	3.285	HOMO→LUMO	0.856
19	pcseg-2	3.290	HOMO−1→LUMO	0.901	3.268	HOMO→LUMO	0.853
20	6-31+G(d,p)	2.975	HOMO→LUMO	0.638	3.023	HOMO→LUMO	0.625
20	pcseg-2	3.043	HOMO→LUMO	0.759	3.010	HOMO→LUMO	0.625
21	6-31+G(d,p)	3.328	HOMO−2→LUMO	0.854	3.346	HOMO→LUMO	0.816
21	pcseg-2	3.323	HOMO−2→LUMO	0.869	3.325	HOMO→LUMO	0.816
22	6-31+G(d,p)	3.294	HOMO−2→LUMO	0.535	3.366	HOMO→LUMO	0.487
22	pcseg-2	3.292	HOMO−2→LUMO	0.549	3.347	HOMO→LUMO	0.490
		Electrostatic Embedding:					
20	pcseg-2	2.99	HOMO→LUMO	0.73	2.90	HOMO→LUMO	0.56

Appendix E

One and Two-photon Absorption Cross-Sections of the Fruits Series Fluorescent Proteins using QM and QM/pol-MM approaches

Table E.1. For the First Lowest Excited State S_1 of a Set of FP Chromophores, the Vertical Excitation Energies (VEEs), σ^{2PA}, Ground State Permanent Electric Dipole Moment (μ_{00}), Transition Electric Dipole Moment (μ_{01}), as Well as the Difference between the Excited and Ground State Permanent Electric Dipole Moments ($\Delta\mu_{10}$). All Computations Were Done Using the CAM-B3LYP Functional and the pcseg-2 Basis Set

Chromophore in vacuum						
S_1						
Chromophore	VEE (eV)	OS (au)	σ^{2PA} (GM)	μ_{00} (au)	μ_{01} (au)	$\Delta\mu_{10}$ (au)
DsRed	2.61	0.96	3	5.02	4.13	0.67
mCherry pH = 11	2.63	0.86	7	5.87	4.08	0.92
mCherry pH = 8[†]	2.71	0.94	9	6.83	4.06	1.07
mCherry pH = 8[‡]	2.61	0.91	7	5.84	4.16	0.88
mCherry pH = 8	2.61	1.02	8	6.98	4.26	0.91
mOrange	2.66	0.88	9	5.28	4.17	1.00
mPlum a	2.63	0.83	4	4.61	3.89	0.79
mPlum b	2.63	0.93	10	5.18	4.13	1.07
mStrawberry	2.65	0.90	6	5.77	4.10	0.86
nCC 20[ξ]	2.89	0.60	60	1.35	3.14	3.40

† Larger QM region in the ONIOM optimization.

‡ Glu 210, located above the chromophore, is deprotonated.

ξ Optimized within the mCherry pH = 11 protein barrel.

Table E.2. For the Second Lowest Excited State S_2 of a Set of FP Chromophores, the Vertical Excitation Energies (VEEs), σ^{2PA}, Ground State Permanent Electric Dipole Moment (μ_{00}), Transition Electric Dipole Moment (μ_{01}), as Well as the Difference between the Excited and Ground State Permanent Electric Dipole Moments ($\Delta\mu_{10}$). All Computations Were Done Using the CAM-B3LYP Functional and the pcseg-2 Basis Set

Chromophore in vacuum						
S_2						
Chromophore	VEE (eV)	OS (au)	σ^{2PA} (GM)	μ_{00} (au)	μ_{01} (au)	$\Delta\mu_{10}$ (au)
DsRed	3.47	0.00	0	5.02	0.01	3.15
mCherry pH = 11	3.52	0.00	0	5.87	0.05	3.23
mCherry pH = 8[†]	3.51	0.00	0	6.83	0.07	9.56
mCherry pH = 8[‡]	3.50	0.00	0	5.84	0.03	3.27
mCherry pH = 8	3.51	0.00	0	6.98	0.01	3.31
mOrange	3.55	0.00	0	5.28	0.04	3.36
mPlum a	3.52	0.00	0	4.61	0.07	3.04
mPlum b	3.54	0.00	0	5.18	0.01	3.33
mStrawberry	3.52	0.00	0	5.77	0.01	3.23
nCC 20[ξ]	3.30	0.01	12	1.35	0.27	3.22

† Larger QM region in the ONIOM optimization.

‡ Glu 210, located above the chromophore, is deprotonated.

ξ Optimized within the mCherry pH = 11 protein barrel.

Table E.3. For the Third Lowest Excited State S_3 of a Set of FP Chromophores, the Vertical Excitation Energies (VEEs), σ^{2PA}, Ground State Permanent Electric Dipole Moment (μ_{00}), Transition Electric Dipole Moment (μ_{01}), as Well as the Difference between the Excited and Ground State Permanent Electric Dipole Moments ($\Delta\mu_{10}$). All Computations Were Done Using the CAM-B3LYP Functional and the pcseg-2 Basis Set

Chromophore in vacuum						
S_3						
Chromophore	VEE (eV)	OS (au)	σ^{2PA} (GM)	μ_{00} (au)	μ_{01} (au)	$\Delta\mu_{10}$ (au)
DsRed	4.02	0.11	153	5.02	1.17	2.96
mCherry pH = 11	3.85	0.00	1	5.87	0.13	9.93
mCherry pH = 8[†]	3.57	0.00	0	6.83	0.00	3.13
mCherry pH = 8[‡]	4.02	0.17	242	5.84	1.57	1.41
mCherry pH = 8	3.84	0.00	1	6.98	0.09	9.56
mOrange	4.05	0.00	1	5.28	0.15	0.31
mPlum a	3.93	0.08	100	4.61	1.07	4.98
mPlum b	3.91	0.04	55	5.18	0.76	7.52
mStrawberry	3.80	0.01	6	1.35	0.47	5.91
nCC 20[ξ]	3.99	0.00	0	5.77	0.14	0.21

† Larger QM region in the ONIOM optimization.

‡ Glu 210, located above the chromophore, is deprotonated.

ξ Optimized within the mCherry pH = 11 protein barrel.

Table E.4. For the First Lowest Excited State S_1 of a Set of FPs, the Vertical Excitation Energies (VEE), σ^{2PA}, Ground State Permanent Electric Dipole Moment (μ_{00}), Transition Electric Dipole Moment (μ_{01}), as Well as the Difference between the Excited and Ground State Permanent Electric Dipole Moments ($\Delta\mu_{10}$). All Computations Were Done Using the CAM-B3LYP Functional and the pcseg-2 Basis Set

Chromophore—Protein						
			S_1			
Chromophore	VEE (eV)	OS (au)	σ^{2PA} (GM)	μ_{00} (au)	μ_{01} (au)	$\Delta\mu_{10}$ (au)
DsRed	2.63	0.52	1	582	3.1	0.4
mCherry pH = 11	2.71	0.48	3	439	3.2	0.6
mCherry pH = 8[†]	2.40	0.01	2	415	0.4	7.5
mCherry pH = 8[‡]	2.64	0.50	2	430	3.3	0.5
mCherry pH = 8	2.38	0.04	3	402	0.9	7.1
mOrange	2.67	0.44	3	347	3.1	1.2
mPlum a	2.61	0.48	1	476	3.2	0.5
mPlum b	2.22	0.00	0	408	0.2	7.2
mStrawberry	2.64	0.08	12	457	1.2	7.5
nCC 20[ξ]	2.63	0.52	3	440	3.3	7.8

† Larger QM region in the ONIOM optimization.
‡ Glu 210, located above the chromophore, is deprotonated.
ξ Optimized within the mCherry pH = 11 protein barrel.

Table E.5. For the Second Lowest Excited State S_2 of a Set of FPs, the Vertical Excitation Energy (VEE), σ^{2PA}, Ground State Permanent Electric Dipole Moment (μ_{00}), Transition Electric Dipole Moment (μ_{01}), as Well as the Difference between the Excited and Ground State Permanent Electric Dipole Moments ($\Delta\mu_{10}$). All Computations Were Done Using the CAM-B3LYP Functional and the pcseg-2 Basis Set

Chromophore—Protein						
			S_2			
Chromophore	VEE (eV)	OS (au)	σ^{2PA} (GM)	μ_{00} (au)	μ_{01} (au)	$\Delta\mu_{10}$ (au)
DsRed	4.10	8.81E-04	7.75E-02	582	0.1	9.5
mCherry pH = 11	3.39	3.66E-05	3.21E-02	439	0.0	11.3
mCherry pH = 8[†]	2.71	5.04E-01	2.12E+00	415	3.1	0.8
mCherry pH = 8[‡]	2.79	2.17E-03	2.09E-01	430	0.2	10.8
mCherry pH = 8	2.56	5.07E-01	1.81E+00	402	3.1	1.1
mOrange	2.71	3.91E-02	1.12E-01	347	0.9	10.0
mPlum a	3.68	5.60E-04	1.04E+00	476	0.1	10.2
mPlum b	2.55	5.49E-01	5.60E+00	408	3.3	1.0
mStrawberry	2.71	2.54E-01	2.24E+01	457	2.1	4.4
nCC 20[ξ]	3.83	1.04E-03	5.25E-01	440	0.1	0.8

† Larger QM region in the ONIOM optimization.
‡ Glu 210, located above the chromophore, is deprotonated.
ξ Optimized within the mCherry pH = 11 protein barrel.

Table E.6. For the Third Lowest Excited State S_3 of a Set of FPs, the Vertical Excitation Energy (VEE), σ^{2PA}, Ground State Permanent Electric Dipole Moment (μ_{00}), Transition Electric Dipole Moment (μ_{01}), as Well as the Difference between the Excited and Ground State Permanent Electric Dipole Moments ($\Delta\mu_{10}$). All Computations Were Done Using the CAM-B3LYP Functional and the pcseg-2 Basis Set

Chromophore—Protein						
			S_3			
Chromophore	VEE (eV)	OS (au)	σ^{2PA} (GM)	μ_{00} (au)	μ_{01} (au)	$\Delta\mu_{10}$ (au)
DsRed	4.21	0.11	0	582	1.2	2.5
mCherry pH = 11	3.48	0.00	0	439	0.0	9.8
mCherry pH = 8[†]	3.23	0.00	0	415	0.1	1.6
mCherry pH = 8[‡]	3.16	0.00	0	430	0.1	9.4
mCherry pH = 8	3.24	0.00	0	402	0.1	1.6
mOrange	3.56	0.00	0	347	0.0	9.9
mPlum a	3.88	0.01	7	476	0.3	8.9
mPlum b	3.27	0.00	0	408	0.2	1.3
mStrawberry	3.22	0.00	0	457	0.1	2.2
nCC 20[ξ]	3.98	0.00	0	440	0.2	2.3

† Larger QM region in the ONIOM optimization.
‡ Glu 210, located above the chromophore, is deprotonated.
ξ Optimized within the mCherry pH = 11 protein barrel.
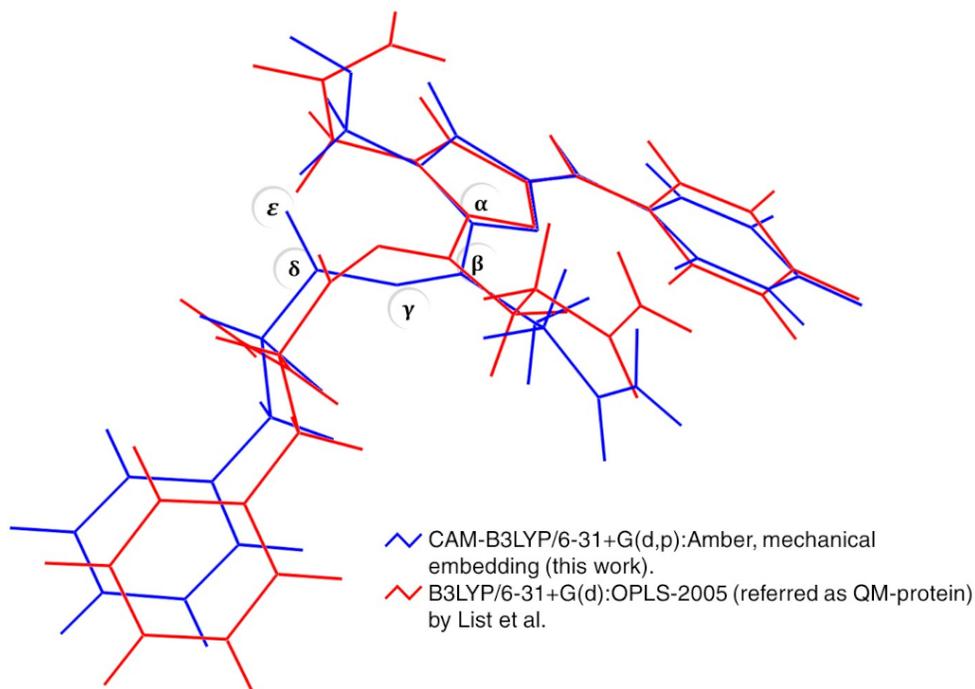
Figure E.1. Comparison of the chromophore structure obtained in this work at the ONIOM[CAM-B3LYP/6-31+G(d,p):Amber] level of theory with respect to the one obtained by List et al. at the B3LYP/6-31+G(d,p):OPLS-2005 level of theory.[86] The pair of angles ($\alpha - \beta - \gamma - \delta$, $\beta - \gamma - \delta - \epsilon$) are (–87°, 5°) and (106°, 103°), for this work and List et al.'s,[86] respectively.
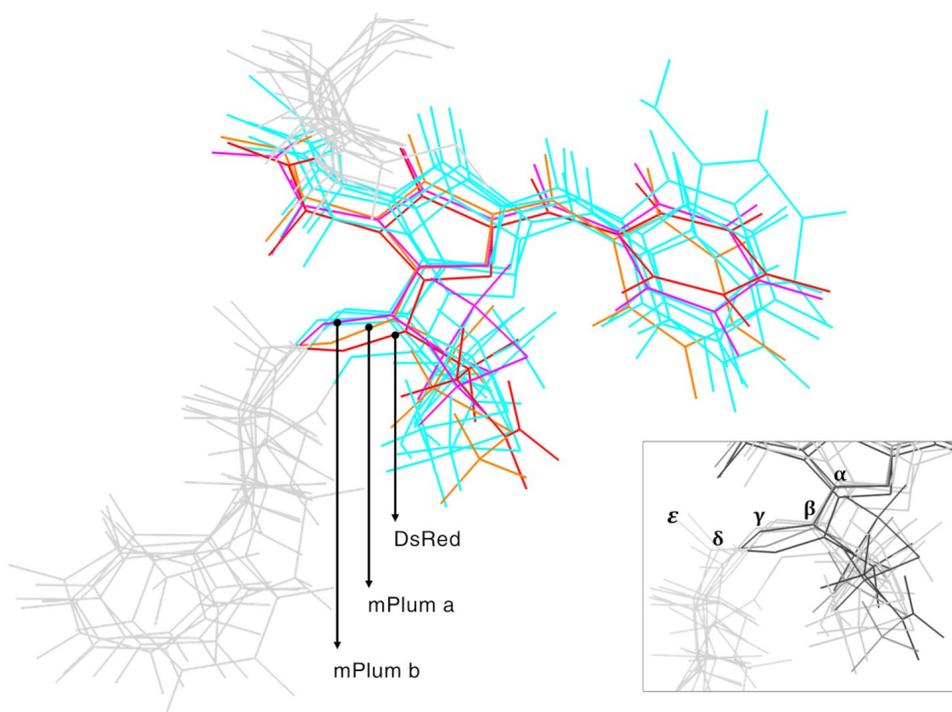
Figure E.2. Comparison of the acylimine moiety in the chromophore structure among the different FPs studied in this work. The optimization of the structures was done using the ONIOM QM/MM approach, mechanical embedding, and the TD-CAM-B3LYP/6-31+G(d,p):Amber level of theory. The acylimine conformation is present in our structures in two main conformations (angles are given in Table E.7), i) mCherry (pH = 8, pH = 11, and nCC20–mCherry), mStrawberry, as well as mOrange, and ii) DsRed, mPlum a, and mPlum b.

Table E.7. For a Curated Set of the FPs Discussed in This Work, the Dihedral Angles (in °) $\alpha - \beta - \gamma - \delta$ and $\beta - \gamma - \delta - \epsilon$ (Figure E.2) Containing the Acylimine Moiety of the RFP-type Chromophore

FP	$\alpha - \beta - \gamma - \delta$	$\beta - \gamma - \delta - \epsilon$
DsRed	4.7	−87.4
mCherry pH = 11	177.4	84.9
mCherry pH = 8[†]	−178.0	73.5
mOrange	177.5	95.5
mPlum a	7.2	−88.4
mPlum b	174.6	110.7
mStrawberry	176.3	92.6
nCC 20[ξ]	176.3	56.1
DsRed[86]	166.1	103.4

† The size of the QM region in the optimization is the same as in the rest of FPs.

ξ Optimized within the mCherry pH = 11 protein barrel.

Table E.8. For mCherry at pH = 11, the Effect of the Neighbouring Residues (see Figure 2 in the Main Text) in the VEE (eV), OS, σ^{2PA} (GM), Ground State Permanent Dipole Moment (μ_{00} [D]), Transition State Dipole Moment (μ_{0n} [D]), and the Difference between the Ground and the Excited State Permanent Dipole Moments ($\Delta\mu$ [D]). The Deviation Percentage of the σ^{2PA} with Respect to the CH6 Isolated Chromophore (Second Row of the Table) is Also Provided. In all Cases where the Residue PHE 62 is not Included, the Chromophore Was Truncated in Such a Way to Preserve its Acylimine Double Bond. Most of the Data Corresponds to the Transition S_1 (See Footnotes). All Computations Were Done Using the CAM-B3LYP Functional and the pcseg-2 Basis Set

Residue	Water	VEE	OS	σ^{2PA}	σ^{2PA} %	μ_{00}	μ_{0n}	$\Delta\mu$
CH6 63	—	2.6	1.1	13.4	N.A.	10.4	10.4	2.3
ARG 90 (16)	343 (23)	2.9	1.1	7.9	−41.1	20.0	9.9	1.8
ARG 90 (16)	—	2.9	1.1	8.9	−34.0	20.5	10.0	1.9
SER 64 (2)	343 (23), 384(13)	2.8	0.8	10.0	−25.4	7.6	10.3	2.1
TRP 88 (15)	232 (11), 343 (23), 384 (13)	2.7	0.9	10.2	−23.9	11.7	10.3	2.2
SER 64 (2)	283*, 343 (23), 374*, 384 (13)	2.7	1.0	10.3	−23.1	7.8	10.3	2.1
GLN 61 (21)	—	2.5	0.9	11.2	−16.4	11.7	9.8	2.2
SER 59 (19)	—	2.7	0.7	11.4	−14.9	12.1	9.8	2.2
PRO 60 (20)	—	2.6	0.8	11.5	−14.2	10.2	9.5	2.3
GLN 61 (21)	324*	2.5	0.8	11.6	−13.4	12.2	9.8	2.2
LEU 194 (6)	—	2.6	0.7	11.6	−13.4	12.8	10.0	2.3
SER 64 (2)	—	2.7	1.0	11.7	−12.7	11.5	10.4	2.2
TRP 88 (15)	—	2.6	0.8	11.9	−11.2	18.5	10.3	2.3
PRO 60 (20)	324*	2.6	0.8	11.9	−11.2	8.3	9.5	2.3
LYS 65 (3)	—	2.7	0.7	11.9	−11.2	8.5	10.1	2.2
—	343 (23)	2.7	1.1	12.0	−10.4	3.4	10.3	2.2
SER 64 (2)	287*, 374*	2.6	1.0	12.2	−9.0	12.3	10.3	2.2
GLU 210 (7)	—	2.7	0.9	12.3	−8.2	11.9	10.1	2.2
LEU 194 (6)	256 (14)	2.6	0.8	12.9	−3.7	10.1	10.0	2.3
GLN 39 (8)	—	2.6	0.9	13.1	−2.2	18.3	10.3	2.3
GLN 39 (8)	287*, 374*	2.6	0.8	13.4	0.0	20.3	10.2	2.3
ILE 192 (4)	—	2.6	1.0	14.2	6.0	3.9	10.4	2.3
PHE 62 (22)	232 (11)	2.6	0.9	14.3	6.7	11.8	10.4	2.3
ALA 41 (10)	—	2.6	0.8	14.6	9.0	22.8	10.4	2.3
—	256 (14)	2.6	1.1	14.7	9.7	8.9	10.4	2.3
ILE 156 (17)	—	2.6	0.8	14.7	9.7	3.5	10.4	2.3
PHE 62 (22)	—	2.6	0.8	15.1	12.7	13.9	10.4	2.4
PHE 11 (1)	—	2.6	1.0	15.3	14.2	28.2	10.6	2.3
GLN 158 (18)	—	2.6	1.1	16.6	23.9	7.6	10.7	2.4
SER 141 (5)	—	2.6	1.2	18.7	39.6	4.6	10.8	2.4
LYS 65 (3), SER 141 (5) GLN 158 (18), GLU 210 (7)	—	2.7†	1.0	19.3	44.0	14.2	10.6	2.4
SER 141 (5)	256 (14)	2.6	1.2	21.8	62.7	4.7	10.9	2.6
SER 141 (5), GLN 158 (18) GLU 210 (7)	—	2.7†	1.0	24.5	82.8	19.1	10.8	2.7
SER 141 (5), GLN 158 (18)	—	2.6	1.2	24.9	85.8	3.6	11.1	2.7
SER 141 (5), GLN 158 (18)	256 (14), 324*	2.6	1.2	30.9	130.6	5.5	11.0	2.9

* Not shown in Figure 2 in the main text.
† Data corresponds to the transition S_2.
— The cases where residues or water molecules are included only.

208

Table E.9. For mCherry at pH = 8, the Effect of the Neighbouring Residues (see Figure 2 in the Main Text) in the VEE (eV), OS, σ^{2PA} (GM), Ground State Permanent Dipole Moment (μ_{00} [D]), Transition State Dipole Moment (μ_{0n} [D]), and the Difference between the Ground and the Excited State Permament Dipole Moments ($\Delta\mu$ [D]). The Deviation Percentage of the σ^{2PA} with Respect to the Isolated Chromophore CH6 (Second Row of the Table) is Also Provided. In All Cases where the Residue PHE 62 is not Included, the Chromophore was Truncated in Such a way to Preserve its Acylimine Double Bond. All Computations Were Done Using the CAM-B3LYP Functional and the pcseg-2 Basis Set

Residue	Water	VEE	OS	σ^{2PA}	σ^{2PA} %	μ_{00}	μ_{0n}	$\Delta\mu$
CH6 63	—	2.618	1.117	13.5	N.A.	10.3	10.6	2.2
LYS 65 (3)		2.7[†]	0.9	7.5	−44.1	8.2	10.0	0.5
ARG 90 (16)	343 (23)	2.9	1.1	9.2	−31.9	20.1	10.2	1.8
ARG 90 (16)	—	2.8	1.1	10.0	−25.9	20.5	10.3	1.9
SER 64 (2)	343 (23), 384 (13)	2.7	0.9	11.0	−18.5	7.3	10.5	2.1
TRP 88 (15)	232 (11), 343 (23), 384 (13)	2.7	1.0	11.1	−17.8	13.0	10.5	2.2
GLU 210 (7)	—	2.6	0.9	11.5	−14.8	10.0	10.2	2.2
LEU 194 (6)	—	2.6	0.7	11.7	−13.3	12.3	10.2	2.2
SER 64 (2)	287*, 343 (23), 374*, 384 (13)	2.7	0.9	11.8	−12.6	7.0	10.4	2.2
SER 64 (2)	287*, 374*	2.6	1.0	12.0	−11.1	0.5	10.5	2.2
TRP 88 (15)	—	2.6	1.0	12.0	−11.1	17.6	10.6	2.2
SER 59 (19)	—	2.6	1.0	12.1	−10.4	8.1	10.2	2.2
SER 64 (2)	—	2.6	1.0	12.2	−9.6	2.5	10.6	2.2
—	343 (23)	2.7	1.1	12.5	−7.4	4.0	10.6	2.2
PRO 60 (20)	—	2.6	0.9	12.9	−4.4	13.3	9.9	2.3
GLN 61 (21)	—	2.5	1.0	13.0	−3.7	13.0	10.3	2.3
GLN 39 (8)	275*, 287*, 374*	2.6	0.8	13.0	−3.7	24.0	10.4	2.2
GLN 39 (8)	—	2.6	0.9	13.2	−2.2	18.5	10.5	2.2
LEU 194 (6)	256 (14)	2.6	0.8	13.3	−1.5	9.2	10.3	2.3
PRO 60 (20)	324*	2.6	0.9	13.5	0.0	11.4	9.9	2.3
GLN 158 (18)	—	2.6	1.1	14.2	5.2	13.1	11.0	2.2
THR 40	—	2.6	0.8	14.2	5.2	21.1	10.7	2.3
ILE 192 (4)	—	2.6	1.0	14.3	5.9	4.8	10.8	2.2
ILE 156 (17)	—	2.6	1.0	14.9	10.4	5.8	10.9	2.3
PHE 11 (1)	—	2.6	1.0	15.0	11.1	26.8	10.9	2.3
—	256 (14)	2.6	1.1	15.1	11.9	9.1	10.7	2.3
GLN 208	—	2.6	1.0	15.1	11.9	15.2	10.6	2.4
ALA 41 (10)	—	2.6	0.8	15.1	11.9	23.1	10.7	2.3
SER 141 (5), GLN 158 (18) GLU 210 (7)	—	2.6	1.1	15.5	14.8	11.2	10.9	2.3
PHE 62 (22)	232 (11)	2.6	1.0	15.6	15.6	15.2	10.9	2.3
SER 141 (5)	—	2.6	1.2	16.9	25.2	1.5	11.1	2.3
SER 141 (5), GLN 158 (18)	—	2.6	1.3	17.9	32.6	13.8	11.4	2.3
SER 141 (5)	256 (14)	2.6	1.2	20.3	50.4	1.3	11.1	2.5
SER 141 (5), GLN 158 (18)	256 (14), 324*	2.6	1.3	23.1	71.1	15.0	11.4	2.5

* Not shown in Figure 2 in the main text.

† Corresponding to the transition S_2.

— The cases where residues or water molecules are included only.

E.1   Coordinates for the Clusters of the Ser192 Mutants

CH6+Ser141+Ser192 Using UFF

C 1.26212 4.34839 0.21767
C 0.13102 3.87872 -0.65412
O -0.43717 4.61786 -1.44880
N -0.22128 2.57187 -0.44883
C -0.16538 4.82116 -4.85385

S -0.12971 3.07416 -5.34481
C -0.72190 2.32395 -3.79196
C 0.42776 2.07915 -2.82279
C -0.10781 1.74141 -1.43886
C -0.55282 0.30941 -1.20834
N -0.19078 -0.64781 -2.01289
O 3.11420 -5.62229 -4.99789
C 0.60226 -5.13056 -2.53748
C 1.53233 -5.76111 -3.26187
C 2.29298 -5.03481 -4.30189
C 1.98484 -3.60048 -4.50536
C 1.05713 -2.98813 -3.75511
C 0.34545 -3.69442 -2.68977
C -0.59460 -3.06522 -1.76244
C -0.77911 -1.77082 -1.47982
C -1.60510 -1.39504 -0.29866
O -2.26309 -2.13072 0.40858
N -1.36284 -0.05830 -0.17302
C -1.73952 0.65472 1.04156
C -2.97522 1.54273 0.93626
O -3.28827 2.29653 1.85658
H -1.50599 2.95177 -3.35661
H -1.19290 1.36941 -4.05020
H 1.09047 2.94777 -2.75324
H 1.05956 1.25623 -3.17646
H 0.04233 -5.68196 -1.78744
H 1.75702 -6.80975 -3.12386
H 2.53229 -3.08928 -5.28677
H 0.83328 -1.94164 -3.93235
H -1.15367 -3.78032 -1.15609
H -0.89573 1.25577 1.38822
H -1.95234 -0.08032 1.82724
H 0.19143 5.43523 -5.68552
H -1.18812 5.12585 -4.61616
H 0.48442 4.99916 -3.99426
N -3.74511 1.38709 -0.19387
C -4.94106 2.16604 -0.39573
H -3.42599 0.78670 -0.94046
C 2.82953 -8.89405 -9.92743
C 2.64591 -9.63401 -8.62970
O 3.38119 -10.55559 -8.29153
N 1.57304 -9.20154 -7.89640
C 1.34827 -9.61247 -6.51491
C -0.14860 -9.87824 -6.27395
O -1.03325 -9.50224 -7.04318

210

H 0.96465 -8.47958 -8.29089
H 1.90975 -10.53273 -6.31971
C 1.79227 -8.50246 -5.55281
H 1.33322 -7.54978 -5.84204
H 1.48478 -8.72412 -4.52628
O 3.20713 -8.34593 -5.53616
H 3.39246 -7.38999 -5.40573
N -0.43539 -10.49232 -5.06869
C -1.79072 -10.80629 -4.68768
H 0.31606 -10.97157 -4.58470
C -4.43257 -9.25566 -8.08199
C -3.99561 -7.81929 -8.19039
O -4.80025 -6.89797 -8.28909
N -2.63901 -7.63110 -8.17022
C -2.04833 -6.30823 -8.00967
C -0.72514 -6.21118 -8.78353
O 0.04721 -7.15914 -8.91517
H -2.02872 -8.43364 -7.98403
C -1.77797 -6.02782 -6.51844
O -0.94189 -4.89687 -6.28046
N -0.40707 -4.92828 -9.18416
C 0.86563 -4.59440 -9.76803
H -1.10736 -4.19886 -9.10000
H 1.87599 3.49891 0.53275
H 1.89582 5.03240 -0.35275
H 0.85971 4.85275 1.09853
H -5.59666 1.62368 -1.08045
H -4.65937 3.12646 -0.83467
H -5.45589 2.33595 0.55390
H -1.83146 -10.90093 -3.59993
H -2.07368 -11.75344 -5.15312
H -2.47495 -10.01872 -5.01577
H 3.89263 -8.87767 -10.18378
H 2.26572 -9.39428 -10.71790
H 2.49477 -7.85645 -9.83843
H 0.90945 -4.98293 -10.78782
H 1.66780 -5.03509 -9.16997
H 0.96552 -3.50610 -9.77576
H -4.67975 -9.64193 -9.07330
H -5.31423 -9.30680 -7.43580
H -3.65590 -9.88003 -7.63307
H -2.73423 -5.55083 -8.40695
H -1.31129 -6.89232 -6.03232
H -2.72650 -5.84769 -6.00186
H -0.03872 -5.14038 -6.55948

CH6+Ser192 From the Optimized Structure Using UFF

C 1.26212 4.34839 0.21767
C 0.13102 3.87872 -0.65412
O -0.43717 4.61786 -1.44880
N -0.22128 2.57187 -0.44883
C -0.16538 4.82116 -4.85385
S -0.12971 3.07416 -5.34481
C -0.72190 2.32395 -3.79196
C 0.42776 2.07915 -2.82279
C -0.10781 1.74141 -1.43886
C -0.55282 0.30941 -1.20834
N -0.19078 -0.64781 -2.01289
O 3.11420 -5.62229 -4.99789
C 0.60226 -5.13056 -2.53748
C 1.53233 -5.76111 -3.26187
C 2.29298 -5.03481 -4.30189
C 1.98484 -3.60048 -4.50536
C 1.05713 -2.98813 -3.75511
C 0.34545 -3.69442 -2.68977
C -0.59460 -3.06522 -1.76244
C -0.77911 -1.77082 -1.47982
C -1.60510 -1.39504 -0.29866
O -2.26309 -2.13072 0.40858
N -1.36284 -0.05830 -0.17302
C -1.73952 0.65472 1.04156
C -2.97522 1.54273 0.93626
O -3.28827 2.29653 1.85658
H -1.50599 2.95177 -3.35661
H -1.19290 1.36941 -4.05020
H 1.09047 2.94777 -2.75324
H 1.05956 1.25623 -3.17646
H 0.04233 -5.68196 -1.78744
H 1.75702 -6.80975 -3.12386
H 2.53229 -3.08928 -5.28677
H 0.83328 -1.94164 -3.93235
H -1.15367 -3.78032 -1.15609
H -0.89573 1.25577 1.38822
H -1.95234 -0.08032 1.82724
H 0.19143 5.43523 -5.68552
H -1.18812 5.12585 -4.61616
H 0.48442 4.99916 -3.99426
N -3.74511 1.38709 -0.19387
C -4.94106 2.16604 -0.39573

H -3.42599 0.78670 -0.94046
C -4.43257 -9.25566 -8.08199
C -3.99561 -7.81929 -8.19039
O -4.80025 -6.89797 -8.28909
N -2.63901 -7.63110 -8.17022
C -2.04833 -6.30823 -8.00967
C -0.72514 -6.21118 -8.78353
O 0.04721 -7.15914 -8.91517
H -2.02872 -8.43364 -7.98403
C -1.77797 -6.02782 -6.51844
O -0.94189 -4.89687 -6.28046
N -0.40707 -4.92828 -9.18416
C 0.86563 -4.59440 -9.76803
H -1.10736 -4.19886 -9.10000
H 1.87599 3.49891 0.53275
H 1.89582 5.03240 -0.35275
H 0.85971 4.85275 1.09853
H -5.59666 1.62368 -1.08045
H -4.65937 3.12646 -0.83467
H -5.45589 2.33595 0.55390
H 0.90945 -4.98293 -10.78782
H 1.66780 -5.03509 -9.16997
H 0.96552 -3.50610 -9.77576
H -4.67975 -9.64193 -9.07330
H -5.31423 -9.30680 -7.43580
H -3.65590 -9.88003 -7.63307
H -2.73423 -5.55083 -8.40695
H -1.31129 -6.89232 -6.03232
H -2.72650 -5.84769 -6.00186
H -0.03872 -5.14038 -6.55948

nCC+Ser141+Ser192 Using UFF

C 1.88921 4.92835 -0.10350
C 0.92082 4.50010 -1.17163
O 0.59284 5.24642 -2.08557
H 1.35699 5.51156 0.65177
N 0.38015 3.26156 -0.95402
C 0.01582 5.60157 -5.47815
S 0.34213 3.84292 -5.78713
C -0.32869 3.14303 -4.23972
C 0.79180 2.62094 -3.34813
C 0.29331 2.40698 -1.92583
C -0.33699 1.05977 -1.60663
N -0.13572 0.02389 -2.37176

```
C  -0.75865  -2.32827  -1.91394
C  -0.78254  -1.00339  -1.72151
C  -1.49058  -0.47266  -0.52544
O  -2.19703  -1.07983   0.24906
N  -1.12060   0.84225  -0.50842
C  -1.42394   1.65399   0.66488
C  -2.77221   2.37312   0.66146
O  -3.11647   3.05769   1.62457
H  -0.95065   3.88416  -3.72866
H  -0.99119   2.31605  -4.51621
H   1.64689   3.30569  -3.32326
H   1.18448   1.67853  -3.74708
H  -1.29772  -2.90712  -1.16874
H  -0.64364   2.39005   0.84903
H  -1.43678   0.98755   1.53667
H   0.40906   6.19461  -6.30873
H  -1.06144   5.77682  -5.41170
H   0.50239   5.93738  -4.56053
C  -0.03784  -3.11466  -2.86330
C  -0.11341  -4.54717  -3.02637
C   0.94606  -2.64659  -3.70644
C   0.94435  -4.88965  -3.87747
C  -0.92947  -5.59923  -2.51338
H   1.32769  -1.65022  -3.88606
C   1.29550  -6.21059  -4.16777
C  -0.54187  -6.92786  -2.77259
H   2.30962  -3.68294  -4.95403
C   0.54766  -7.22701  -3.58048
H   2.12872  -6.44776  -4.82262
H  -1.11077  -7.74650  -2.33782
H   0.80977  -8.26367  -3.77137
N   1.52540  -3.72350  -4.31735
N  -2.06972  -5.41387  -1.70461
H  -2.39658  -4.45750  -1.66929
H  -2.82538  -6.04748  -1.94789
N  -3.56411   2.20923  -0.45106
C  -4.86893   2.81896  -0.54182
H  -3.22406   1.67637  -1.23847
H  -5.30917   2.95687   0.44975
C   2.11906  -8.36080 -11.11624
C   1.91793  -9.31694  -9.97454
O   2.76370 -10.14751  -9.65531
H   1.56894  -7.43024 -10.95369
N   0.72417  -9.17744  -9.31277
C   0.52447  -9.81157  -8.01079
```

214

```
C  -0.96931  -9.84617  -7.64341
O  -1.82847  -9.27103  -8.31168
H   0.10642  -8.39574  -9.54758
H   0.88319 -10.84600  -8.06873
C   1.31279  -9.04559  -6.93944
H   1.16143  -9.47440  -5.94461
H   2.38676  -9.07173  -7.14876
O   0.92759  -7.67014  -6.88676
H   0.75805  -7.36912  -7.80409
N  -1.27952 -10.54818  -6.49592
C  -2.63836 -10.66229  -6.02545
H  -0.57775 -11.13993  -6.06967
H  -3.19169  -9.73870  -6.21648
C  -4.99953  -8.35654  -8.96172
C  -4.41233  -6.96978  -8.94798
O  -5.06513  -5.99382  -9.30631
H  -4.34471  -9.08845  -8.48589
N  -3.12875  -6.88972  -8.47369
C  -2.43957  -5.61558  -8.30077
C  -1.06683  -5.61468  -9.00435
O  -0.39921  -6.63198  -9.20274
H  -2.58185  -7.75133  -8.38845
C  -2.20410  -5.30790  -6.81488
O  -1.90189  -3.92264  -6.62494
N  -0.60785  -4.34773  -9.30658
C   0.70769  -4.09536  -9.82831
H  -1.17227  -3.55731  -9.00531
H   1.44240  -4.70915  -9.30079
H   2.68279   5.53114  -0.55218
H   2.34704   4.05451   0.36963
H  -4.76429   3.78973  -1.03252
H  -5.51265   2.17142  -1.14160
H   3.18247  -8.11041 -11.18193
H   1.79177  -8.82519 -12.04839
H  -3.12017 -11.49189  -6.54880
H  -2.61810 -10.86427  -4.95192
H   0.72032  -4.33915 -10.89309
H   0.93804  -3.03633  -9.68680
H  -5.18823  -8.66457  -9.99277
H  -5.94039  -8.34109  -8.40313
H  -3.09977  -5.53362  -6.22640
H  -1.37351  -5.90074  -6.41617
H  -1.56796  -3.83979  -5.71073
H  -3.04256  -4.81765  -8.75065
```

nCC+Ser192 From the Optimized Structure Using UFF

C 1.88921 4.92835 -0.10350
C 0.92082 4.50010 -1.17163
O 0.59284 5.24642 -2.08557
H 1.35699 5.51156 0.65177
N 0.38015 3.26156 -0.95402
C 0.01582 5.60157 -5.47815
S 0.34213 3.84292 -5.78713
C -0.32869 3.14303 -4.23972
C 0.79180 2.62094 -3.34813
C 0.29331 2.40698 -1.92583
C -0.33699 1.05977 -1.60663
N -0.13572 0.02389 -2.37176
C -0.75865 -2.32827 -1.91394
C -0.78254 -1.00339 -1.72151
C -1.49058 -0.47266 -0.52544
O -2.19703 -1.07983 0.24906
N -1.12060 0.84225 -0.50842
C -1.42394 1.65399 0.66488
C -2.77221 2.37312 0.66146
O -3.11647 3.05769 1.62457
H -0.95065 3.88416 -3.72866
H -0.99119 2.31605 -4.51621
H 1.64689 3.30569 -3.32326
H 1.18448 1.67853 -3.74708
H -1.29772 -2.90712 -1.16874
H -0.64364 2.39005 0.84903
H -1.43678 0.98755 1.53667
H 0.40906 6.19461 -6.30873
H -1.06144 5.77682 -5.41170
H 0.50239 5.93738 -4.56053
C -0.03784 -3.11466 -2.86330
C -0.11341 -4.54717 -3.02637
C 0.94606 -2.64659 -3.70644
C 0.94435 -4.88965 -3.87747
C -0.92947 -5.59923 -2.51338
H 1.32769 -1.65022 -3.88606
C 1.29550 -6.21059 -4.16777
C -0.54187 -6.92786 -2.77259
H 2.30962 -3.68294 -4.95403
C 0.54766 -7.22701 -3.58048
H 2.12872 -6.44776 -4.82262
H -1.11077 -7.74650 -2.33782
H 0.80977 -8.26367 -3.77137

N 1.52540 -3.72350 -4.31735
N -2.06972 -5.41387 -1.70461
H -2.39658 -4.45750 -1.66929
H -2.82538 -6.04748 -1.94789
N -3.56411 2.20923 -0.45106
C -4.86893 2.81896 -0.54182
H -3.22406 1.67637 -1.23847
H -5.30917 2.95687 0.44975
C -4.99953 -8.35654 -8.96172
C -4.41233 -6.96978 -8.94798
O -5.06513 -5.99382 -9.30631
H -4.34471 -9.08845 -8.48589
N -3.12875 -6.88972 -8.47369
C -2.43957 -5.61558 -8.30077
C -1.06683 -5.61468 -9.00435
O -0.39921 -6.63198 -9.20274
H -2.58185 -7.75133 -8.38845
C -2.20410 -5.30790 -6.81488
O -1.90189 -3.92264 -6.62494
N -0.60785 -4.34773 -9.30658
C 0.70769 -4.09536 -9.82831
H -1.17227 -3.55731 -9.00531
H 1.44240 -4.70915 -9.30079
H 2.68279 5.53114 -0.55218
H 2.34704 4.05451 0.36963
H -4.76429 3.78973 -1.03252
H -5.51265 2.17142 -1.14160
H 0.72032 -4.33915 -10.89309
H 0.93804 -3.03633 -9.68680
H -5.18823 -8.66457 -9.99277
H -5.94039 -8.34109 -8.40313
H -3.09977 -5.53362 -6.22640
H -1.37351 -5.90074 -6.41617
H -1.56796 -3.83979 -5.71073
H -3.04256 -4.81765 -8.75065