THE ROLE OF THE ENVIRONMENT AROUND THE CHROMOPHORE IN CONTROLLING THE PHOTOPHYSICS OF FLUORESCENT PROTEINS

by

Tirthendu Sen

A Dissertation Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(CHEMISTRY)

August 2021

Acknowledgements

Graduate school of USC has been a quite learning experience for me since Fall 2016. This will remain as the most defining and fulfilling chapter of my career and life. I was lucky to have Professor Anna Krylov as my advisor. Her invaluable advice and encouragement helped me to understand and appreciate theoretical and computational chemistry. During the lab rotation, she exposed me to various directions that were pursued at the iOpenShell lab at that time. After considering different options I choose to work on excited-state properties of fluorescent proteins. Prof. Krylov is not only a great scientist but also an inspiration for novices like me at the beginning of their scientific research careers. She encouraged me to read, test, and discuss extensively, however crazy idea it might have been. One of her best qualities is the way she gradually integrates her students into the scientific community. I have attended 9 conferences and 1 workshop during my PhD, which allowed me to showcase my work as well as learn from others. All helped me to grow professionally. I cannot appreciate her contributions enough for gradually shaping both my scientific thinking and computational skills, which will stay with me for rest of my life.

I feel lucky to be a part of the Department of Chemistry of USC because of the great scientific environment it provides. I would like to thank Prof. Chi Mak, Prof. Sri Narayan, Prof. Moh El-Naggar, and Prof. Jahan Dawlaty for being on the committee. I appreciate Prof. Chi Mak's encouragement to to dig deep into problems. I also thank Prof. Oleg Prezdo and Prof. Alex Benderskii for teaching us concepts in statistical mechanics and spectroscopy.

I would like to thank all my collaborators, Profs. Alexey Bogdanov, Konstantin Lukyanov, Alexander Nemukhin, Bella Grigorenko, Igor Polyakov, and Mikhail Baranov, as well as Prof. Yingying Ma, who worked with us for a brief period of time and returned to her university in China to continue her independent scientific career.

I have learned a lot from senior group members in Prof. Krylov's lab. Dr. Samer Gozem, Dr. Kaushik Nanda, Dr. Alexandre Barrozo, Dr. Wojciech Skomorowski welcomed me and made me feel comfortable in the group. I cannot thank enough Dr. Atanu Acharya who explained to me and taught me the basics of the project, which we started to work on during my first semester at USC. I would also like to thank all our current group members — Dr. Sven Kaehler, Dr. Maristella Alessio, Dr. Yongbin Kim, Madhubani Mukherjee, Goran Gludetti, Saikiran Kotaru, Ronit Sarangi, Pawel Wojcik, Sourav Dey, and Nayanthara Jayadev. Special thanks to Sahil Gulania for all wonderful shared memories and friendship that will endure in time to come. I am also deeply grateful to Dr. Dibyendu Mondal for his help when I first arrived to the United States.

I treasure the time I spent with my roommates and friends with whom I was fortunate to share so many happy memories. Starting from Michael to Sayan, Tushar, Avik, Diganta, Prakarsh,

we built new families with diverse backgrounds. I have to leave this place in the course of time with many precious memories.

I would like to take this opportunity to acknowledge the support and encouragement that my parents showed over the years. Being a school teacher, my father encouraged me to make my own choices in pursuit of happiness. My mother shares the biggest credit to whatever I have achieved in life so far. Without their support, struggle, sacrifice I would not be able to do what I always wanted to do.

List would not be complete without thanking my teachers from whom I have had the privilege to learn: Profs. Ananta Ghosh and Sukriti Samanta, my first chemistry teachers in undergraduate days, Prof. Naresh Patwari, my master's advisor, Prof. Sourav Pal, who taught us the basics of electronic structure theory, and Prof. Sandip Kar who was source of constant encouragement at any stage when in doubt.

Table of contents

List of tables

List of figures

Abstract

Fluorescent proteins from the family of Green Fluorescent Protein (GFP) are unique in that they are the only fluorescent probes of natural origin. Their photoproperties make them suitable for a wide variety of applications. Fluorescent proteins are useful devices for studying the mechanistic details of various processes in cells, both *in vitro* and *in cellulo*. Excited-state lifetime is of the fundamental importance, as it limits the time-scales of competing relaxation channels of the excited chromophore. Structural changes in an excited state can cause temporary and permanent loss of fluorescence. For example, cis-trans photoisomerization often proceeds via a transient dark state, whereas a stable twisted geometry in excited state leads to a long-living dark state with a loss of fluorescence. This thesis covers studies of different fluorescent proteins and provides a mechanistic/operational insights into phenomena such as photoswitching in Dreiklang and loss of fluorescent quantum yield upon specific mutations in enhanced green fluorescent protein (EGFP). Chapter 1 presents an overview of GFP-like proteins.

Chapter 2 presents the theoretical methods for computing radiative and nonradiative lifetimes and discusses requisite computational tools. An equation for radiative lifetime can be derived from first principles with classical harmonic oscillator, which allows us to estimate radiative lifetime with a help of electronic structure calculations. The dominant channel responsible for nonradiative decay is discussed in Chapter 1. To capture the essential physics and

to quantify nonradiative lifetime, we performed dynamics simulations of excited-state chromophores. Towards this end, we re-parameterised the ground-state forcefield parameters of the chromophore to describe excited-state potential energy surface (PES) with *ab initio* calculations. We then use excited-state lifetime to estimate fluorescent quantum yield in model systems. We then discuss multiple decay channels operational in proteins like EGFP and how the protonation states of the key residues affect excited-state dynamics. We utilize thermodynamic cycles to estimate free-energy changes upon changing the key protonation states to compute relative populations of different protonation states.

In Chapter 3 fluorogenic dyes based on the GFP chromophore are discussed. The compounds contain a pyridinium ring instead of phenolate and feature large Stokes shifts and solvent-dependent variations in the fluorescence quantum yield, which facilitates their use for imaging the membrane structure of endoplasmic reticulum. Electronic structure calculations explain the trends in their solvatochromic behavior.

Chapter 4 focuses on EGFP — one of the most popular genetically encoded fluorescent probes, which carries the threonine-tyrosine-glycine (TYG) chromophore, undergoes efficient green-to-red oxidative photoconversion (redding) with electron acceptors. In contrast, enhanced yellow fluorescent protein (EYFP), a close EGFP homologue (5 amino acid substitutions), with glycine-tyrosine-glycine (GYG) chromophore, is much less susceptible to redding and requires halide ions in addition to the oxidants. We clarified the role of the first chromophore-forming amino acid in photoinduced behavior of these fluorescent proteins. To that end, we compared photobleaching and redding kinetics of EGFP, EYFP, and their mutants with reciprocally substituted chromophore residues, EGFP-T65G and EYFP-G65T. Experimental measurements showed that T65G mutation significantly increases EGFP photostability and inhibits its

excited-state oxidation. Remarkably, while EYFP-G65T demonstrated highly increased spectral sensitivity to chloride, it is also able to undergo redding independent of chloride. Atomistic calculations revealed that the GYG chromophore has an increased flexibility, which facilitates radiationless relaxation leading to the reduced fluorescence quantum yield in the T65G mutant. The GYG chromophore also has larger oscillator strength relative to TYG, which leads to a shorter radiative lifetime (i.e., a faster rate of fluorescence). The faster fluorescence rate partially compensates for the loss of quantum efficiency due to radiationless relaxation. The shorter excited-state lifetime of the GYG chromophore is responsible for its increased photostability and resistance to redding. In EYFP and EYFP-G65T, the chromophore is stabilized by π-stacking with Tyr203, which suppresses its twisting motions relative to EGFP.

Chapter 5 presents the results of high-level electronic structure and dynamics simulations of the photoactive protein Dreiklang. With the goal of understanding the details of Dreiklang's photocycle, we carefully characterized the excited states of the ON- and OFF-forms of Dreiklang. The key finding of our study is the existence of a low-lying excited state of a charge-transfer character in the neutral ON form and that population of this state, which is nearly isoenergetic with the locally excited bright state, initiates a series of steps that ultimately lead to the formation of the hydrated dark chromophore (OFF state). These results allowed us to refine the mechanistic picture of Dreiklang's photocycle and photoactivation.

Chapter 6 introduces BrUSLEE—BRight Ultimately Shorttime Enhanced Emitter— a new fluorescent protein derived from the enhanced green fluorescent protein (EGFP) by 3 mutations: T65G/Y145M/F165Y. BrUSLEE shows an unusual combination of high fluorescence brightness and short fluorescence lifetime. To explain the peculiarities of its photobehavior, we investigated fine structural determinants of the fluorescence lifetime in connection with

brightness by combination of time-resolved fluorescence measurements and atomistic simulations. High-resolution fluorescence measurements revealed 2 distinct subpopulations co-existing in a wide temperature range (4-300 K). The fluorescence lifetimes of these emissive states change considerably with temperature, converging to low temperature (intrinsic) lifetimes that are vastly different from each other and from that of the parental EGFP. The crystal structure and 15N-NMR spectroscopy of BrUSLEE show no obvious structural heterogeneity. Atomistic simulations suggest that the heterogeneity arises due to co-existing populations of different protonation states of chromophore-adjacent titratable residues. Different protonation states of His148 alter the hydrogen-bond network around the chromophore, which significantly affects its twisting flexibility in the excited state. Changes in the hydrogen-bond network also explain the variations in photo-physical properties among EGFP and the T65G, T65G/Y145M, and T65G/Y145M/F165Y (BrUSLEE) mutants.

The result of the research presented in this thesis were summarized in the following publications:

1. Y. G. Ermakova, T. Sen, Y. A. Bogdanova, A. Y. Smirnov, N. S. Nadezha, A. I. Krylov, M. S. Baranov, Pyridinium analogues of green fluorescent protein chromophore: fluorogenic dyes with large solvent-dependent Stokes shift, J. Phys. Chem. lett. **9**, 1958 (2018).
   (Chapter 3).

2. T. Sen, A. V. Mamontova, A. V. Titelmayer, A. M. Shakhov, A. A. Astafiev, A. Acharya, K. A. Lukyanov, A. I. Krylov, and A. M. Bogdanov, Influence of the first chromophore-forming residue on photobleaching and oxidative photoconversion of EGFP and EYFP, Int. J. Mol. Sci. **20**, 5229 (2019).
   (Chapter 4).

3. T. Sen, Y. Ma, I. V. Polyakov, B. L. Grigorenko, A. V. Nemukin, A. I. Krylov, Interplay between Locally Excited and Charge Transfer States Governs the Photoswitching Mechanism in the Fluorescent Protein Dreiklang, J. Phys. Chem. B. **125**, 757 (2021).
   (Chapter 5).

4. E. G. Maksimov, T. Sen, D. V. Zlenko, G. V. Tsoraev, N. V. Pletneva, V. Z. Pletnev, S. A. Goncharuk, K. S. Mineev, A. V. Mamontova, T. R. Simonyan, K. A. Lukyanov, A. M. Bogdanov, A. I. Krylov, BrUSLEE and his shadow: Two persistent excited-state populations within a GFP mutant, in-preparation.
   (Chapter 6).

Chapter 1: Introduction and overview

1.1 Green fluorescent proteins

Capabilities in biological imaging changed dramatically since the first application of GFP to make green-glowing sensory neurons in *C. elegans* in 1994. A genetically encoded fluorescent label for *in vivo* imaging was immediately recognized as a major breakthrough in the domain of cell biology and bioimaging[1]. The unique properties such as low toxicity, ease of use, and the ability to tune its properties by genetic engineering made fluorescent proteins (FPs) powerful tools for *in vivo* observation of protein localization, and interactions, and intracellular pH measurements[2,3]. Two Nobel Prizes in Chemistry (2008 and 2014) emphasize the importance of photophysical properties of FPs.

GFP was first discovered in Pacific Northwest jellyfish *Aequoria Victoria* in 1962[4]. It took more than 30 years to decode GFP gene and to demonstrate that functional GFP can be expressed in various model organisms,[1,5] which opened an era of applications of GFP as a fluorescent label. So far, GFP-like proteins have been found only in multicellular animal species (*Metazoa* kingdom), specifically in hydroid jellyfishes and coral polyps (*phylum Cnidaria*), combjellies (*Ctenophora*), crustaceans (*Arthropoda*), and lancelets (*Chordata*)[6]. Natural GFP-like proteins demonstrate a broad spectral diversity including cyan, green, yellow, orange, and red FPs as well as a colorful palette of non-fluorescent chromoproteins[6] (see Fig.

1.1).

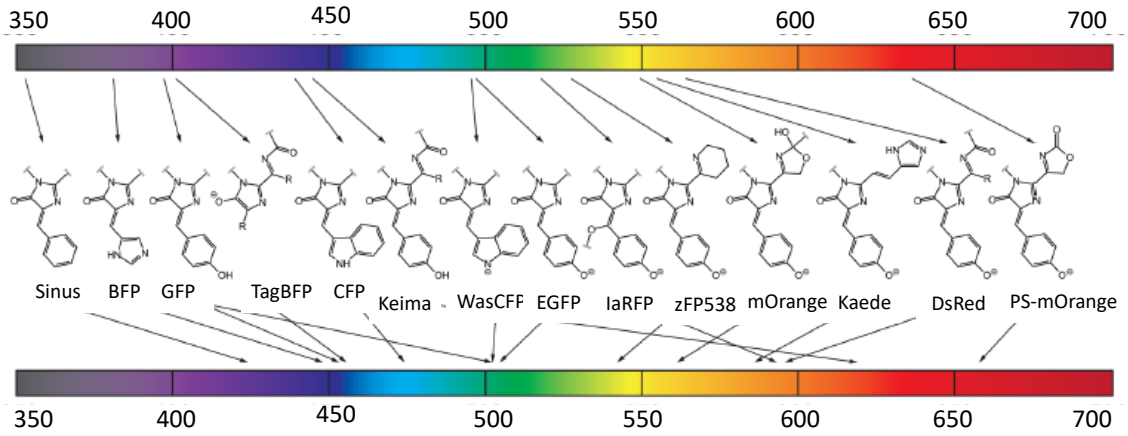


Figure 1.1: Color tuning in fluorescent proteins: Different chemical structures of the chromophore lead to different colors. Main types of chromophore structures are shown together with corresponding excitation (upper bar) and emission (bottom bar) wavelengths designated by arrows. The size of π-conjugated system is particularly important for determining the color: more extensive conjugation leads to red-shifted absorption (compare, for example, blue, green, and red chromophores). Changes in protonation states of the chromophore also affect the energy gap between the ground and the excited states. Excited-state deprotonation of the chromophore is one of the mechanisms of achieving large Stokes shifts. Absorption/emission can be red shifted by π-stacking of the chromophore with other aromatic groups (e.g., tyrosine), as in YFP (not shown). Specific interactions with nearby residues also affect the hue (for example, additional red shift in mPlum fluorescence is attributed to a hydrogen bond formed by acylimines oxygen). [Reproduced from Ref. 7].

As illustrated in Figure 1.1, color is the key aspect of FP's application to bioimaging. Color tuning is achieved by varying the length of the extended π-conjugated system, change in protonation state of the chromophore, and interactions with nearby residues. Another key aspect is brightness, which is highly desirable for labeling. Bright FPs possess large extinction coefficients (EC), and high fluorescent quantum yields (FQY). Properties like phototoxicity and photostability are also important. Due to a wide variety of applications, there is no single

best FP. Depending on the applications, different combinations of properties are desirable.

Consider, for example, photostability. In many applications, bleaching, a gradual loss of optical output upon repeated irradiation, is undesirable. Consequently, protein engineering often aims at more photostable fluorescent proteins. On the other hand, bleaching is exploited in super-resolution imaging[2,24–27]. Methods based on fluorescence loss and recovery are used to trace protein dynamics; photoconversions and photoswitching enable optical highlighting and timing of biochemical processes[23,28] In a similar vein, phototoxicity, which is undesirable for *in vivo* imaging applications, can be exploited in photodynamic therapies and targeted protein/cell inactivation[29].

The photophysics of fluorescent proteins has inspired numerous experimental and theoretical studies[15–23]. However, the details of the photocycle and chromophore formation, the effect of mutations, and the role of the chromophore's surroundings are not fully understood due to the complexity of the system. Molecular-level understanding of these processes provides a crucial advantage in the design of new FPs with properties suitable for particular applications. Knowledge of structure-function relationship and detailed molecular-level mechanistic understanding of the photocycle are essential prerequisites for controlling properties of FPs.

To investigate these properties two strategies have been followed in the last two decades. In the first approach, a particular property is studied across a wide range of FPs and the variations of the property are then rationalized in terms of crystal structures, conjugation in chromophore, local environment, protonation state *etc.* The second approach focuses on one protein and a series of point mutations are introduced to understand the operational mechanisms and properties. For example, scientists have worked on GFP and built several variants of it with

tuned photoproperties. Mutations, leading to different variants may also cause critical structural changes in Fps.

The crystal structure of the wild type GFP (PDB ID: 1W7S)[10] was reported in 1996. It contains 238 amino acids with 11-stranded β-barrel around a single helix. The approximate molar weight of is 25 to 30 kDa. The diameter of the barrel is approximately 24 Å and its height is 42 Å[8,9]. The chromophore resides inside a relatively tight β- barrel. The chromophore is formed by an autocatalytic cyclization of the polypeptide backbone between residue Ser65 and Gly67 and an oxidation of the α-β bond of Tyr66 (SYG) upon protein folding[7]. EYFP, on the other hand, possesses Gly65 instead of Ser65 making the chromophore GYG (see Fig. 1.2).

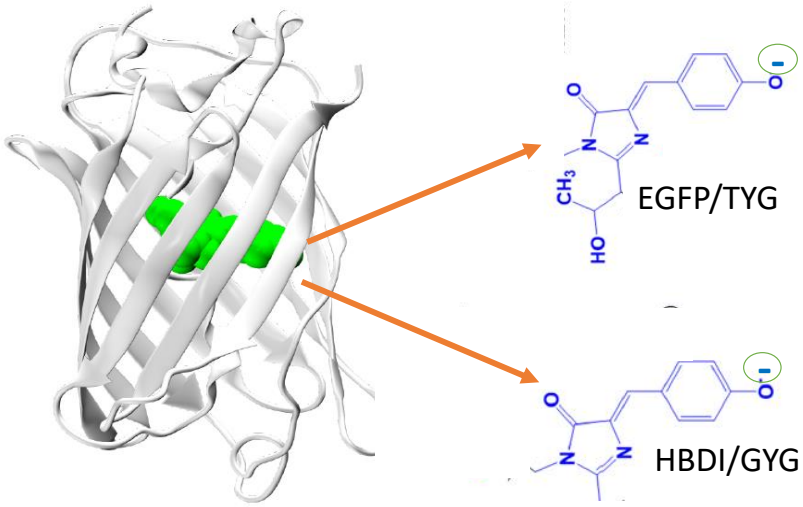


Figure 1.2: Left: A typical structure of a fluorescent protein represented by EGFP. Right: The chromophores in EGFP/GFP-S65T and EYFP/EGFP-T65G (HBDI). Reproduced from Refs. 7 and 40.

In wt-GFP, the chromophore exists in two protonation states, which are in equilibrium with each other: neutral form (A) and anionic form (B). Crystal structure of the GFP-S65T variant (PDB ID: 1EMA)[7] was published almost simultaneously, with resolution of approximately 1.9 Å. The S65T mutation shifts the protonation equilibrium to nearly 100% B-form. This was the first step towards understanding the role of the nearby residues around the chromophore in controlling GFP's photophysical properties.

In 2011, another crystal structure (2Y0G)[30] of a GFP variant was reported with enhanced fluorescent intensity. The enhanced variant, named EGFP, differs from GFP-S65T by the F64L mutation. Being directly connected to the chromophore, residue 64 can be an important element of the local structure. The F64L mutation involves a less bulky side chain, increases van der Waals interaction energy, and results in a tighter packing of the helix during protein folding[30]. A crystal structure (6j6i)[32] of the same variant was published in 2019, with ultrahigh resolution. The paper not only reported an accurate structure, but also analyzed the protonation states of key residues around the chromophore. It is important to understand how these structural changes due to the mutations are related to photoproperties, of FPs.

1.2 Fluorescent protein photocycle

GFP, EGFP, EYFP, Dreiklang, and many other FPs contain a hydroxy benzylidene imidazolone (HBDI)[33,34] type chromophore. HBDI belongs to a class of cyanine dyes, owing to the following structural features: a phenol and imidazolinone moieties connected via a methylene bridge (See Fig. 1.2). This highly conjugated molecule may exist in various protonation states and various resonance structures, depending on the chemical environment around it. The photophysics of the isolated or solvated chromophore is very different from that

of the protein-bound chromophore. The chromophore is nonfluorescent in solution because of reduced excited-state lifetime[35,36] due to fast radiationless decay facilitated by a twisting motion. The rigid protein environment restricts the chromophore's motions in the excited state and limits the accessibility of solvent and other species (oxidants, reductants *etc*) to the chromophore. The ability of FP-chromophores to display different fluorescence in different environments can be utilized in developing fluorogenic and solvatochromic dyes. Fluorogenic dye changes fluorescent intensity upon binding with a target object (becomes fluorescent from nonfluorescent). Solvatochromic dyes, on the other hand, are dyes that change their fluorescent color depending on the solvent.

Figure 1.3 outlines various excited-state processes in FPs. The photocycle is initiated by light absorption producing an electronically excited chromophore. The excited chromophore can decay via several competing relaxation channels from the electronically excited state(s). One of the dominant channels is fluorescence, which restores the ground-state chromophore. The color of emitted light often differs from the absorbed light. This color change, called Stokes shift, arises due to structural relaxation of the chromophore in the excited state, change in hydrogen-bonding network, or excited-state proton transfer (ESPT). For example, in mPlum, a far-red-shifted FPs (RFP), large Stokes shift arises because two different hydrogen-bonding networks around the chromophore in ground state collapse to one in the excited state[38]. ESPT is an established mechanism, operational in GFP, where change in protonation state of the chromophore causes large Stokes shift. In wt-GFP, the neutral chromophore is the dominant form whereas the anionic form becomes dominat in GFP-S65T[39]. Further investigation of the chromophore pocket reveals a hydrogen bonding network (tyrosylO-W-Ser205-Glu222-Ser/Thr66), as shown in Fig. 1.4, plays a key role in transferring the proton from neutral chromophore to the anionic glutamate (GLU), resulting in the anionic chromophore and neutral

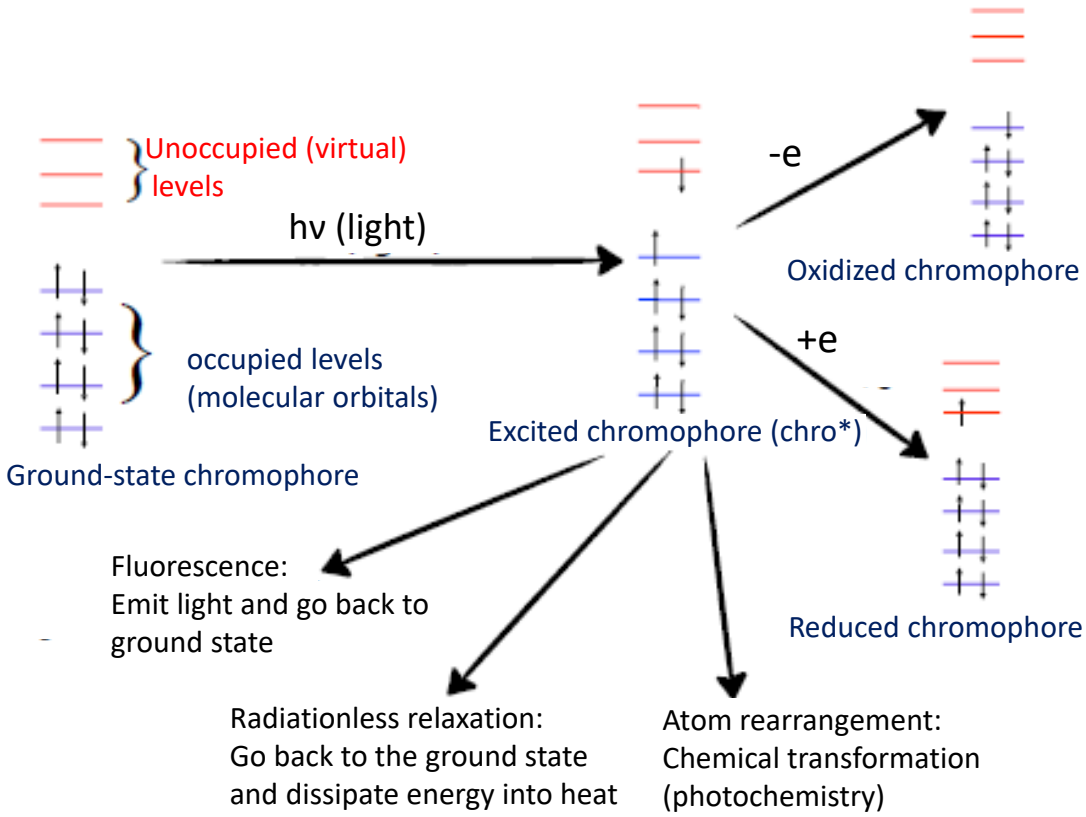


Figure 1.3: Excited-state processes in fluorescent proteins. The main relaxation channel is fluorescence. Radiationless relaxation, a process in which the chromophore relaxes to the ground state by dissipating electronic energy into heat, reduces quantum yield of fluorescence. Other competing processes, such as transition to a triplet state via inter-system crossing (not shown), excited-state chemistry and electron transfer, alter the chemical identity of the chromophore thus leading to temporary or permanent loss of fluorescence (blinking and bleaching) or changing its color (photoconversion). Reproduced from Ref. 7.

glutamic acid (GLUP) via ESPT.[44]

As the bonding pattern changes significantly in electronically excited FPs, the chromophore can readily undergo reactions like ET/CT, often resulting in permanent loss of fluorescence, especially in the presence of external agents. One such photoconversion is oxidative redding in

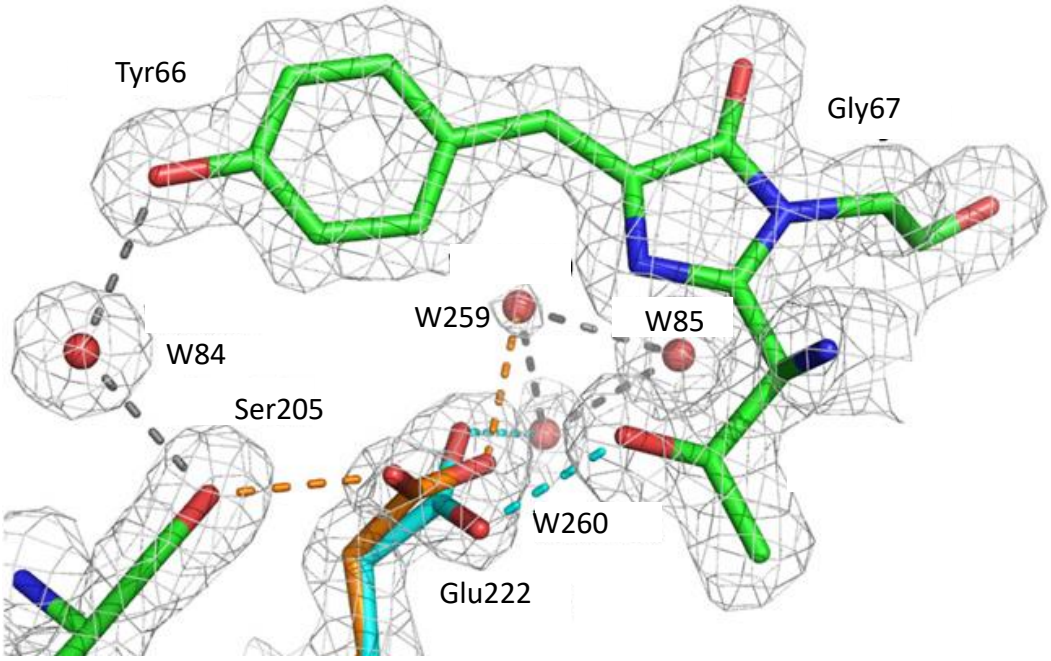


Figure 1.4: Structural analysis of EGFP showing electron density overlaid on the chromophore and the neighbouring residues. The tridentate density around Glu222 clearly indicates alternate conformations of the side chain represented in orange and cyan. hydrogen bonds for each conformation are correspondingly shown in orange or cyan. Glu222 is either hydrogen bonded to Ser205 or to Thr65, but not to both residues at the same time. Reproduced from Ref. 30.

EGFP. Interestingly, in EYFP which differs from EGFP by three mutations T65G, T203Y, and H148L the redding is suppressed. Bogdanov *et al.* found that a π-stacking between Y203 and the chromophore in EYFP increases the oxidation potential of the chromophore[40].

Other processes, which may be involved in the photocycle, are radiationless relaxation, photo-isomerization, and chemical transformations. In many FPs, fluorescence is the main channel, competing with radiationless relaxation. However, the yield of the processes such as bleaching, blinking, photostability, phototoxicity, photoswitching *etc.* are determined by the competition between the main relaxation channels (fluorescence versus radiationless relaxation) and various photoinduced transformations. The timescales of different channels

are key to accurate understanding of yields and branching ratios. These processes are limited by a finite excited-state lifetime, which varies between 1-10 ns in FPs. To play a role in FP photocycle, the excited-state process should have a lifetime comparable with the excited-state lifetime. Below we briefly discuss photoswitching phenomenon and excited-state lifetimes of FPs.

1.3 Excited-state lifetime

The branching ratios of photoinduced processes are determined by excited-state lifetime (which varies between 1-10 ns). Excited-state lifetime determines FQY, relative brightness (RB), relative photostability (RP), *etc*. In the most basic case of a single emissive state, the population of excited fluorophores (Chro*) decays via two competing first-order processes[25]:

$$Chro^* \xrightarrow{k_r} Chro + h\nu, \tag{1.1}$$

$$Chro^* \xrightarrow{k_{nr}} Chro, \tag{1.2}$$

where k_r is the radiative (intrinsic fluorescence) rate constant and k_{nr} describes all quenching channels. The overall decay of the excited-state chromophore is also described by the first-order kinetics with $k = k_r + k_{nr}$ and the corresponding apparent (measured) fluorescence lifetime $\tau = \frac{\ln(2)}{k}$.

It is well established that the dominant radiationless decay pathway is internal conversion (IC). The rate of IC was found to depend weakly on solvent viscosity. Separation of viscosity and thermal effects for HBDI (and a related model compound lacking the hydroxyl group) showed that IC is barrierless at room temperature, but exhibits an apparent activation barrier in rigid media. It was suggested that the coordinate promoting IC must displace only a small

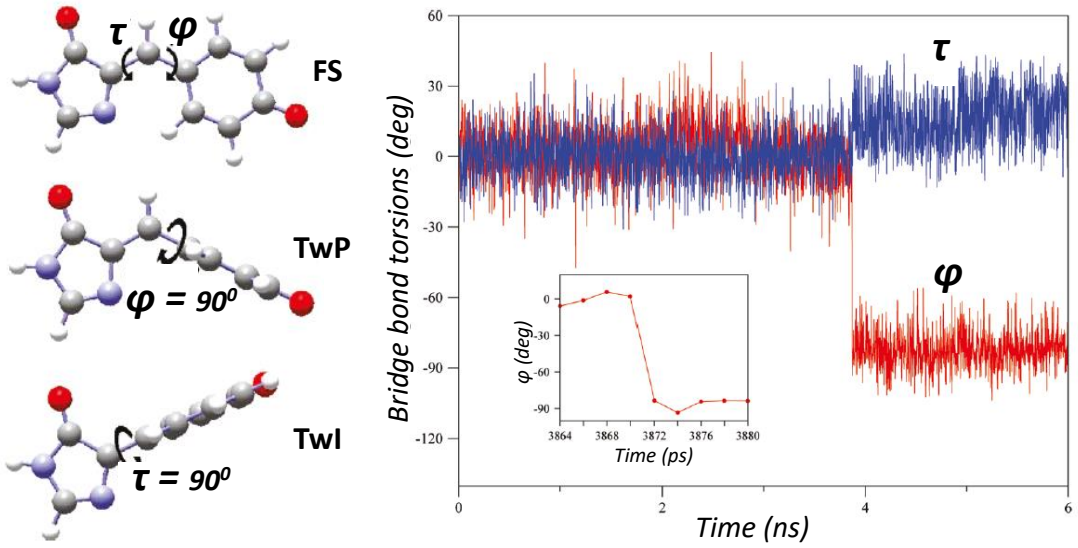


Figure 1.5: Left: The anionic GFP chromophore model p-hydroxybenzylidene-imidazolinone (HBI) in three different geometries: the planar fluorescent state (FS) minimum (τ = ϕ = 0°), the TwP geometry twisted 90° around the phenol bridge bond and the TwI geometry twisted 90° around the imidazolinone bridge bond. Right: Bridge bond torsions (τ and ϕ) from an excited-state MD simulation of the solvated GFP with the anionic chromophore. The inset zooms on the time window where the ϕ twist occurs. Reproduced from Ref. 48.

solvent volume. One possibility is the concerted twist about the bridging C=C-C bonds (hula twist) proposed on the basis of quantum chemical calculations by Weber *et al.*[45] Martinez *et al,* performed high-level QM/MM calculations for the HBDI chromophore and concluded that the bare chromophore is nonfluorescent because of short lifeitme in the excited state in gas phase and in water. Upon excitation, the chromophore undergoes twist around the methylene bridge, relaxing back to ground state, resulting in the loss of fluorescence[35,36].

In GFP-S65T, the radiationless relaxation of the chromophore is restricted due to rigid protein environment. Excited-state simulations revealed twisting around ϕ within a few nanoseconds (0.2-12.9 ns) in all simulations (see Fig. 1.5) , but showed no twisting around τ. This was attributed to the constrain created by the residues connected to the imidazolinone ring.[47,48].

However, the effect of the local environment on nonradiative relaxation is not fully understood. Similarly, the effect of the environment on intrinsic radiative lifetime is not understood.

1.4 Photoswitches

Nonradiative relaxation, leading to long-living dark states, results in loss in fluorescence. The phenomenon enables molecular photoswitching behavior. Switching can be a result of cis-trans isomerization, which is one of the most important photoinduced transformations involved in photoresponse in biological systems (such as rhodopsin). In FPs, this process may lead to reversibly photoswitchable FPs (RS-FPs). The first efficient RS-FP, Dronpa[42,43], was discovered by Ando *et al*. Dronpa absorbs at 503 nm (2.46 eV) and emits at 518 nm (2.39 eV). It converts to the nonfluorescent form upon irradiation at 488 bm (2.54 eV) with poor QY. The reverse process occurs upon illumination with light of 405 bm (3.06 eV). Brakemann *et al,* who investigated the chemistry of photoswitching proposed that a primary factor determining the fluorescence ability of the chromophore in different conformations is its flexibility. The flexibility can be described in terms of tortional angle ϕ and τ around the methylene bridge. The cis-trans isomerization in Dronpa leads to blinking, a temporary loss of fluorescence.

In this thesis, we investigated Dreiklang, a unique photoswitch. In contrast to other photoswitchable proteins, the mechanism in Dreiklang does not involve cis-trans isomerization. Instead, the chromophore undergoes a reversible hydration/dehydration reaction at the imidazolilone ring. Owing to this unique switching mechanism, the wavelengths used for photoswitching and for excitation inducing fluorescence are decoupled in Dreiklang, leading to important advantages for super-resolution microscopy.

Since the first application of FPs, many efforts we made to better understand the effect of mutations on its photophysical properties, such as lifetime and FQY. In collaboration with Bogdanov and his colleagues, we introduced BrUSLEE. It differs from EGFP in three mutations: T65G, Y145M, and F165Y. The BrUSLEE is unique in exhibiting short fluorescence lifetime (820 ps) and relatively high brightness (0.78) with FQY 0.3. This was a result of systamatic structural evolution in preparing and designing new variants in EGFP[51,53]. However, the reason behind such drop in lifetime and FQY is not well understood. The key questions are how does radiative lifetime changes upon mutations, and what is the main channel in nonradiative relaxation. Why such channels are so dominant in BrUSLEE compared to EGFP? We made an effort to find answers to these questions in Chapters 4 and 6 of the thesis. We hope that our findings would be helpful in rational design of new FPs with desirable photoproperties.

1.5 Chapter 1 references

[1] M. Chalfie, Y. Tu, G. Euskirchen, W.W. Ward, D.C. Prasher, Green fluorescent protein as a marker for gene expression, Science **263**, 802 (1994).

[2] K. Nienhaus and G.U. Nienhaus, Fluorescent proteins for live-cell imaging with super-resolution, Chem. Soc. Rev. **43**, 1088 (2014).

[3] H.C. Ishikawa-Ankerhold, R. Ankerhold, and G.P.C. Drummen, Advanced fluorescent microscopy techniques — FRAP, FLIP, FLAP, FRET and FLIM, Molecules **17**, 4047 (2012).

[4] F. H. Jhonson, O. Shimoura, Y. Saiga, L. C. Gershman, G. T. Reynolds, J. R. Waters. Quantum efficiency of Cypridina luminescence, with a note on that of Aequorea, J. Chem. Theory Comput. **60**, 85 (1962).

[5] D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast, M. J. Cormier, Primary structure of the Aequorea victoria green-fluorescent protein, Gene **111**, 229 (1992).

[6] D. M. Chudakov, M. V. Matz, S. Lukyanov, K. A. Lukyanov, Fluorescent proteins and their applications in imaging living cells and tissues, Phys. Rev. **90**, 1103 (2010).

[7] M. Ormö, A.B. Cubitt, K. Kallio, L.A. Gross, R.Y. Tsien, and S.J. Remington, Crystal structure of the aequorea victoria green fluorescent protein, Science **273**, 1392 (1996).

[8] R.Y. Tsien, The green fluorescent protein, Annu. Rev. Biochem. **67**, 509 (1998).

[9] A. Acharya, A. M. Bogdanov, K. B. Bravaya, B. L. Grigorenko, A. V. Nemukhin, K. A. Lukyanov, and A. I. Krylov, Photoinduced chemistry in fluorescent proteins: Curse or blessing?, Chem. Rev. **117**, 758 (2017).

[10] F. Yang, L.G. Moss, G. Larry, G.N. Phillips, The molecular structure of green fluorescent protein, Nat. Bio. **14**, 1246 (1996).

[11] J. S. Remington, Green fluorescent protein: a perspective, Prot. Sci. **20**, 1509 (2011).

[12] N. O. Alieva, K. A. Konzen, S. F. Field, E. A. Meleshkevitch, M. E. Hunt, V. Beltran-Ramirez, D. J. Miller, J. Wiedenmann, A. Salih, M. V. Matz, Diversity and evolution of coral fluorescent proteins, PloS one **3**, e2680 (2008).

[13] A. Salih, A. Larkum, G. Cox, M. Kuhl, O. Hoegh-Guldberg, Fluorescent pigments in corals are photoprotective, Nature **408**, 850 (2000).

[14] S. HD. Haddock, C. W. Dunn, Fluorescent proteins function as a prey attractant: experimental evidence from the hydromedusa Olindias formosus and other marine organisms, Bio. open **4**, 1094 (2015).

[15] M. Zimmer, Green fluorescent protein (GFP): applications, structure, and related photophysical behavior, Chem. rev. **102**, 759 (2002).

[16] S. R. Meech, Excited state reactions in fluorescent proteins, Chem. Soc. Rev. **38**, 2922 (2009).

[17] V. Sample, R. H. Newman, J. Zhang, The structure and function of fluorescent proteins, Chem. Soc. Rev. **38**, 2852 (2009).

[18] T. Van, J. Jasper, Photoreactions and dynamics of the green fluorescent protein, Chem. Soc. Rev. **38**, 2935 (2009).

[19] R. M. Wachter, The family of GFP-like proteins: structure, function, photophysics and biosensor applications. Introduction and perspective, Photochem. photobiol. **82**, 339 (2006).

[20] H. E. Seward, C. R. Bagshaw, The photochemistry of fluorescent proteins: implications for their biological applications, Chem. Soc. Rev. **38**, 2842 (2009).

[21] K. B. Bravaya, B. L. Grigorenko, A. V. Nemukin, A. I. Krylov, Quantum chemistry behind bioimaging: insights from ab initio studies of fluorescent proteins and their chromophores, Acc. Chem. Rev. **45**, 265 (2012).

[22] K. A. Lukyanov, D. M. Chudakov, S. Lukyanov, V. V. Verkhusha, Photoactivatable fluorescent proteins, Nat. Rev. Mol. Cell Biol. **6**, 885 (2005).

[23] K. A. Lukyanov, E. O.Serebrovskaya, S. Lukyanov, D. M. Chudakov, Fluorescent proteins as light-inducible photochemical partners, Photochem. Photobiol. Sci. **9**, 1301 (2010).

[24] M. Hofmann, C. Eggeling, S. Jakobs, S. W. Hell, Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins, Pro. Nat. Acad. Sci. **102**, 17565 (2005).

[25] T. Ha, P. Tinnefeld, Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging, Annu. Rev. Phys. Chem. **63**, (2012).

[26] D. T. Burnette, P. Sengupta, Y. Dai, J. Lippincott-Schwartz, B. Kachar, Bleaching/blinking assisted localization microscopy for superresolution imaging using standard fluorescent molecules, Pro. Nat. Acad. Sci. **108**, 21081 (2011).

[27] P. Sengupta, S. B. Van Engelenburg, J. Lippincott-Schwartz, Superresolution imaging of biological systems using photoactivated localization microscopy, Chem. Rev. **114**, 3189 (2014).

[28] V. Adam, R. Berardozzi, M. Byrdin, D. Bourgeois, Phototransformable fluorescent proteins: Future challenges, Curr. Opinion Chem. Biol. **20**, 92 (2014).

[29] M. E. Bulina, K. A. Lukyanov, O. V. Britanova, D. Onichtchouk, S. Lukyanov, D. M. Chudakov, Chromophore-assisted light inactivation (CALI) using the phototoxic fluorescent protein KillerRed, Nat. Prot. **1**, 947 (2006).

[30] A. Royant, M. Noirclerc-Savoye, Stabilizing role of glutamic acid 222 in the structure of Enhanced Green Fluorescent Protein, J. Str. Biol. **174**, 385 (2011).

[31] J. AJ. Arpino, P. J. Rizkallah, D. D. Jones, Crystal structure of enhanced green fluorescent protein to 1.35 Å resolution reveals alternative conformations for Glu222, PloS one, **7**, e47132 (2012).

[32] K. Takaba, Y. Tai, H. Eki, H.-A. Dao, Y. Hanazono, K. Hasegawa, K. Mikia, and K. Takeda, Subatomic resolution x-ray structures of green fluorescent protein, IUCrJ **6**, 387 (2019).

[33] N. S. Baleeva, M. S. Baranov, Synthesis and Properties of 5-Methylidene-3, 5-dihydro-4H-imidazol-4-ones (Microreview), Chem. Heterocycl. Comp. **52**, 444 (2016).

[34] H. Deng, Y. Su, M. Hu, X. Jin, L. He, Y. Pang, R. Dong, X. Zhu, Multicolor Fluorescent Polymers Inspired from Green Fluorescent Protein. Macromol. **48**, 5969 (2015).

[35] T. J. Martinez, Insights for light-driven molecular devices from ab initio multiple spawning excited-state dynamics of organic and biological chromophores, Acc. chem. res. **39**, 119 (2006).

[36] S. Olsen, K. Lamothe, T. J. Martinez, Protonic gating of excited-state twisting and charge localization in GFP chromophores: a mechanistic hypothesis for reversible photoswitching, J. Am. Chem. Soc. **132**, 1192 (2010).

[37] A. S. Klymchenko, Solvatochromic and Fluorogenic Dyes as Environment-Sensitive, Probes: Design and Biological Applications, Acc. Chem. Res. **50**, 366 (2017).

[38] S. Faraji, A. I. Krylov, On the nature of an extended Stokes shift in the mPlum fluorescent protein, J. Phys. Chem. B. **119**, 13052 (2015).

[39] B. L. Grigorenko, A. V. Nemukin, I. V. Polyakov, D. I. Morozov, A. I. Krylov, First-principles characterization of the energy landscape and optical spectra of green fluorescent protein along the A I B proton transfer route, J. Am. Chem. Soc. **135**, 11541 (2013).

[40] A. M. Bogdanov, A. Acharya, A. V. Titelmayer, A. V. Mamontova, K. B. Bravaya, A. B. Kolomeisky, K. A. Lukyanov, A. I. Krylov, Turning on and off photoinduced electron transfer in fluorescent proteins by π-stacking, halide binding, and Tyr145 mutations, J. Am. Chem. Soc. bf 138, 4807 (2016).

[41] R. B. Vegh, K. B. Bravaya, D. A. Bloch, A. S. Bommarius, L. M. Tolbert, M. Verkhovsky, A. I. Krylov, Chromophore photoreduction in red fluorescent proteins is responsible for bleaching and phototoxicity, J. Phys. Chem. B. **118**, 4527 (2014).

[42] R. Ando, H. Hama, Y. H. Hiroshi, M. Yamamoto-Hino, H. Mizanu, A. Miyawaki, An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein, Proc. Nat. Acad. Sci. **99**, 12651 (2002).

[43] R. Ando, H. Mizanu, A. Miyawaki, Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting, Science **306**, 1370 (2004).

[44] G. Stricker, V. Subramaniam, C. AM. Seidel, A. Volkmer, Photochromicity and fluorescence lifetimes of green fluorescent protein, J. Phys. Chem. B. **103**, 8612 (1999).

[45] D. Mandal, T. Tahara, S. R. Meech, Excited-state dynamics in the green fluorescent protein chromophore, J. Phys. Chem. B. **108**, 1102 (2004).

[46] J. Goedhart, L. Van Weeren, M. A. Hink, N. OE Vischer, K. Jalink, T. WJ. Theodorus, Bright cyan fluorescent protein variants identified by fluorescence lifetime screening, Nat. Methods **7**, 137 (2010).

[47] A. A. Heikal, S. T. Hess, W. W. Webb, Multiphoton molecular spectroscopy and excited-state dynamics of enhanced green fluorescent protein (EGFP): acid–base specificity, Chem. Phys. **274**, 37 (2001).

[48] G. Jonasson, J-M. Teuler, G. Vallverdu, F. Merola, J. Ridard, B. Levy, I. Demachy, Excited state dynamics of the green fluorescent protein on the nanosecond time scale, J. Chem. Theory Comput. **7**, 1990 (2011).

[49] J. Goedhart, D. Von Stetten, M. Noirclerc-Savoye, M. Lelimousin, L. Joosen, M. A. Hink, L. Van Weeren, T. WJ. Gadella, A. Royant, Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%, Nat. comm. **3**, 1 (2012).

[50] A. Svendsen, H. V. Kiefer, H. B. Pedersen, A. V. Bochenkova, L. H. Andersen, Origin of the intrinsic fluorescence of the green fluorescent protein, J. Am. Chem. Soc. **139**, 8766 (2017).

[51] A. V. Mamontova, A. M. Shakhov, K. A. Lukyanov, A. M. Bogdanov, Deciphering the Role of Positions 145 and 165 in Fluorescence Lifetime Shortening in the EGFP Variants, Biomolecules **10**, 1547 (2020).

[52] M. A. Mena, T. P. Treynor, S. L. Mayo, P. S. Daugherty, Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library, Nat. Biotech. **24**, 1569 (2006).

[53] A. V. Mamontova, I. D. Solovyev, A. P. Savitsky, A. M. Shakhov, K. A. Lukyanov, A. M. Bogdanov, Bright GFP with subnanosecond fluorescence lifetime, Scientific reports **8**, 1 (2018).

Chapter 2: Methodology

This Chapter provides a brief overview of the theoretical background and computational protocols used in this thesis; full details can be found in the original papers listed in the Abstract. Here we discuss radiative lifetime and how it can be computed from electronic-structure calculations. The next few sections focus on molecular dynamics protocols and types of QM/MM calculations performed to evaluate properites. We also discuss re-parameterization of ground-state forcefield to describe excited states. Finally, we focus on the free energy calculations for different protonation states.

2.1 Radiative lifetime

The process of radiative decay can be understood classically, by analyzing an oscillating charge model. The classical model is constructed by considering an atom of charge q, mass m in one-dimensional potential well described by Newton's equation of motion:

$$\ddot{\mu} + \gamma\dot{\mu} + {\omega_0}^2\mu = 0, \tag{2.1}$$

where, ω_0^2 = k/m, γ is the damping rate, and μ is electric dipole defined as $\mu = q \cdot r$.

The mechanical energy of an oscillator is

$$\langle W_{osc} \rangle_{cycle} = \left\langle \frac{1}{2} k \left(\frac{\mu(t)}{q} \right)^2 + \frac{1}{2} m \left(\frac{\dot{\mu}(t)}{q} \right)^2 \right\rangle_{cycle} = W_0 e^{-\gamma t}, \tag{2.2}$$

where

$$W_0 = \frac{1}{2} \frac{m\omega_0^2}{q} \mu_0^2. \tag{2.3}$$

The above equation means that the energy of the oscillator decays exponentially with time. Radiative lifetime is then defined as the inverse of the rate of damping rate (τ = 1/γ).

Power radiated by an oscillator is

$$P_{rad} = -\frac{d}{dt} \langle W_{osc} \rangle = W_0 \gamma e^{-\gamma t}, \tag{2.4}$$

and average radiative power, according to classical electromagnetic theory is then:

$$P_{rad} = \frac{\mu_0 \omega_0^4}{12\pi\epsilon_0 c^3} e^{-\gamma t}. \tag{2.5}$$

Comparing Eqs. (4) and (5), we obtain classical expression for radiative lifetime and radiative damping rate

$$\gamma = \frac{1}{\tau} = \frac{1}{6\pi\epsilon_0 c^3} \left(\frac{q^2 \omega_0^2}{m} \right), . \tag{2.6}$$

where, ϵ_0 is dielectric constant, c is the velocity of light in vacuum. In a medium of refractive index (n) classical radiative life time is

$$\tau = \frac{1}{\gamma} = 6\pi\epsilon_0 c'^3 \left(\frac{m}{q^2 \omega_0^2} \right), \tag{2.7}$$

where c′= c/n.

The absorption cross-section is defined as

$$\sigma_{abs} = \frac{q^2}{4mc\epsilon_0} \frac{\gamma}{(\omega - \omega_0)^2 + (\gamma/2)^2}. \tag{2.8}$$

By defining a Lorentzian line shape function

$$L(\omega - \omega_0) = \frac{\gamma/2\pi}{(\omega - \omega_0)^2 + (\gamma/2)^2}, \tag{2.9}$$

where L(ω-ω_0) has been normalized so that $\int_{-\infty}^{\infty}$ L(ω-ω_0)dω = 1, Eq. (8) becomes

$$\sigma_{abs}(\omega) = \frac{\pi q^2}{2mc\epsilon_0} L(\omega - \omega_0). \tag{2.10}$$

Full width at half maxima (FWHM) is given by $\Delta\omega$= γ.

Another way of characterizing the interaction of the dipole with the driving field obtained by integrating the absorption cross section over frequency

$$\int_{-\infty}^{\infty} \sigma(\omega) d\omega = \frac{\pi q^2}{2mc\epsilon_0}. \tag{2.11}$$

The above equation eliminates the dependence on frequency.

$$\int_{-\infty}^{\infty} \sigma(\omega) d\omega = \frac{\pi q^2}{2mc\epsilon_0} = \frac{\pi e^2}{2m_e c\epsilon_0} f_{abs}, \tag{2.12}$$

where f_{abs} is the oscillator strength consisting of a single electron of mass m_e and charge e

$$f_{abs} = \left(\frac{m_e}{m}\right)\left(\frac{q}{e}\right)^2, \tag{2.13}$$

From Eq. (6)

$$\gamma = \frac{1}{\tau} = \frac{e\omega_0^2}{6\pi m_e \epsilon_0 c^3} f_{abs}. \tag{2.14}$$

This equation means that for the transitions at the same frequency, the intrinsic fluorescence lifetime should be inversely proportional to f_{abs}.

The absorption cross section becomes:

$$\sigma_{abs} = \frac{\pi e^2}{2m_e c\epsilon_0} L(\omega - \omega_0) f_{abs}. \tag{2.15}$$

To develop QM treatment of absorption a quantum two-level system in presence of an external electric field, described by the following Hamiltonian:

$$\hat{H}_{tot} = \hat{H}_{atom} + \hat{H}_{field} + \hat{H}_{int}, \tag{2.16}$$

where $\hat{H}_{int}$ accounts for the interaction energy between an atom and the field:

$$\hat{H}_{int} = -\hat{\mu} \cdot \hat{E}. \tag{2.17}$$

This leads to the following expression for radiative γ

$$\gamma = \frac{2\pi}{\hbar} |\mu_{if} \cdot u_e|^2 \rho_f(\hbar\omega_{if}), \tag{2.18}$$

where μ_{if} is the transition dipole moment from state i to state f, u_e is the direction of the polarization of the light, ρ_f is the density of the final states at the frequency of transition.

Starting from that, we arrive at the quantum analogues of the classsical relationships for absorption cross-section and for the relationship between the oscillator strength and radiative lifetime. The absorption cross-section in quantum two level system is

$$\sigma_{abs}(\omega) = \frac{\pi\omega_0|\mu_{if}|^2}{3\hbar c\epsilon_0} L(\omega - \omega_0). \tag{2.19}$$

Introducing the quantum oscillator strength in a direct analogy with Eq. (12)

$$\int \sigma_{abs}(\omega) = \frac{\pi e^2}{2m_e c\epsilon_0} f_{abs}, \tag{2.20}$$

where

$$f_{abs} = \frac{2m_e\omega_0|\mu_{if}|^2}{3\hbar e^2}. \tag{2.21}$$

This quantum oscillator strength is directly computed in quantum-chemical calculations.

The classical treatment of light accounts for only absorption and stimulated emission (rate denoted as B_{if}). Spontaneous emission has be to treated quantum mechanically, which is incorporated in the total rate equation as *ad hoc* (denoted as A_{fi}), i and f denotes initial and final states. The rate of absorption and stimulated emission are proportional to the energy density at transition frequency $\rho_\omega(\omega_0)$. Now we consider again a two-level system with two manifolds of degenerate states of g_i and g_f. In thermodynamic equilibrium, the ratio of population in the two manifolds

$$\frac{N_f}{N_i} = \frac{g_f}{g_i} e^{-\hbar\omega_0/kT}. \tag{2.22}$$

Including the contribution from spontaneous emission, the total rate of change of the upper manifold's population is

$$N_f = [N_f B_{if} - N_i B_{fi}]\rho_\omega(\omega_0) - A_{fi}N_f. \tag{2.23}$$

Using the relation $B_{if} = (g_f/g_i)B_{fi}$ we obtain

$$\frac{A_{fi}}{B_{fi}} = \rho_\omega(\omega_0)e^{(\hbar\omega_0/kT-1)}. \tag{2.24}$$

Spectral density of the Plank radiation is given by

$$\rho_\omega(\omega_0) = \frac{\hbar\omega_0^3}{\pi^2 c^3}\frac{1}{e^{(\hbar\omega_0/kT-1)}}. \tag{2.25}$$

Comparing Eqs. (24) and (25) we obtain

$$\frac{A_{fi}}{B_{fi}} = \frac{\hbar\omega_0^3}{\pi^2 c^3}. \tag{2.26}$$

The Einstein cofficient B_{fl} for absorption and stimulated emission are generalized as

$$B_{fi} = \frac{\pi g_i|\mu_{if}|^2}{3\hbar^2\epsilon_0}. \tag{2.27}$$

Comparing Eqn. (26) and (27)

$$A_{fi} = \frac{1}{\tau} = \frac{g_i\omega_0^3|\mu_{if}|^2}{g_f 3\hbar\pi c^3\epsilon_0} = \frac{g_i\omega_0^2 e^2}{g_f 2\pi c^3 m_e\epsilon_0}f_{abs}. \tag{2.28}$$

This is known as the Strickler-Berg equation[1].

In atomic units, these equations become rather simple. Including the effect of the dielectric environment and by redefining τ as τ_r:

$$\frac{1}{\tau_r} = \frac{\omega_0^2 f_{abs}}{2\pi(c')^3\epsilon}, \tag{2.29}$$

We use this equation to compute radiative lifetime. We take oscillator strength and excitation energies from electronic structure calculations. We follow two different strategies. The first one involves molecular dynamics (MD) simulation in ground state of model structure followed by QM/MM calculations and the results are averaged over snapshots. The other one involves QM/MM optimization of the model structure followed by computation of excitation energy. The second strategy is used when MD fails to predict the best possible model structure.

2.2 MD simulations

In our computational studies, we always begin with MD simulations to carry out equilibrium sampling. We also use MD to model excited-state dynamics.

MD simulations employed the CHARMM27[4,5] parameters for standard protein residues and the parameters derived by Reuter *et al.* for the anionic and neutral GFP chromophores[6]. Parameters for the off-state chromophore in Dreiklang were not available in literature. We obtained them by recognizing the similarity between parts of the chromophore and known amino acids. For example, in the off-state of the chromophore, the parts of the hydrated imidazolione ring can be viewed as combination of proline and threonine. The series of electronic structure calculations allowed us to reparameterize the forcefield for the off-state of the chromophore[7]. Chapter 5 provides the details of this parameterization.

We used the TIP3P[8] water model to describe explicit solvent molecules around the protein. The protein was solvated in a box, producing a water buffer of about 15 Å. The surface charges were neutralized with Na^+ and Cl^- ions at appropriate positions. This was the protocol followed consistently in different proteins that we studied and for different protonation states of the nearby residues. MD simulations were performed with NAMD[9] to generate equilibrated geometries (snapshots) that were used for the subsequent QM/MM calculations. Full details of

MD simulation are provided in Chapters 4, 5, and 6.[7,10]

2.3 Protein structures and protonation states

Assigning the correct protonation states of the residues in proteins is challenging and complicated. Crystal structures provide only an indirect information about the protonation states, unless they are obtained with super-resolution crystallography. Experimental kinetics studies, especially isotope effects and the pH dependence of optical properties, are often used to elucidate protonation states. Several computational methods can be used to evaluate correct protonation states. The most rigorous approach of identifying the most stable form[11,12] is to compute Gibbs free energies of various protonation states. This approach requires extensive thermodynamic averaging. As a shortcut, one can use optimized geometries in different protonation states, assuming that the most stable structure represents the thermodynamically favorable one. This approach accounts for the stabilization provided by hydrogen-bonding[13,14], but ignores entropic effects.

In EGFP, GLU222 and HIS148 are two residues near the chromophore that can exist in different protonation states. GLU can exist in two different protonation states: anionic (GLU) and neutral (GLUP). HIS can exist in three protonation states: neutral HSD (protonated at δ N), neutral HSE (protonated at ϵ N), positively charged HSP (protonated at both N). The structures of these residues are shown in Chapter 6. In Dreiklang, GLU222 and HIS145 are two residues near the chromophore that can exist in different protonation states. Note that the most favorable protonation state may not represent 100% population and different populations may coexist in the protein. This aspect is explored in detail in Chapter 6.

We followed three different approaches to understand protonation state of HIS and GLU in

the FPs we studied (EGFP, EGFP-T65G, EGFP-T65G-Y145M (Duo), EGFP-T65G-Y145M-F165Y (BrUSLEE), EYFP, Dreiklang).

1. Protonation states for all proteins were first checked with *Propka*[15,16] software and then the residues near the chromophore were checked manually. On the basis of pKa predicted by *Propka*, we concluded that GLU222 is in the neutral protonated GLUP form and HIS 148/145 is in the neutral HSD/HSE form (HSD is the most favorable due to a strong hydrogen bond with the chromophore.)

2. We then prepared model systems with different combinations of protonation states and performed MD simulations. Averaged structural parameters can be compared with the crystal structure.The key distances around the chromophores are measured and compared with the crystal structure to figure out the best possible protonation state. The distance cut off was set to 0.5 Å. This analysis is presented in detail in Chapters 4, 5, and 6. However, this analysis did not yield definitive conclusions due to the uncertainty of the forcefield parameters.

3. We used QM/MM optimization with mechanical embedding to optimize the geometry of a region around the chromophore and to compare the relative energies to identify the lowest-energy structure. Then the model structure with the lowest energy was concluded to represent the most probable protonation state.

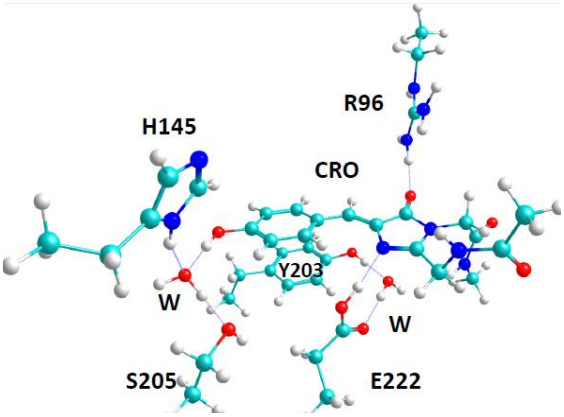


Figure 2.1: Residues included in the extended QM part in the excited-state calculations of ON state of Dreiklang. TYR203 is Y203, GLUP222 is E222, ARG96 is R96. Only the chromophore was kept in small QM region. In the medium QM region ARG96, CRO, HIS145, SER205. TYR203, GLUP222, W were included in the QM region. In extended QM region LEU64, and VAL68 were added. Reproduced from Chapter 5.

2.4 QM/MM optimization

Distance analysis from equilibrated MD trajectories is not always conclusive in determining protonation states of the key residues. Hence, we prepared model systems (as shown in Fig. 6.11) with different combinations of protonation states. QM/MM optimizations were carried out using ONIOM. ONIOM is a mechanical embedding scheme, which describes a model system in two parts X and Y, where X is treated classically and Y is treated quantum-mechanically. The total energy is defined as:

$$E_{tot}(X,Y) = E_{MM}(X) + E_{QM}(Y) + E(X,Y). \tag{2.30}$$

where *E(X, Y)* is intermolecular interaction energy[17–19]. If X and Y are connected via a covalent bond, a link atom (L) is introduced.

$$E_{tot}(X-Y) = E_{MM}(X-Y) - E_{MM}(Y-L) + E_{QM}(Y-L), \tag{2.31}$$

$$E_{MM}(X-Y) = E_{MM}(X) + E_{MM}(X,Y) + E_{MM}(Y), \tag{2.32}$$

$$E_{MM}(Y-L) = E_{MM}(L) + E_{MM}(Y,L) + E_{MM}(Y), \tag{2.33}$$

After simplifying:

$$E_{tot}(X-Y) = E_{MM}(X) + E_{MM}(X,Y) + E_{QM}(Y-L) + E_{LINK}. \tag{2.34}$$

where E_{LINK} denotes link atom correction term and this is given by:

$$E_{LINK} = -E_{MM}(L) - E_{MM}(Y,L). \tag{2.35}$$

In the course of optimization, all coordinates were allowed to relax, except for the positions of link atoms (C-α carbons of the amino-acid residues shown in Fig. 6.11), which were pinned to the positions from the MM-relaxed structures. The QM part was described by ωB97X-D/aug-cc-pVDZ. This functional[20,21] belongs to the family of long-range corrected functionals in which the notorious self-interaction error is greatly reduced. The benchmarks illustrated excellent performance of ωB97X-D for structures and energy differences of a broad range of compounds[20,21]. Optimizations were performed with Q-Chem[22].

2.5 QM/MM protocols for excitation energy

Optimized geometries and equilibrium trajectories were used to compute excitation energy. In these calculations we used electrostatic embedding and described the MM part by point charges. One drawback of this approach is that it neglects the effect of polarization in the MM part. To prevent the overpolarization of the QM part, the charges on the boundary atoms were set to zero. Bonds before -CONH were cut and capped with hydrogen atoms and charge on

CONH was set to be zero; the excess charge was then redistributed over other atoms of the immediate residue to maintain the total charge of the amino acid. This is explained in detail in Chapters 5, and 6.

In geometry optimizations, we used a finite cluster approach (see Chapter 5). To reduce the cost of calculations, a smaller model system was prepared (this does not affect geometry of the QM region in mechanical embedding scheme). We note that this smaller cluster is not sufficient for excitation-energy calculations because the electrostatic effect of the solvated ions and bulk solvent are not taken care of properly. We consider different size of the QM region (small, medium, large; as shown in Fig. 6.11) to check the convergence of electronic properties with respect to the size of the QM region. This approach is called a finite cluster approach. The effect of MM can be added by using an extrapolation scheme as follows:

$$\Delta = \langle E_{ex}(QM/MM) - E_{ex}(QM)\rangle_{MD}, \tag{2.36}$$

$$E_{ex}(QM/MM - corr) = E_{ex}(QM - opt) + \Delta, \tag{2.37}$$

where (E_{ex}(QM/MM-corr) is the extrapolated energy. Accurate computation of excitation energiess and oscillator strengths allow us to evaluate the spectroscopic properties as well as radiative lifetimes. We followed this approach to study the photoswitching mechanism in Dreiklang, which is presented in Chapter 5.

2.6 Nonradiative lifetimes

Excited-state PES of GFP-like chromophores was investigated in several studies by Martinez *et al.*, Jonasson *et al.*, and many others. Electronically excited chromophore can decay via two competing first-order processes: radiative (intrinsic) and nonradiative. Nonradiative

decay can occur through many quenching channels. It is believed that twisting of the GFP-chromophore in the excited state is the dominant non-radiative decay channel. Jonasson *et al.* studied excited-state dynamics of GFP with high-level quantum-chemical calculations[24]. Energy of the S_1 excited state of the anionic chromophore (HBDI) at different geometries was computed to build two-dimensional torsional potential V(τ, ϕ). This study suggested that a twisted structure is energetically favorable in the excited state. The twist around ϕ and (or) τ leads to energy lowering of an isolated chromophore. However, in the protein the twist around τ is restricted due to covalent bonding to nearby residues.

We also optimized the geometry of the HBDI chromophore in first excited state with ωB97X-

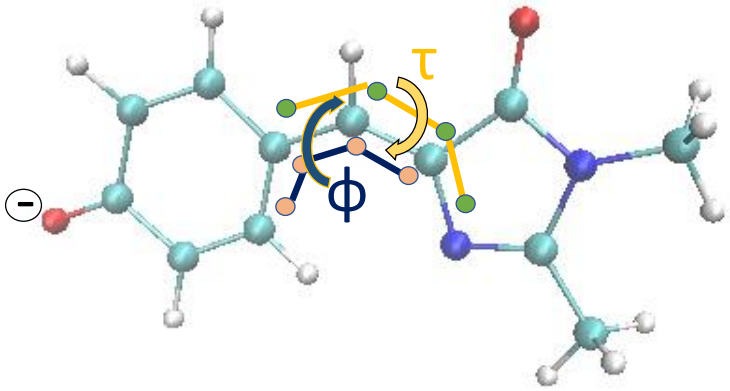


Figure 2.2: Definition of the two torsional angles ϕ and τ describing chromophore twisting. ϕ describes twist around the single bond (phenolate flip) and τ describes twist around the double bond (imidozalinone flip). Reproduced from Chapter 6.

D/aug-cc-pVDZ and obtained geometries twisted around the methylene bridge (see Fig. 2.2) . This is consistent with previous studies[24,25]. Using these structures, we reparameterized the forcefield parameters to describe PES in the excited-state[6]. First, we computed the NBO charges[26] of the HBDI chromophore in the ground and excited states. Partial charges, bond lengths, angles and dihedral angles were computed by performing electronic-structure calculations in the ground and excited state of the bare HBDI chromophore. Details of the protocol is given in Chapter 4.

The force constants (k) are calibrated as:

$$k_{ex-charmm} = \frac{k_{ex-computed}}{k_{gs-computed}} \times k_{gs-charmm}. \tag{2.38}$$

The most important parameter is the torsional angle ϕ. The PES scans show that the chromophore is planar in the ground state and twisted in the excited state. We fitted the excited-state potential with a fitting potential with the calculated force constant, which enables the flip around ϕ. Partial charges and other force-field parameters are listed in Chapter4. Fitting the potential for excited-state PES (right) with respect to ϕ:

$$E = k[1 + cos(n\phi - 180)]; ground - state, n = 2, \tag{2.39}$$

$$E = k[1 + cos(n\phi - 180)]; excited - state, n = 4. \tag{2.40}$$

The major difference in the ground- and excited-state PES (other than force constants) is the change in periodicity (*n*) of the fitting potentials, with much lower value of force constant for the torsional angle ϕ (see Fig. 2.3).

The quality of the reparameterized forcefield was examined by comparing with the AIMD calculations on several snapshots as discussed in Chapter 6. An excellent agreement validated our forcefield parameters. We analyzed the trajectories and the population of planar conformation is fitted to an exponent corresponding to first-order kinetics to evaluate nonradiative lifetime:

$$A(t) = e^{-kt}, \tag{2.41}$$

$$\tau_{nr} = \frac{ln2}{k}, \tag{2.42}$$

where radiationless (non-radiative) half-life is τ_{nr}. The protocol of the analysis is provided in detail in Chapter 4.

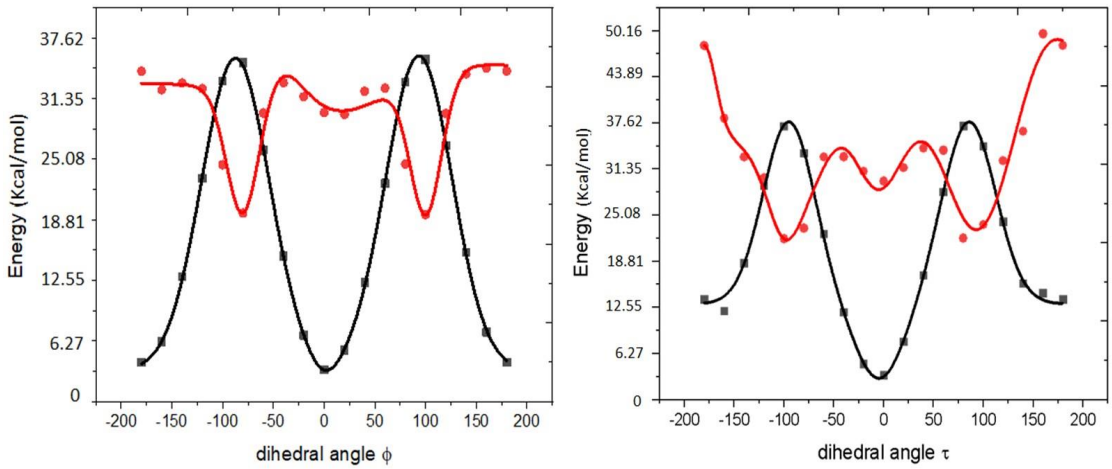


Figure 2.3: Ground- and excited-state torsional potentials for ϕ (twisting of the phenolic ring) and τ (twisting of the imidazolinone ring) of the bare HBDI chromophore. Black dots are *ab initio* calculations whereas red and black lines mark *ab initio* force-field. The barrier heights for twisting along ϕ and τ in the excited state are 3.5 kcal/mol and 3.2 kcal/mol, respectively. The respective ground-state barriers are 32.1 and 34.9 kcal/mol. Reproduced from Chapter 4.

Then we compute apparent excited-state lifetime for each form as:

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}}, \tag{2.43}$$

$$\tau = \frac{\tau_{nr}\tau_r}{\tau_r + \tau_{nr}}, \tag{2.44}$$

$$\tag{2.45}$$

and FQY as:

$$FQY = \frac{\tau_{nr}}{\tau_r + \tau_{nr}}. \tag{2.46}$$

We compute the macroscopic extinction coefficient using the following expression[27]:

$$\epsilon(\tilde{\omega}) = \sum_i \frac{N_a e^2}{4 m_e c^2 \epsilon_0 \ln 10 \sqrt{\pi}} \frac{f_i}{\Gamma} \exp\left[-\left(\frac{\tilde{\omega} - \tilde{\omega}_i}{\Gamma}\right)^2\right], \tag{2.47}$$

where $\epsilon(\tilde{\omega})$ is the molar extinction coefficient in $\mathrm{Lmol}^{-1}\mathrm{cm}^{-1}$; $\tilde{\omega}$ is the excitation wavelength, N_a is the Avogadro number, e is the electron charge, m_e is the electron mass, c is the speed of light in cm s^{-1}, ϵ_0 is the vacuum permittivity in F cm^{-1}, f_i is the oscillator strength of the state i, and Γ is the broadening factor in cm^{-1}. We used wavenumbers, since the units are L$\cdot$ $mol^{-1}\cdot$ cm^{-1} and so Γ is in cm^{-1}. The coefficient is:

$$\frac{N_a e^2}{4m_e c^2 \epsilon_0 \ln 10\sqrt{\pi}} = 1.277 \times 10^8 \ \mathrm{L} \cdot \mathrm{mol}^{-1} \cdot \mathrm{cm}^{-2}. \tag{2.48}$$

The choice of Γ is the biggest uncertainty in the calculations, as we cannot compute it from first principles. In calculation we use Γ= 0.3 eV. Brightness in given by:

$$B = EC \cdot FQY \tag{2.49}$$

Using these expressions we are able to compute all properties that we are interested in EGFP and its mutants.

2.7 Free energies of different protonation states

Although the fluorescence obeys first-order kinetics, in the presence of multiple distinct populations of the fluorophore, the observed decay becomes multi-exponential. In this case, the average lifetime is given by:

$$\langle \tau \rangle = \sum_i A_i \tau_i. \tag{2.50}$$

Multi-exponential fluorescence decay (spectral heterogeneity) arises due to structural heterogeneity, such as different conformations or protonation states of fluorophores, or different local

environments. Hence, calculation of free energy differences between different states is a critical prerequisite in understanding the spectral heterogeneity. Various approaches have been proposed in the literature to compute accurate QM/MM free energies for chemical reactions in solutions[28–33]. In the quantum mechanical free energy (QM-FE) approach by Jorgensen and co-workers,[28,29] a reaction pathway for atoms in QM region is calculated in a vacuum. Free energies for the interaction between the QM and MM atoms are then calculated along the reaction pathway by MM free energy perturbation (FEP) or thermodynamic integration, with electrostatic interactions between the QM and MM atoms are described by point charges.

An alternative approach is the *ab initio* QM/MM apprach (QM(ai)/MM) by Warshel and co-workers.[31,32] The phase space is sampled by MD simulations with a reference potential given by empirical valence bond (EVB) method. We used a quantum mechanical thermodynamic cycle perturbation (QTCP), a combination of QM-FE and Warshel's approach, which employs a thermodynamic cycle (shown in Fig. 6.15) to estimate QM/MM free energy change[33].

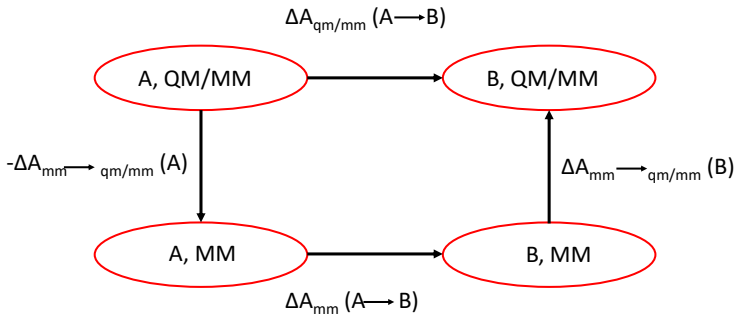


Figure 2.4: The quantum mechanical thermodynamic cycle perturbation (QTCP) method employing a thermodynamic cycle to calculate QM/MM free-energy changes[33].

To compute free-energy differences, we employ the thermodynamic cycle shown in Fig. 6.15. This approach[33], called QTCP, allows one to compute high-level QM/MM free energy changes between two states A and B based on classical (MM) sampling and a relatively modest amount of QM/MM calculations. In this approach, the free energy change between A and B

described by QM/MM is calculated as the sum of three terms: (1) free energy change between A described by MM and by QM/MM (-ΔA$_{mm\rightarrow qm/mm}$ (A)), (2) the free energy change between A and B, with both described by the MM potential (ΔA$_{mm}$(A$\rightarrow$B)), and (3) the free energy change between B described by the MM potential and by QM/MM (ΔA$_{mm\rightarrow qm/mm}$(B)). Hence,

$$\Delta A_{qm/mm}(A \rightarrow B) = -\Delta A_{mm\rightarrow qm/mm}(A) + \Delta A_{mm}(A \rightarrow B) + \Delta A_{mm\rightarrow qm/mm}(B), \quad (2.51)$$

$$\Delta A_{mm}(A \rightarrow B) = -k_B T \ln\langle e^{-[E^{tot}_{mm}(B) - E^{tot}_{mm}(A)]/k_B T}\rangle_{mm,A}, \quad (2.52)$$

$$\Delta A_{mm\rightarrow qm/mm} = -k_B T \ln\langle e^{-[E^{tot}_{qm/mm}(X) - E^{tot}_{mm}(X)]/k_B T}\rangle_{mm,X}, \quad (2.53)$$

Once free energies are computed, one can evaluate the populations of different forms by using the Maxwell-Boltzmann equation:

$$\frac{P_A}{P_B} = e^{-\frac{\Delta A_{qm/mm}(A\rightarrow B)}{k_b T}}. \quad (2.54)$$

We performed a series of QM/MM calculation with ONIOM to evaluate each terms of the equation above. Details of the protocol that we followed are described in Chapter 6. One of the drawbacks of such protocol is that we only sampled in the wells of two protonation states (A and B), but not along the reaction coordinate (as exercised in FEP).

2.8 Chapter 2 references

[1] T. R. Gosnell, Fundamentals of Spectroscopy and Laser Physics, Camb. Univ. Press 3 (2002).

[2] D. B. Hand, The refractivity of protein solutions, J. Biol. Chem. **108**, 703 (1935).

[3] T. L. MacMeekin, M. L. Merton, N. J. Hipp, Refractive indices of amino acids, proteins, and related substances, Advances in Chemistry **44**, 54 (1964).

[4] N. Foloppe and A. D. MacKerell, All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data, J. Comput. Chem. **21**, 86 (2000).

[5] A. D. JR MacKerell, N. Banavali, N. Foloppe, Development and current status of the CHARMM force field for nucleic acids, Biopolymers: Original Research on Biomolecules **56**, 257 (2000).

[6] N. Reuter, H. Lin, and W. Thiel, Green fluorescent proteins: Empirical force field for the neutral and deprotonated forms of the chromophore. Molecular dynamics simulations of the wild type and S65T mutant, J. Phys. Chem. B **106**, 6310 (2002).

[7] T. Sen, Y. Ma, I. V. Polyakov, B. L. Grigorenko, A. V. Nemukin, A. I. Krylov, Interplay between Locally Excited and Charge Transfer States Governs the Photoswitching Mechanism in the Fluorescent Protein Dreiklang, J. Phys. Chem. B. **125**, 757 (2021).

[8] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. **79**, 926 (1983).

[9] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. **26**, 1781 (2005).

[10] T. Sen, A. V. Mamontova, A. V. Titelmayer, A. M. Shakhov, A. A. Astafiev, A. Acharya, K. A. Lukyanov, A. I. Krylov, and A. M. Bogdanov, Influence of the first chromophore-forming residue on photobleaching and oxidative photoconversion of EGFP and EYFP, Int. J. Mol. Sci. **20**, 5229 (2019).

[11] J. Aqvist, A. Warshel, Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches, Chem. Rev. **93**, 2523 (1993).

[12] Y. Sham, Z. Chu, A. Warshel, Consistent calculations of p K a's of ionizable residues in proteins: semi-microscopic and microscopic approaches, J. Phys. Chem. B. **101**, 4458 (1997).

[13] K. B. Bravaya, O. M. Subach, N. Korovina, V. V. Verkhusha, A. I. Krylov, Insight into the common mechanism of the chromophore formation in the red fluorescent proteins: the elusive blue intermediate revealed, J. Am. Chem. Soc. **134**, 2807 (2012).

[14] V. A. Mironov, M. G. Khernova, B. L. Grigorenko, A. P. Savitsky, A. V. Nemukin, Thermal isomerization of the chromoprotein asFP595 and its kindling mutant A143G: QM/MM molecular dynamics simulations, J. Phys. Chem. B. **117**, 13507 (2013).

[15] M. H. M. Olsson, C. R. Sondergaard, M. Rostkowski, and J. H. Jensen, PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions, J. Chem. Theory Comput. **7**, 525 (2011).

[16] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, Nucleic acids research **32**, W665 (2004).

[17] D. Bakowies, W. Thiel, Hybrid models for combined quantum mechanical and molecular mechanical approaches, J. Phys. Chem. **100**, 10580 (1996).

[18] U. Eichler, C. M. Kolmel, J. Sauer, Combining ab initio techniques with analytical potential functions for structure predictions of large systems: Method and application to crystalline silica polymorphs, J. Comput. Chem. **18**, 463 (1997).

[19] A. O. Dohn, Multiscale electrostatic embedding simulations for modeling structure and dynamics of molecules in solution: A tutorial review, Int. J. Quant. Chem. **120**, e26343 (2020).

[20] J.-D. Chai and M. Head-Gordon, Systematic optimization of long-range corrected hybrid density functionals, J. Chem. Phys. **128**, 084106 (2008).

[21] J.-D. Chai and M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom-atom dispersion interactions, Phys. Chem. Chem. Phys. **10**, 6615 (2008).

[22] Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A.T.B.; Wormit, M.; Kussmann, J.; Lange, A.W.; Behn, A.; Deng, J.; Feng, X., et al., Advances in molecular quantum chemistry contained in the Q-Chem 4 program package, Mol. Phys. **113**, 184 (2015).

[23] A. M. Bogdanov, A. Acharya, A. V. Titelmayer, A. V. Mamontova, K. B. Bravaya, A. B. Kolomeisky, K. A. Lukyanov, A. I. Krylov, Turning on and off photoinduced electron transfer in fluorescent proteins by π-stacking, halide binding, and Tyr145 mutations, J. Am. Chem. Soc. **138**, 4807 (2016).

[24] G. Jonasson, J-M. Teuler, G. Vallverdu, F. Merola, J. Ridard, B. Levy, I. Demachy, Excited state dynamics of the green fluorescent protein on the nanosecond time scale, J. Chem. Theory Comput. **7**, 1990 (2011).

[25] S. Olsen, K. Lamothe, T. J. Martinez, Protonic gating of excited-state twisting and charge localization in GFP chromophores: a mechanistic hypothesis for reversible photoswitching, J. Am. Chem. Soc. **132**, 1192 (2010).

[26] E. D. Glendening, C. R. Landis, F. Weinhold, NBO 6.0: Natural bond orbital analysis program, J.Comput. Chem. **34**, 1429 (2013).

[27] M. de Wergifosse, C. G. Elles, A. I. Krylov, Two-photon absorption spectroscopy of stilbene and phenanthrene: Excited-state analysis and comparison with ethylene and toluene, J. Chem. Phys. **146**, 174102 (2017).

[28] J. Chandrasekhar, W. L. Jorgensen, Energy profile for a nonconcerted SN2 reaction in solution, J. Am. Chem. Soc. **107**, 2974 (1985).

[29] J. Chandrasekhar, S. F. Smith, W. L. Jorgensen, Theoretical examination of the SN2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution, J. Am. Chem. Soc. **107**, 154 (1985).

[30] W. L. Jorgensen, Free energy calculations: a breakthrough for modeling organic chemistry in solution, Acc. Chem. Res. **22**, 184 (1989).

[31] R. P. Muller, A. Warshel, Ab initio calculations of free energy barriers for chemical reactions in solution, J. Phys. Chem. **99**, 17516 (1995).

[32] M. strajbl, G. Hong, A. Warshel, Ab initio QM/MM simulation with proper sampling:first principle calculations of the free energy of the autodissociation of water in aqueous solution, J. Phys. Chem. B. **106**, 13333 (2002).

[33] T. H. Rod, U. Ryde, Accurate QM/MM free energy calculations of enzyme reactions: methylation by catechol O-methyltransferase, J. Chem. Theory Comput. **1**, 1240 (2005).

Chapter 3: Pyridinium Analogues of Green Fluorescent Protein Chromophore: Fluorogenic Dyes with Large Solvent-Dependent Stokes Shift

3.1 Introduction

Fluorogenic dyes—compounds that are non-fluorescent in free state but show fluorescence enhancement upon binding with target objects—are very attractive in bioimaging[1], because they can be used in fluorescence microscopy for staining various parts of living systems including proteins[2], nucleic acids[3], and other components[4–6]. Among numerous fluorogenic dyes particularly interesting are structurally modified analogues of the chromophores of the fluorescent proteins from the GFP family[7–9]. These compounds, representing diverse benzylidene imidazolones (BDI)[10–12], have intense and multifarious colors, are small, highly soluble in water, and are easy to synthesize[13]. Despite being highly emissive inside intact proteins, the chromophores have an extremely low fluorescence quantum yield (FQY) in a free state[14], which suggests their potential utility as fluorogens. The low FQY of free chromophores is attributed to the flexibility

of benzylidene moiety[7–9,12]. Immobilizing the chromophores in a rigid matrix results in a several orders of magnitude increase in fluorescence[11,12,15–19]. The applicability of GFP-derived chromophores as fluorogenic labels in living systems has been illustrated, however, they were used to stain RNA[19–23] or proteins[24,25] that were optimized for interaction with specific compounds. Their use for staining the cells' components has been limited to staining all cell membranes indiscriminately[26]. Here we present synthesis, spectroscopic and computational characterization of novel highly fluorogenic pyridinium analogues of the GFP chromophore (Fig. 3.16) designed for measuring local polarity in cells. Their unique properties—large Stokes shifts and large environment-dependent variations of the FQY—enable their use for selective staining of endoplasmic reticulum (ER).

ER plays a key role in cellular metabolism, protein synthesis, and transport of intermediates and signaling molecules. Characterization of the ER structure in living cells is challenging due to a wide three-dimensional interconnected network of flattened, membrane-enclosed sacks or tube-like cisterns and tubules with different thicknesses. A number of fluorescent dyes for imaging ER structure have been reported[27], including commercial dyes ER Tracker^{TM} Red[28], Green[29], and Thermo Fisher Scientific (E34250, E34251, and E12353). However, most of these dyes mainly fill the ER cavities (ER-Tracker Red, Blue, and Green produced by Invitrogen), leaving its membrane structure unknown. Yet, membrane transport and changes in the composition of membranes determine the most important ER functions.

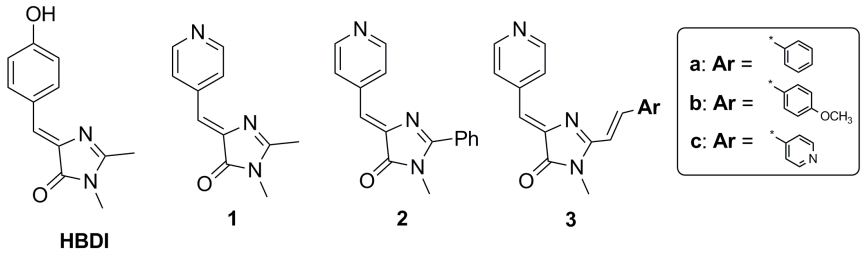


Figure 3.1: HBDI (core of the GFP chromophore) and analogous pyridine chromophores 1-3

Serendipitously, we found that one of the HBDI derivatives containing the pyridinic cycle, compound **1** in Fig. 3.16, shows dramatic solvent-dependent variations of FQY. Such behavior was only observed in 2,5-disubstituted BDIs[30–32]. Since the emission wavelength of **1** is too short for imaging applications, we synthesized several of its analogues featuring extended conjugated π-system, compounds **2** and **3a-c** (Fig. 3.16). All these molecules belong to the class of cyanine dyes owing to their common structural feature, a methyne bridge connecting conjugated aromatic moieties. In HBDI the two aromatic groups are imidozalinone and phenol, whereas in **1**-**3** the phenol ring is replaced by pyridinium. Compounds **1** and **2** were synthesized according to a standard procedure using corresponding carboxymidates[4–6,10,13,33]. Compounds **3a-c** containing an additional double bond were synthesized by condensation of **1** with a range of aromatic aldehydes[34]. The synthetic pathways, compounds properties and characterization are described in Appendix A and B.

3.2 Results and discussion

An important feature of **1**-**3**, which is essential to their use as fluorescent reporters of local polarity, is a strong dependence of their spectroscopic properties on solvent polarity. All synthesized compounds have extremely large Stokes shifts, which increase significantly in polar environment (Table 3.1). The increase in the Stokes shifts is accompanied by a marked decrease of the fluorescence intensity (Fig. 3.2). The absorption maxima in various solvents are very close and the increase in the Stokes shifts is due to bathochromic shifts in emission. As expected, the absorption and emission maxima of **2** and **3** are shifted to longer wavelengths relative to **1** (Figs. 3.2 and S3, Table 3.1). Unfortunately, the FQYs of **3** are low and nearly the same in all solvents, which precludes their use as fluorogenic dyes. In contrast, **2** has the largest variation of FQY: more than two orders of magnitude upon the transition from water to dioxane (Fig. 3.2).

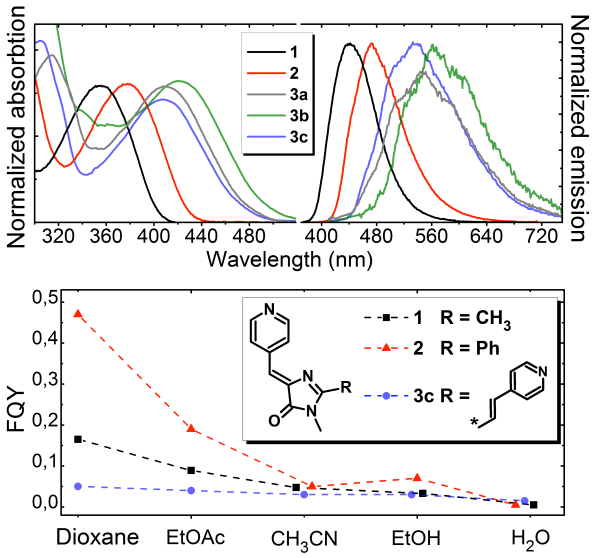


Figure 3.2: Optical properties of compounds 1, 2, and 3c. Top: Absorption and emission spectra in EtOAc. Bottom: FQY in various solvents.

The solvatochromic behavior of **1** and **2** can be described by the Kamlet-Taft model[35], which correlates the spectral shift ν of the solute with the solvent parameters describing its acidic (α), basic (β), and polar (π^*) solvating properties:

$$\nu(\mathrm{cm}^{-1}) = \nu_0 + p\pi^* + a\alpha + b\beta. \tag{3.1}$$

The relative magnitude of solute's parameters p, a, and b reflect the sensitivity of a particular property (e.g., absorption maxima) to solvent polarity, hydrogen-bond donating or accepting abilities, respectively. The results of the analysis are summarized in Appendix (Tables 3.2 and 3.3). An increase in parameter p upon excitation of both compounds suggests a significant increase in the dipole moment, which is typical for other BDIs[24,30,36] and is confirmed by electronic structure calculations. Also in both cases, we observe a change in the parameter a, which indicates changes in proton-accepting properties upon excitation and suggests high photoacidity of the corresponding protonated form[37], which is explained by calculations (Section

Table 3.1: Optical properties of 1, 2, and 3c in various solvents.

Solvent		Abs, nm λ_{max}	Abs, eV E_{ex}	Ext. coeff. $(M \cdot cm)^{-1}$	Em, nm λ_{max}	Em, eV E_{ex}	QY	SS, nm	SS, eV
water	1	347	3.573	10500	464	2.672	0.65	117	0.901
	2	368	3.369	16500	477	2.599	0.69	109	0.770
	3a	408	3.039	11000	560	2.214	0.40	152	0.825
	3b	425	2.197	13500	577	2.149	0.29	152	0.768
	3c	402	3.084	14000	542	2.287	1.3	140	0.797
EtOH	1	351	3.532	11500	453	2.737	3.3	102	0.795
	2	377	3.289	17500	477	2.599	7.3	100	0.690
	3a	411	3.017	10500	550	2.254	9.9	139	0.763
	3b	433	2.863	12000	570	2.175	0.27	137	0.688
	3c	408	3.039	17000	545	2.275	2.8	137	0.764
CH_3CN	1	356	3.483	10000	452	2.743	4.7	96	0.740
	2	377	3.289	16000	475	2.610	4.8	98	0.679
	3a	408	3.039	11000	550	2.254	1.1	142	0.785
	3b	428	2.897	13000	574	2.160	0.26	146	0.737
	3c	408	3.039	15000	541	2.292	3.3	133	0.747
EtOAc	1	353	3.512	11500	442	2.805	8.9	89	0.707
	2	379	3.271	17500	472	2.627	18.6	93	0.644
	3a	408	3.039	10500	544	2.279	1.7	136	0.760
	3b	423	2.931	12000	566	2.191	0.3	143	0.740
	3c	408	3.039	16000	535	2.317	4.1	127	0.722
dioxane	1	357	3.473	11000	439	2.824	16.5	82	0.648
	2	379	3.271	17000	471	2.632	46.6	92	0.639
	3a	411	3.017	10000	546	2.271	2.9	135	0.746
	3b	422	2.938	12500	564	2.198	0.50	142	0.740
	3c	408	3.039	16000	537	2.309	4.8	129	0.730

4.6 of Appendix B). Similar analysis of the FQYs reveals that the increase in the polarity and acidity/basicity of the solvent results in fluorescence quenching. This behavior is typical for many other fluorophores, including GFP chromophore derivatives[26,30–32] and the compounds containing the pyridinium moiety[38]. A likely cause of reduced FQY in polar solvents is partial bond-order flipping associated with strong charge transfer character, which is stabilized in polar media. Changes in bond orders lead to reduced barriers for torsional motion in the excited state thus facilitating radiationless relaxation. However, the behavior of **1** and **2** is different, in particular **1** has an extremely low FQY in hexane. This indicates possible changes in the nature of the excited state and/or in the quenching mechanism. The change in electronic state is confirmed by calculations.

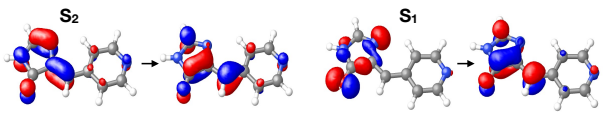


Figure 3.3: NTOs for the two lowest excited states of 1 in gas phase.

To understand the nature of the large Stokes shift and the effect of solvent polarity on the fluorescent properties of the chromophores, we carried out electronic structure calculations. As shown in Appendix, the computed excitation energies correlate reasonably well with the experimental peaks maxima: although the theoretical values of absorption and emission are systematically blue-shifted relative to the experiment, the magnitude of the Stokes shifts is reproduced well by calculations. Fig. 3.17 shows natural transition orbitals (NTOs) corresponding to the two lowest excited states of **1** at the ground-state geometry (see also Fig. 3.17 and 3.18 in Appendix). Importantly, the lowest excited state of **1** is dark at the ground-state geometry; it can be described as an $n \rightarrow \pi^*$ transition. The bright state corresponds to a $\pi \rightarrow \pi^*$ transition; the respective NTOs resemble those in HBDI[36,39]. In all three model compounds, the NTOs of the bright state are localized on the methyne bridge and imidazolone ring (see Fig. 3.18), with only minor contributions from the pyridinium moiety. In compounds **2** and **3c**, the bright state is always the lowest. To assess fluorescent properties of the chromophores, we optimized the structures of the lowest excited state. In isolated chromophore **1**, the structural relaxation does not change the character of the state and S_1 remains dark, which, by virtue of Kasha's rule, means low FQY. The calculations including solvent reveal that while solvent has a small effect on the energies of the excited states at the ground-state geometry (i.e., vertical excitation energies of the S_1 and S_2 states shift by $\sim$0.12 eV), it profoundly affects structural relaxation, leading to the reversal of the state ordering in the polar solvents, such that **1**, which has very low FQY in hexane, becomes fluorescent in polar solvents. This behavior can be explained by the change in the dipole moment in the excited state and by the trends in the transition dipole moments. $\pi \rightarrow \pi^*$ electronic excitation results in a dipole moment increase by 2-3 D, which

further increases upon structural relaxation. The change in dipole moment is associated with the changes in bondlengths, i.e., in the excited state, formally double bonds elongate and formally single bond slightly contract (largest changes occur on the methyne bridge). These trends are well documented[39–41] in GFP-like chromophores and can be explained by the Hückel model[36]. At the ground-state geometry, neither permanent nor transition dipole moments are affected by solvent polarity. However, structural relaxation in polar solvents leads to even higher charge separation in excited state, which results in noticeable solvent-induced variations of respective permanent and transition dipole moments. This explains large solvatochromic shifts in emission that lead to solvent-induced variations in Stokes shifts. Fig. 3.4, which shows the computed Stokes shifts versus the difference of the permanent dipole moment ($\Delta\mu$) in S_1 and S_0, illustrates that the main factor responsible for large solvent-induced variations in the Stokes shift is the change of the dipole moment. The optimized excited-state structures (Figs. 3.19, 3.22, and 3.25 in Appendix) reveal that in polar solvents changes in bondlengths are more pronounced for all 3 model compounds. Significant solvent effect on the shape of excited-state potential energy surfaces has been observed in the HBDI chromophores[40]; in this study[40] the calculations revealed that in the anionic form of HBDI, polar solvent increases changes in bond alternation upon photoexcitation, which is similar to the trends observed here. The changes in bondlengths are related to partial bond-order flipping, in particular, on the methyne bridge. The scans along torsional degrees of freedom in anionic HBDI have shown that the changes in bond alternation are accompanied by the reduced barriers to rotation, which is ultimately responsible for the enhanced radiationless relaxation in polar solvents. Given the similarity in the solvent-induced structural changes in chromophores **1**-**3** and HBDI[40], a similar effect is likely to be operational here (see Appendix D6).

Remarkable solvatochromism of the emission maxima and FQY of **2** suggests that it can be used for measurements of local polarity in living cells or as a fluorogenic dye for labeling cell

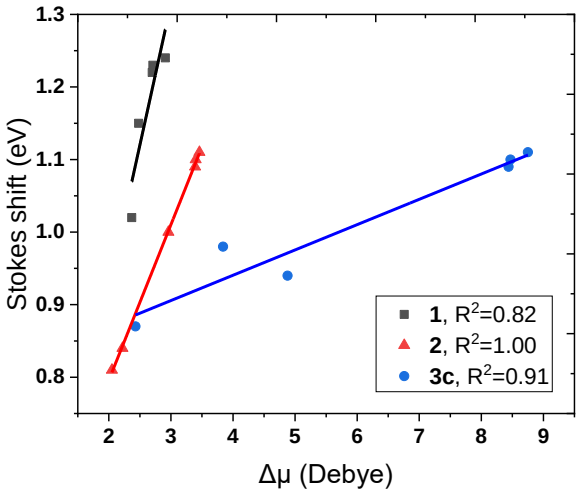


Figure 3.4: Computed Stokes shifts versus the change in permanent dipole moment ($\Delta\mu = \mu(S_1) - \mu(S_0)$).

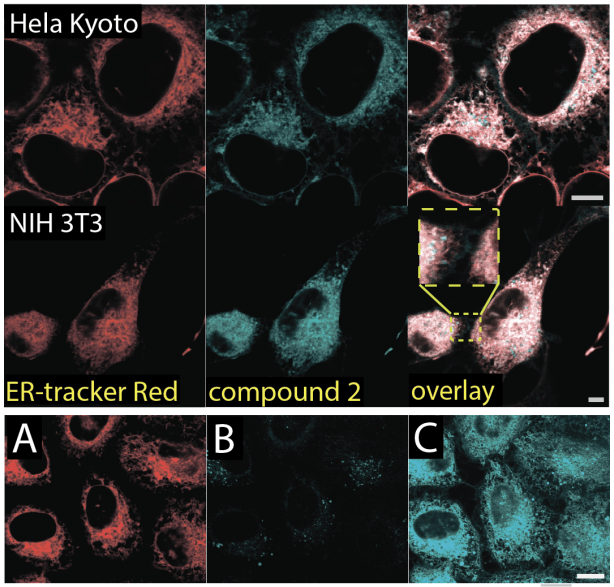


Figure 3.5: Confocal microscopy of the Hela-Kyoto and NIH 3T3 cells labeled with ER-Tracker Red (0.5 mkM) and 2 (5.0 mkM). 559 nm excitation and TRIC for ER-tracker Red and 405 nm excitation and 450-550 nm emission window for 2 with 60X magnification were used. Top: Labeled alive cells; scale 10 mkM. Bottom: Stained cells fixed with formaldehyde right after the fixation (A, B) and after the addition of an extra portion of compound 2 (C); scale 15 mkM.

organelles or lipids. Addition of **2** to cellular media leads to the instantaneous appearance of intense fluorescence inside the cells. This staining is much more selective than with other analogues of the GFP chromophore[26] and the main stained cellular part is ER. Various lipid droplets and the small vacuole-like structure were stained too, but the addition of 20% of Pluronic F-127 in DMSO stock of **2** decreases dramatically the percentage of the stained vacuoles and lipid drops. The addition of 3-20 mkM of **2** from x1000 DMSO-Pluronic stock to Hela-Kyoto and NIH 3T3 cells results in a selective staining of ER, as confirmed by colocalization with 0.5 mkM ER-tracker Red (Invitrogen) (Fig. 3.5, see Section Appendix A2. for more details). A slight difference in the staining was detected in NIH 3T3 fibroblasts, cells having lamelopodia at the head-part of the cell, for which **2** stains additional attached to ER structures or additional ER parts in the plasma membrane region (Fig. 3.5). We assume that these structures can be the thinnest or the most dense ER compartments into which the selected tube-filling ER-tracker penetrates poorly. The fixation of Hela-Kyoto cells stained with 0.5 mkM ER-tracker Red and 5 mkM of **2** using 4% formaldehyde leads to **2** elution, while ER-tracker Red is retained in ER (Fig. 3.5, bottom). However, upon adding another 10 mkM of **2**, the staining of the ER membranes was again observed, but the proportion of non-segregated membranes of vacuole-like structures increased (Fig. 3.5, bottom). Furthermore, when 90% methanol or Triton X-100 was used for cell permeabilization, both dyes were washed out of the cells, presumably due to the destruction of the membranes and the integrity of the membrane cisterns. All this suggests that the observed staining is due to an increase of FQY and accumulation of **2** in the ER membranes.

At the conditions of cell microscopy, **2** is remarkably photostable. We analyzed the performance of **2** in Hela-Kyoto cells compared to the ER-localized blue fluorescent protein BFP-KDEL. The fluorescence of BFP-KDEL was photobleached two-fold in 1.5-2.0 min of 14% 405 nm laser irradiation, while **2** showed no fluorescence decrease at all (Fig. 3.6). The apparent

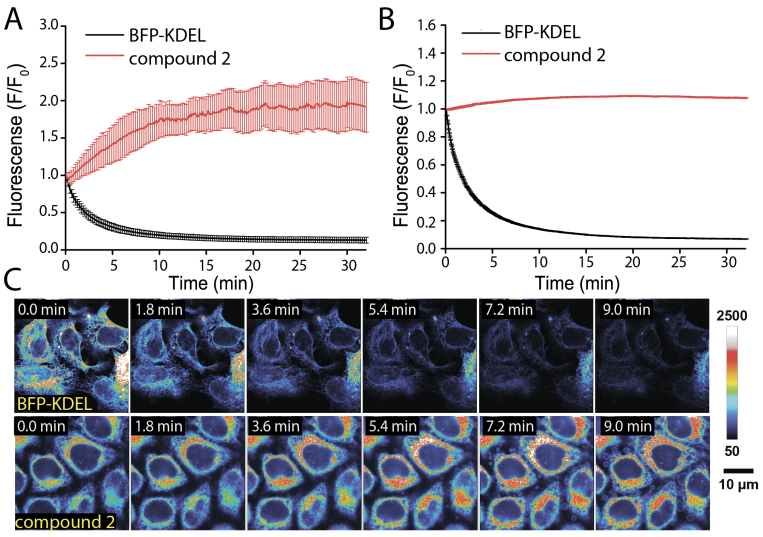


Figure 3.6: Bleaching behavior of Hela-Kyoto cells labeled with 2 and with ER-localized BFP-KDEL protein. Top: Fluorescence intensity of alive (A) and fixed (B) cells in a time-lapse fluorescence microscopy. Bottom: Fluorescent images of alive HeLa cells during the photobleaching.

interminable photostability of **2** suggests its high mobility in cell membranes, causing permanent exchange of bound and free dye molecules from the solution in the field of bleaching, as observed for other fluorogens[20,26]. Note that during a more extended irradiation the fluorescence intensity of **2** only increases (Fig. 3.6A). The effect can be explained by the photodamage of the membranes due to a high laser power and recruiting more dye molecules into damaged membranes from the surroundings. This effect disappears in fixed cells upon decreasing the laser power or exposure to irradiation (Fig. 3.6B). Such a rapid exchange probably indicates that the staining does not occur due to some chemical interactions of **2**. Moreover, the staining is reversible — the replacement of the medium leads to a noticeable weakening of fluorescence, which can be reversed by adding a new portion of the dye. Also, the staining did not result in cells death, which were stable for 10-12 hours (at concentrations 20 mkM), while the test with trypan blue dye did not demonstrate any membrane damage (not shown). We also investigated the pH dependence of optical properties of **2** (Fig. 3.10) and found that its imaging

utility should not be affected by pH-dependent transitions in cells. The pKa of **2** is 3.6, which lies far beyond the physiological range (the value and nature of the optical transition is typical for many other BDIs[42]). Since one of the fluorogenic staining mechanics is the formation of fluorescent agglomerates, we tested **2** for possible aggregation-induced emission and showed that this compound demonstrates no visible emission in crystal solid or in freshly precipitated forms. Thus, we conclude that **2** stains ER membranes because of its unique combination of hydrophobic and fluorogenic properties.

To assess a possibility of using these chromophores in the two-photon excitation regime, we computed[43] 2PA (two-photon absorption) cross sections for several excited states for **1** and **2** and compared it with the prototype HBDI; the results are collected in Table 3.13. The calculations suggest that although **2** is less bright than HBDI in 2PA regime, the cross-sections are sufficiently large for it to be used in two-photon excitation imaging. In contrast, **1** has very bright 2PA transition at 458 nm. If the intensity of this band spills over to longer wavelengths due to inhomogeneous broadening and vibronic interactions (as observed, for example, in a recent study of stilbenes[44]) **1** might actually be brighter than HBDI in two-photon excitation regime.

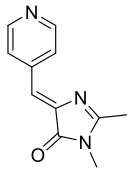
3.3 Conclusion

We presented a novel group of fluorogenic dyes derived from the GFP chromophore. The compounds containing a pyridinium ring in the original chromophore's core feature large solvent-dependent Stokes shifts and solvent-induced variations in the FQY. The calculations explain the observed trends in terms of the increase of the dipole moment upon excited-state relaxation in polar solvents, which is associated with the changes in bondlengths and partial bond-order flipping in the excited state. A unique combination of such optical characteristics and lipophilic properties enables using one of the new dyes for the ER staining. Owing to its

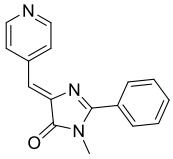
extremely high photostability (ensured by a dynamic exchange between the free and absorbed compound's states) and selectivity (demonstrated by several examples), in combination with pH-independence in the physiological range, **2** is a promising label for this type of cellular organelles.

3.4 Appendix A: Experimental details

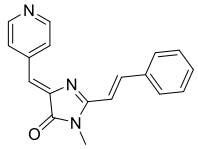
3.4.1 Appendix A1: Materials and methods



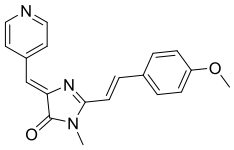
Compound **1**:
(Z)-4-((Pyridin-4-yl)methylene)-1,2-dimethyl-1H-imidazol-5(4H)-one. Yellow solid, 1.5 g (73%), mp = 175-178°C.



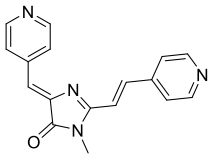
Compound **2**:
(Z)-4-((Pyridin-4-yl)methylene)-1-methyl-2-phenyl-1H-imidazol-5(4H)-one. Yellow solid, 1.7 g (65%), mp = 190-192°C.



Compound **3a**:
(Z)-4-((Pyridin-4-yl)methylene)-1-methyl-2-(E)-styryl-1H-imidazol-5(4H)-one.Yellow solid, 94 mg (33%), mp = 201-203°C.



Compound **3b**:
(Z)-4-((Pyridin-4-yl)methylene)-1methyl-2-((E)-2-(4-methoxyphenyl)vinyl)-1H-imidazol-5(4H)-one. Yellow solid, 67 mg (21%), mp = 223-225°C.



Compound **3c**:
(Z)-4-((Pyridin-4-yl)methylene)-1-methyl-2-((E)-2-(pyridin-4-yl)vinyl)-1H-imidazol-5(4H)-one. Yellow solid, 100 mg (35%), mp∼255°C with decomposition.

Figure 3.7: Structures and properties of compounds 1-3.

Commercially available reagents were used without additional purification. E. Merck Kieselgel 60 was used for column chromatography. Thin layer chromatography (TLC) was performed on silica gel 60F_{254} glass-backed plates (MERCK). Visualization was effected by UV light (254 or 312 nm) and staining with $KMnO_4$. NMR spectra were recorded on a 700 MHz Bruker Advance III NMR at 293 K, 800 MHz Bruker Advance III NMR at 333 K, and Bruker Fourier 300. Chemical shifts are reported relative to the residue peaks of $CDCl_3$ (7.27 ppm for ^{1}H and 77.0 ppm for ^{13}C) or DMSO-d_6 (2.51 ppm for ^{1}H and 39.5 ppm for ^{13}C). Melting points were measured on an SMP 30 apparatus. High-resolution mass spectra (HRMS) spectra were recorded on AB Sciex TripleTOF 5600+ equipped with a DuoSpray (ESI) source.

General method for the preparation of (Z)-4-((Pyridin-4-yl)methylene)-1-methyl-1H-imidazol-5(4H)-ons (1,2). The corresponding aromatic aldehyde (10 mmol) was dissolved in $CHCl_3$ (50 mL) and mixed with methylamine solution (40% aqueous, 2.5 mL) and anhydrous Na_2SO_4 (10 g). The mixture was stirred for 48 h at room temperature, filtered, and dried over the additional Na_2SO_4. The solvent was evaporated, the ethyl((1-methoxy)amino)acetate or benzoate (20 mmol) was added and the mixture was stirred for 24 h at room temperature. The mixture was dried in vacuum and the product was purified by column chromatography ($CHCl_3$/EtOH 100:1).

General method for the preparation of (Z)-4-((Pyridin-4-yl)methylene)-1-methyl-2-aryl-1H-imidazol-5(4H)-ones (3). The product of previous stage 1 (1 mmol) was dissolved in dioxane (20 mL), $ZnCl_2$ (2 g) and the corresponding aldehyde (20 mmol) were added and the reaction mixture was heated to 80°C for 2-20 h. The mixture was cooled and dissolved with EtOAc (100 mL) washed by $NaHPO_4$ solution (5%, 2×50mL), EDTA solution (5%, 2×50 mL), water (2×50 mL), and brine (2×50 mL) and dried over Na_2SO_4. The solvent was

evaporated and product **2** was purified by column chromatography ($CHCl_3$/EtOH 100:2).

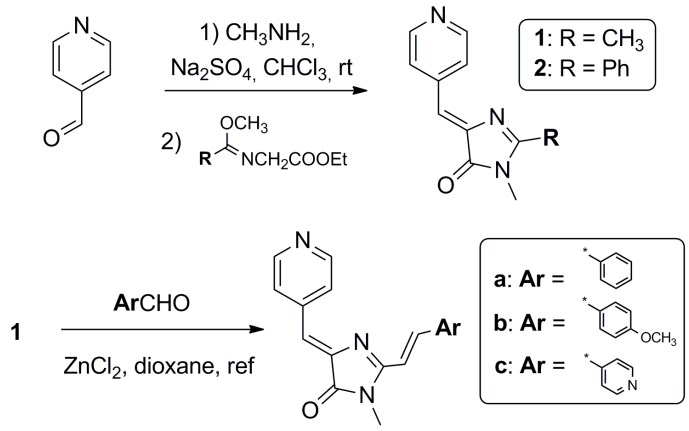


Figure 3.8: Synthesis of compounds 1-3.

3.4.2 Appendix A2: Fluorescent imaging in cells

Fluorescent imaging of non-transfected cells

HeLa-Kyoto and NIH 3T3 cells (the both from EMBL collection) were seeded into 35-mm glass bottom dishes (MatTek Corporation) and cultured in DMEM (Hela-Kyoto) or RPMI (NIH 3T3) media with 10% FBS, 20 mM potassium pyruvate at 37°C in a 5% CO_2 atmosphere. After the 24-48 hours cells were placed into an environmental chamber in 2 mL of Hank's Balanced Salt Solution with calcium and magnesium (HBSS/Ca/Mg, Gibco cat.# 14025-092) supplemented with 20 mM HEPES (pH 7.2) and 20 mM glucose at 37°C. Afterwards cells were incubated with the 5.0 μM of compound **2** (added from 5 mM stock in 80% DMSO and 20% Pluronic F-127 (Thermo Fisher Scientific cat.# P3000MP)) and/or 0.5 μM ER-tracker^{TM} Red (added from 1 mM stock Invitrogen cat. #E34250) for 30 min. Images were captured

using Olympus FluoView 1200 confocal Microscope at 559 nm excitation and TRIC for ER-tracker^{TM} Red and 405 nm excitation and 450-550 nm emission window for compound **2** channel using 60X magnification. Images were processed and analyzed for co-localization in Image J.

Cells fixation and fixed-cells imaging

HeLa-Kyoto cells were grown in 35-mm glass bottom dishes and stained by compound **2** and/or 0.5 μM ER-tracker^{TM} Red as described above. Afterwards cells were fixed with 4% formaldehyde, washed 3 times with 1 ml of PBS and imaged. Images were captured using Olympus FluoView 1200 confocal Microscope at 559 nm excitation and TRIC for ER-tracker^{TM} Red and 405 nm excitation and 450-550 nm emission window for compound **2** using 60X magnification (Fig. 5 bottom, A and B). Afterwards 10 mkM of compound **2** was added to cell media, and new image of compound **2** staining was obtained (Fig. 5 bottom, C).

Photobleaching analysis

We conducted a bleaching analysis of compound **2** in Hela-Kyoto cells compared to the ER-localized blue fluorescent protein BFP-KDEL (Addgene #49150). HeLa-Kyoto cells were seeded into 35-mm glass bottom dishes (MatTek Corporation) and cultured in DMEM (Hela-Kyoto) or RPMI (NIH 3T3) media with 10% FBS, 20 mM potassium pyruvate at 37°C in a 5% CO_2 atmosphere. After 24-36 h, cells were transfected by a mixture of 1 ng DNA and 3 mkL FuGene HD transfection reagent in 100 mkL OptiMEM solution per one dish. After 14 h, cell medium was replaced by 2 mL fresh medium. For imaging 24-48 hours after transfection cells were placed into an environmental chamber in 2 mL of Hank's Balanced Salt Solution with calcium and magnesium (HBSS/Ca/Mg, Gibco cat.# 14025-092) supplemented with 20 mM HEPES (pH 7.2) and 20 mM glucose at 37°C. After that cells were imaged using a Olimpus

FluoView 1200 confocal microscope equipped with a 60X oil objective using 405 nm excitation and 450-550 nm emission window. Images were processed and analyzed for co-localization in Image J. All graphs were processed in OriginPro8.1 (OriginLab).

3.5 Appendix B: Experimental results

3.5.1 Appendix B1: Solvatochromic properties of compounds 1-3

UV-VIS spectra were recorded with a Varian Cary 100 spectrophotometer. Fluorescence excitation and emission spectra were recorded with Agilent Cary Eclipse fluorescence spectrophotometer. Table 1 of the paper shows optical properties of compounds **1**-**3** in various solvents.

3.5.2 Appendix B2: Solvatochromic analysis of absorption and emission spectra of compounds 1 and 2

Kamlet-Taft's model[35] correlates the spectral shift ν of the solute with the solvent parameters that are responsible for the acidic (α), basic (β), and polar (π^*) solvating properties:

$$\nu(\mathrm{cm}^{-1}) = \nu_0 + p\pi^* + a\alpha + b\beta \tag{3.2}$$

Table 3.2 summarizes parameters π, α, and β for various solvents[35] and absorption/emission maxima and fluorescence quantum yields of **1** and **2** in various solvents. Table 3.3 presents Kamlet-Taft-type linear regression analyses.

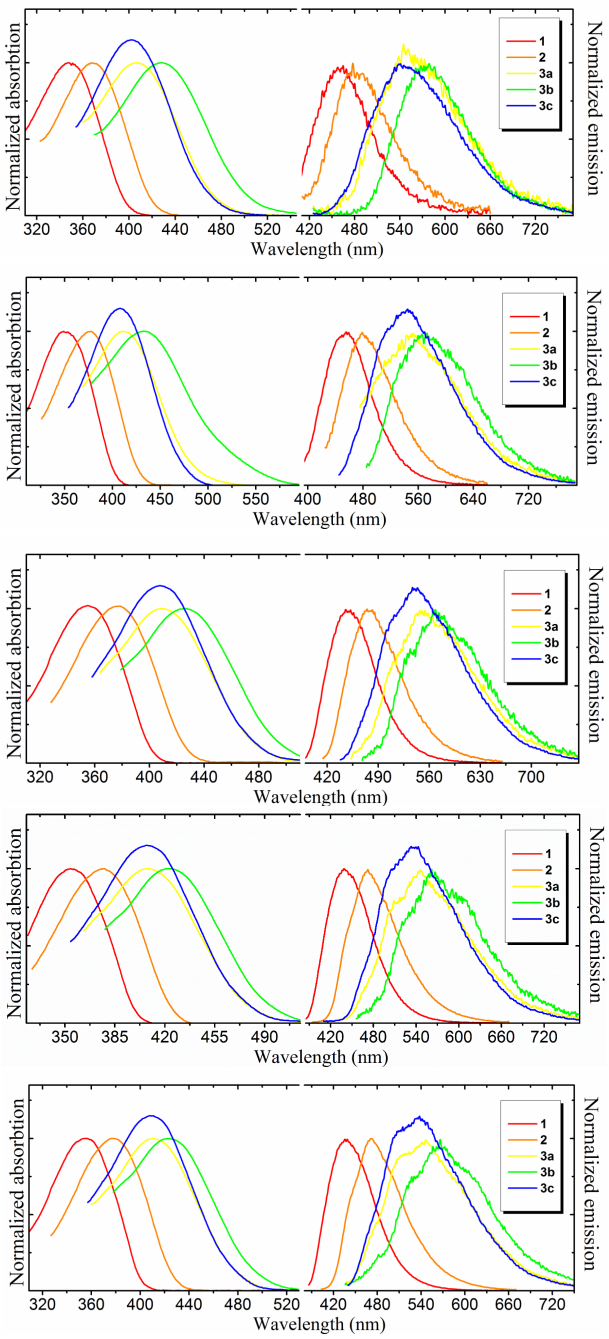


Figure 3.9: From top to bottom: Fluorescence and absorption spectra of 1-3 in water, ethanol, acetonitrile, actetate, and dioxane.

Table 3.2: Kamlet-Taft's parameters and absorption/emission maxima (in nm) and fluorescence quantum yields (in %) of 1 and 2 in various solvents.

Solvent				**1**			**2**		
	π^*	β	α	λ^{abs}	λ^{em}	FQY	λ^{abs}	λ^{em}	FQY
Et_2O	0.24	0.47	0	351	434	6.06	376	470	26.27
EtOH	0.54	0.77	0.83	351	453	3.27	377	477	7.27
EtOAc	0.45	0.45	0	353	442	8.90	379	472	18.64
MeOH	0.60	0.62	0.93	349	458	1.75	375	482	3.68
ACN	0.66	0.31	0.19	356	452	4.72	377	475	4.76
CH_2Cl_2	0.73	0	0.3	355	438	8.50	378	472	16.76
DMF	0.88	0.69	0	361	455	4.32	382	481	6.83
DMSO	1.00	0.76	0	362	458	0.55	383	485	2.37
Water	1.09	0.4	1.17	347	464	0.65	366	477	0.69
Acetone	0.62	0.48	0.08	355	443	4.67	378	476	11.24
THF	0.55	0.55	0	359	445	6.47	381	476	43.35
Hexane	0	0	-0.04	353	416	0.008	374	458	54.27
Toluene	0.49	0.11	0	356	437	12.97	381	476	53.12
Dioxane	0.49	0.37	0	357	439	16.46	379	471	46.61
PY	0.87	0.64	0	360	454	5.00	378	483	23.58

Table 3.3: Solvatochromic spectral parameters (in 10^3/cm^{-1}) of 1 and 2.

		a	b	p	ν_0/FQY$_0^a$	R
1	Abs	0.8	- 0.2	-0.6	28.5	0.92
	Em	-0.4	-1.0	-1.3	23.8	0.97
	FQY^b	0	-0.1	0	0.09	0.50
	FQY^c	0	-0.1	-0.1	0.16	0.80
2	Abs	0.6	-0.4	-0.1	26.6	0.80
	Em	0.0	-0.5	-0.7	21.7	0.91
	FQY^b	-0.2	-0.2	-0.3	0.53	0.79

[a] ν in cm^{-1} and FQY in %. For FQY the corresponding coefficients were calculated using multivariative linear regression analogously to Kamlet-Taft's equation.
[b] All solvents.
[c] The data for hexane excluded.

3.5.3 Appendix B3: pH-titration of compound 2

pKa values of protonated forms of compound **2** were measured by titration of 15 μM solution in water. Absorption spectra of **2** in water at various pH and its titration curves are shown in Fig. 3.10.

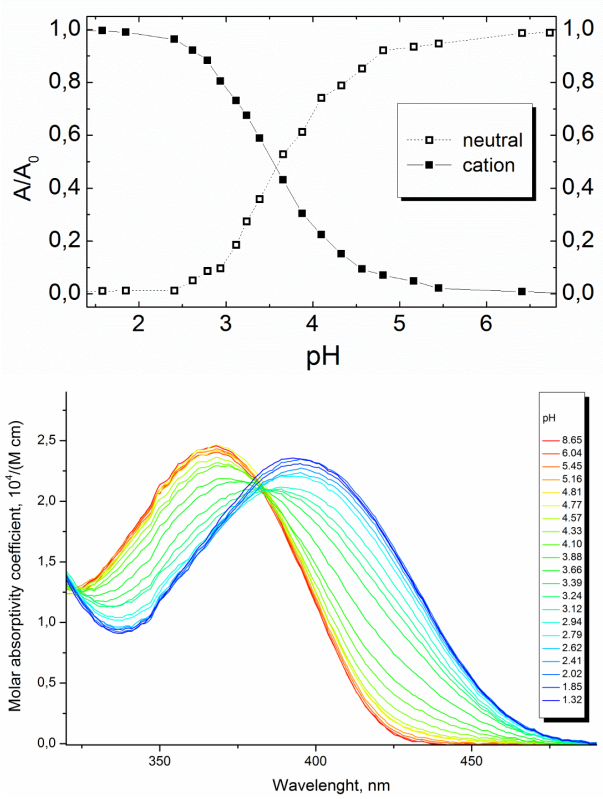


Figure 3.10: Left: pH-titration of compound 2. Right: Absorption spectra of 2 at different pH values. Neutral: λ^{abs}=368 nm; Cation: λ^{abs}=395 nm; pKa(Abs)=3.6.

3.5.4 Appendix B4: ^{1}H and ^{13}C NMR spectra

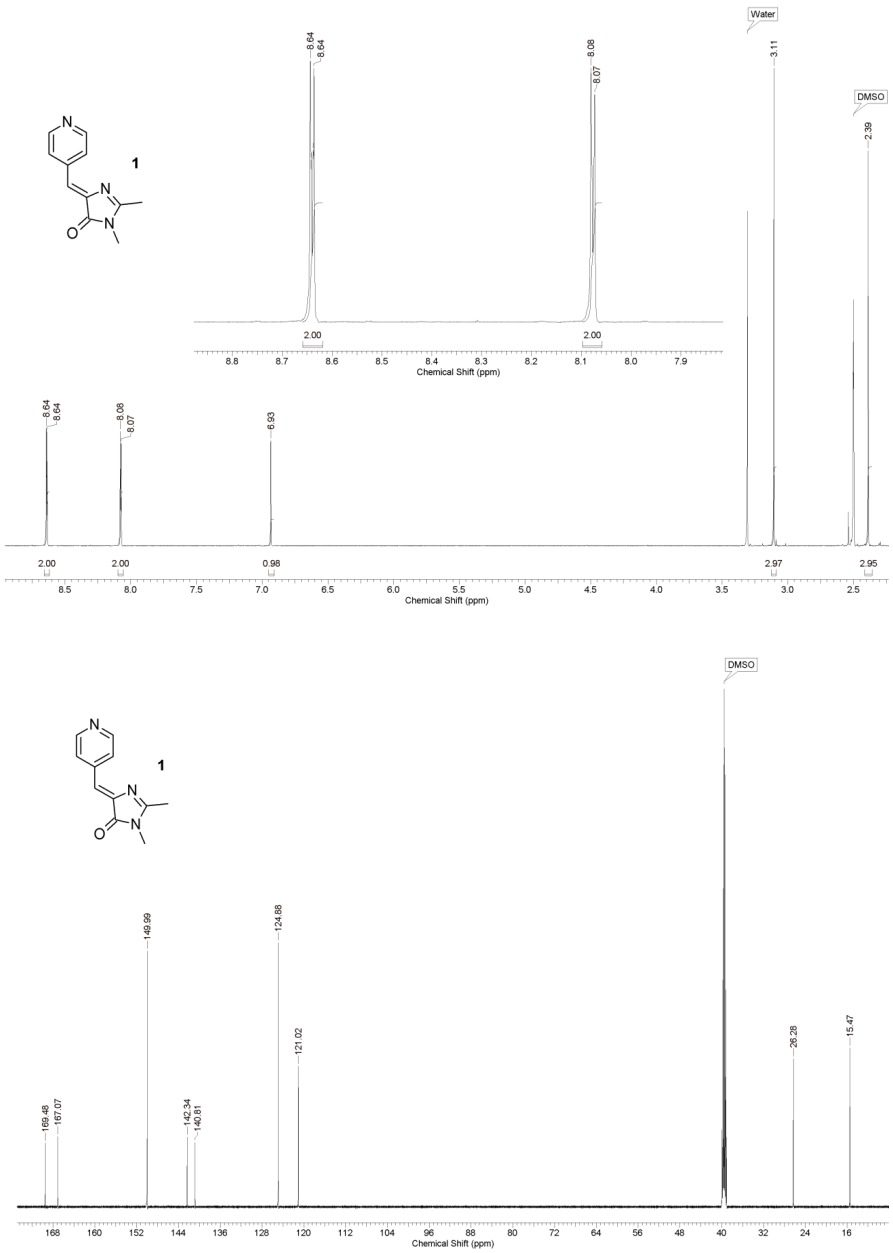


Figure 3.11: Compound 1: ^{1}H NMR (800 MHz, DMSO-d_6) δ=8.64 (d, J=5.9 Hz, 2 H, Ar), 8.08 (d, J=6.11 Hz, 2 H, Ar), 6.93 (s, 1 H, Ar-CH), 3.11 (s, 3 H, CH_3), 2.39 (s, 3 H, CH_3); ^{13}C NMR (176 MHz, DMSO-d_6) δ = 15.5, 26.3, 121.0, 124.9, 140.82, 142.3, 150.0, 167.1, 169.5; HRMS (m/z) calc-d. $C_{11}H_{12}N_3O$ for $[M+H]^+$ 202.0975, found 202.0978.

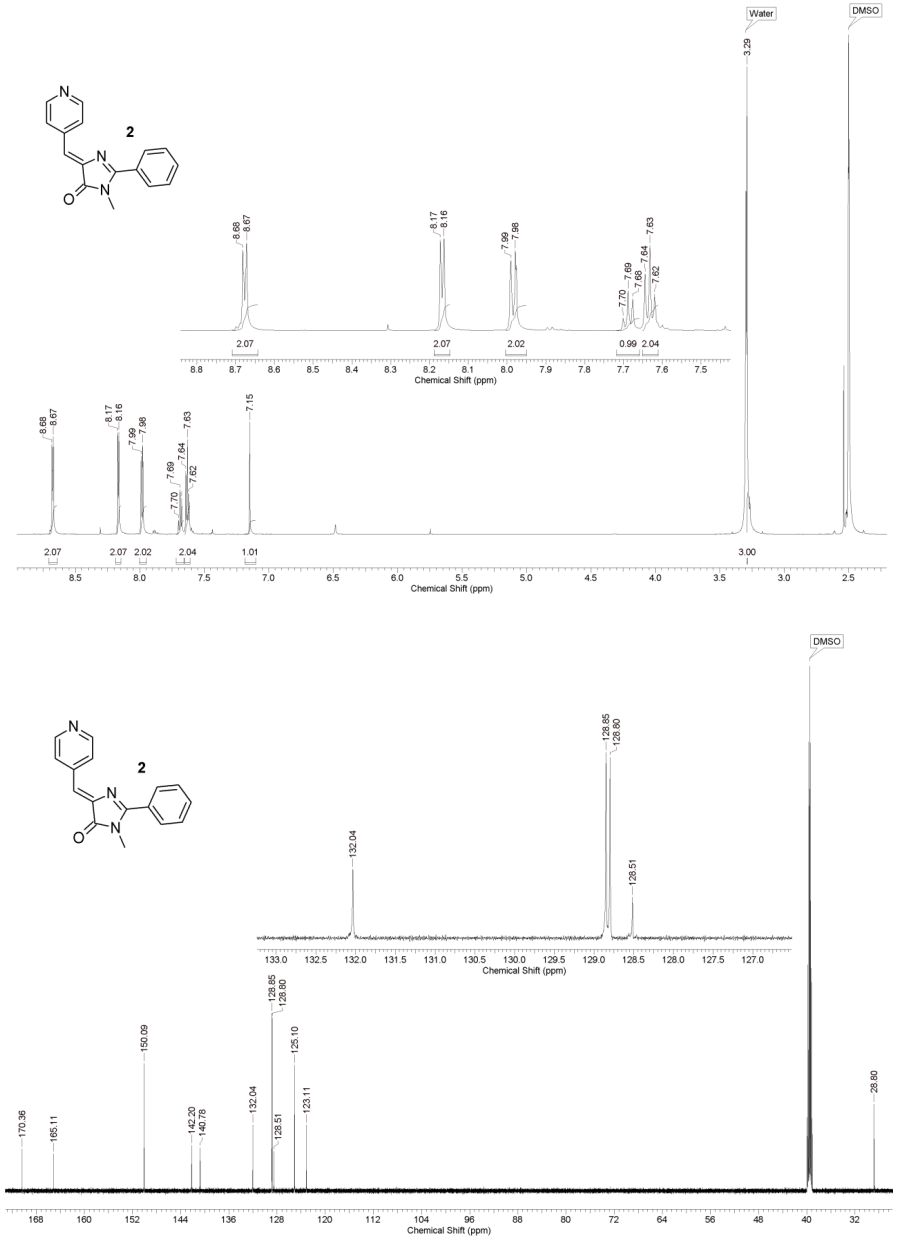


Figure 3.12: Compound 2: ^{1}H NMR (700 MHz, DMSO-d_6) δ=8.68 (d, J=6.2 Hz, 2 H, Ar), 8.17 (d, J=5.9 Hz, 2 H, Ar), 7.98 (d, J=6.9 Hz, 2 H, Ar), 7.69 (t, J=7.3 Hz,1 H, Ar), 7.63 (t, J=7.5 Hz, 2 H, Ar), 7.15 (s, 1 H, Ar-CH), 3.29 (s, 3 H, CH_3); ^{13}C NMR (176 MHz, DMSO-d_6) δ = 28.8, 123.1, 125.1, 128.5, 128.8, 128.9, 132.0, 140.8, 142.2, 150.1, 165.1, 170.4; HRMS (m/z) calc-d. $C_{16}H_{14}N_3O$ for $[M+H]^+$ 264.1131, found 264.1135.

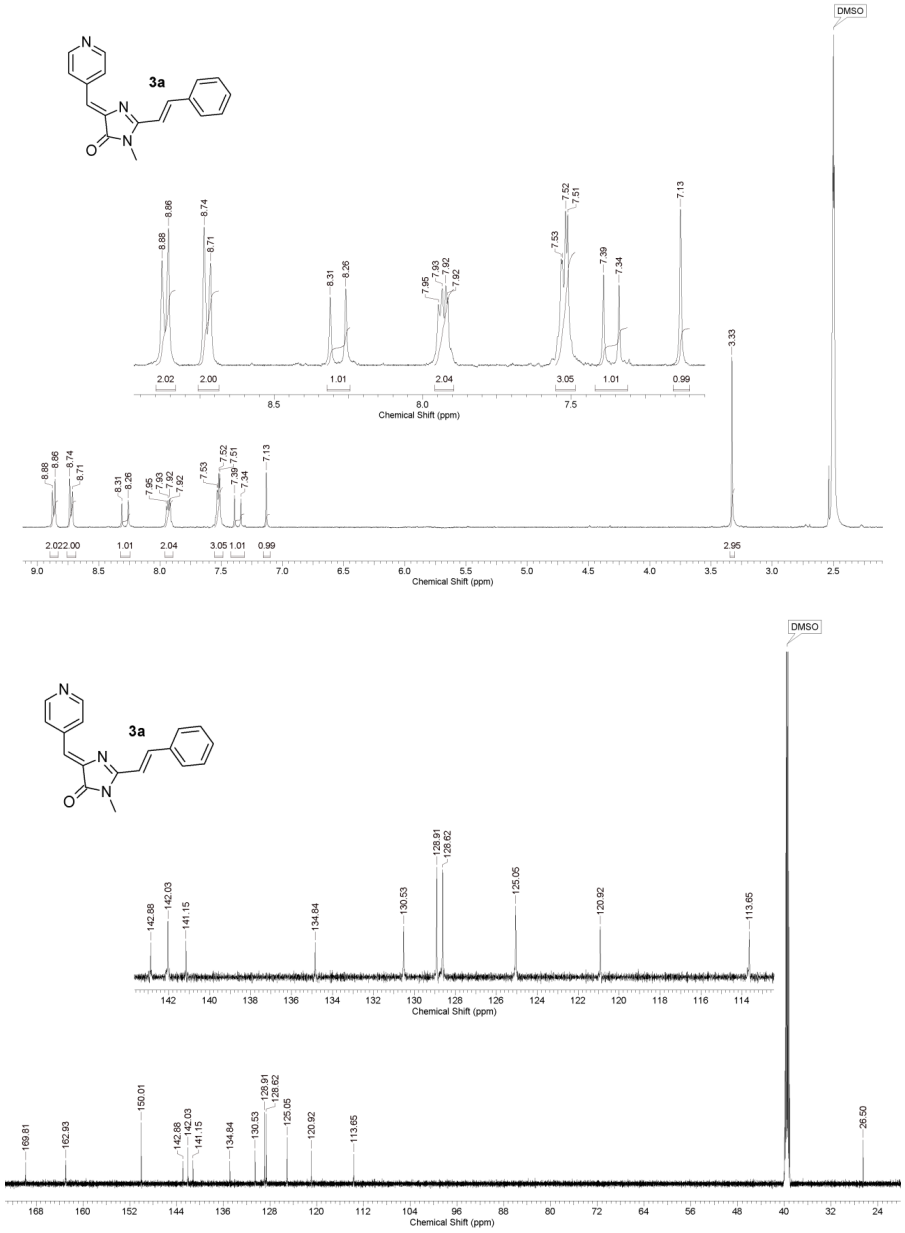


Figure 3.13: Compound 3a: ^{1}H NMR (300 MHz, DMSO-d_6) δ=8.87 (d, J=6.6 Hz, 2 H, Ar), 8.73 (d, J=6.5 Hz, 2 H, Ar), 8.29 (d, J=15.8 Hz, 1 H, CH=CH), 7.89 - 7.96 (m, 2 H, Ar), 7.50 - 7.55 (m, 3 H, Ar), 7.36 (d, J=15.7 Hz, 1 H, CH=CH), 7.13 (s, 1 H, Ar-CH), 3.33 (s, 3 H, CH_3); ^{13}C NMR (176 MHz, DMSO-d_6) δ = 26.5, 113.7, 120.9, 125.1, 128.6, 128.9, 130.5, 134.8, 141.2, 142.0, 142.9, 150.0, 162.9, 169.8; HRMS (m/z) calc-d. $C_{18}H_{16}N_3O$ for $[M+H]^+$ 290.1288, found 290.1292.

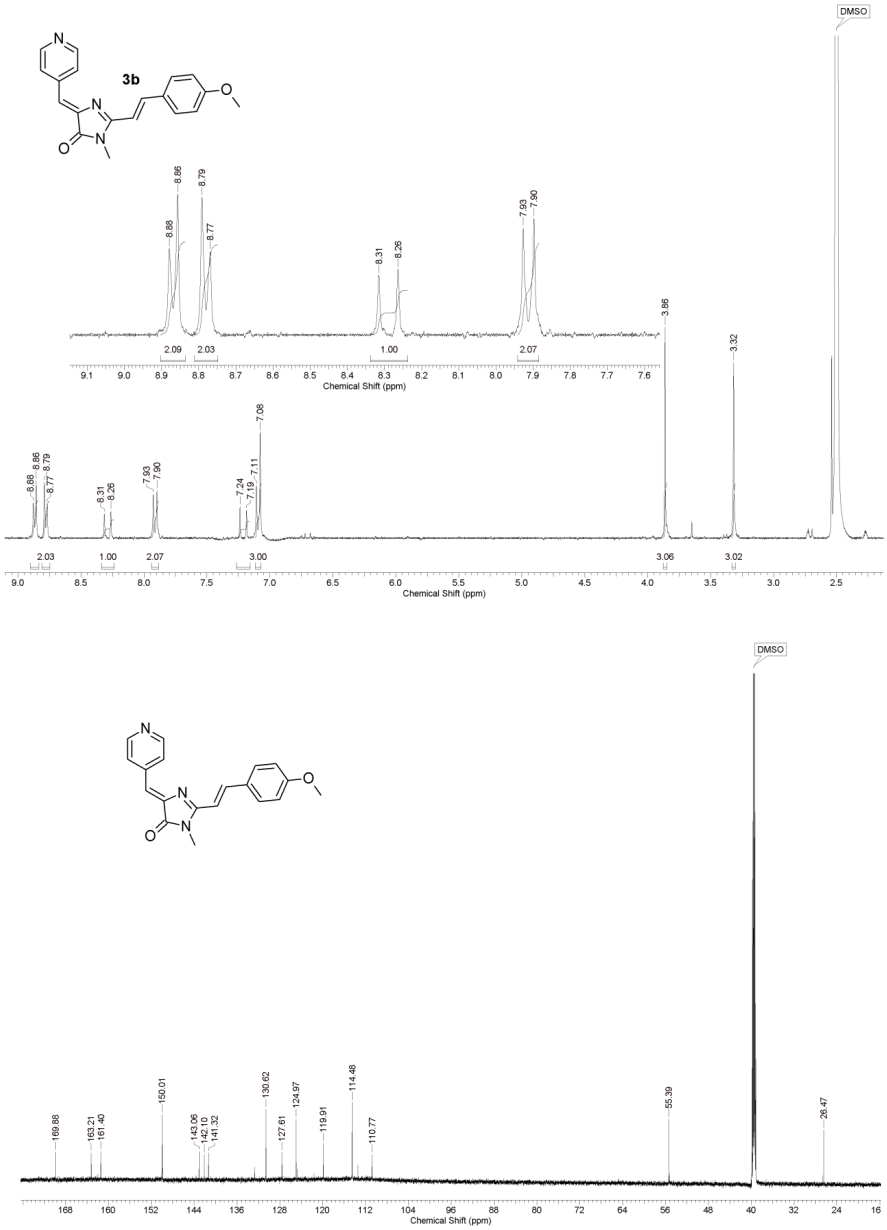


Figure 3.14: Compound 3b: ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.87 (d, J=6.8 Hz, 2 H, Ar), 8.78 (d, J=6.7 Hz, 2 H, Ar), 8.29 (d, J=15.5 Hz, 1 H, CH=CH), 7.91 (d, J=8.8 Hz, 2 H, Ar), 7.21 (d, J=15.6 Hz, 1 H, CH=CH), 7.07 - 7.11 (m, 3 H, Ar, Ar-CH), 3.86 (s, 3 H, CH_3), 3.32 (s, 3 H, CH_3); ^{13}C NMR (176 MHz, DMSO-d_6) δ=26.5, 55.4, 110.8, 114.5, 119.9, 125.0, 127.6, 130.6, 141.3, 142.1, 143.1, 150.0, 161.4, 163.2, 169.9; HRMS (m/z) calc-d. $C_{19}H_{18}N_3O_2$ for $[M+H]^+$ 320.1394, found 320.1397.

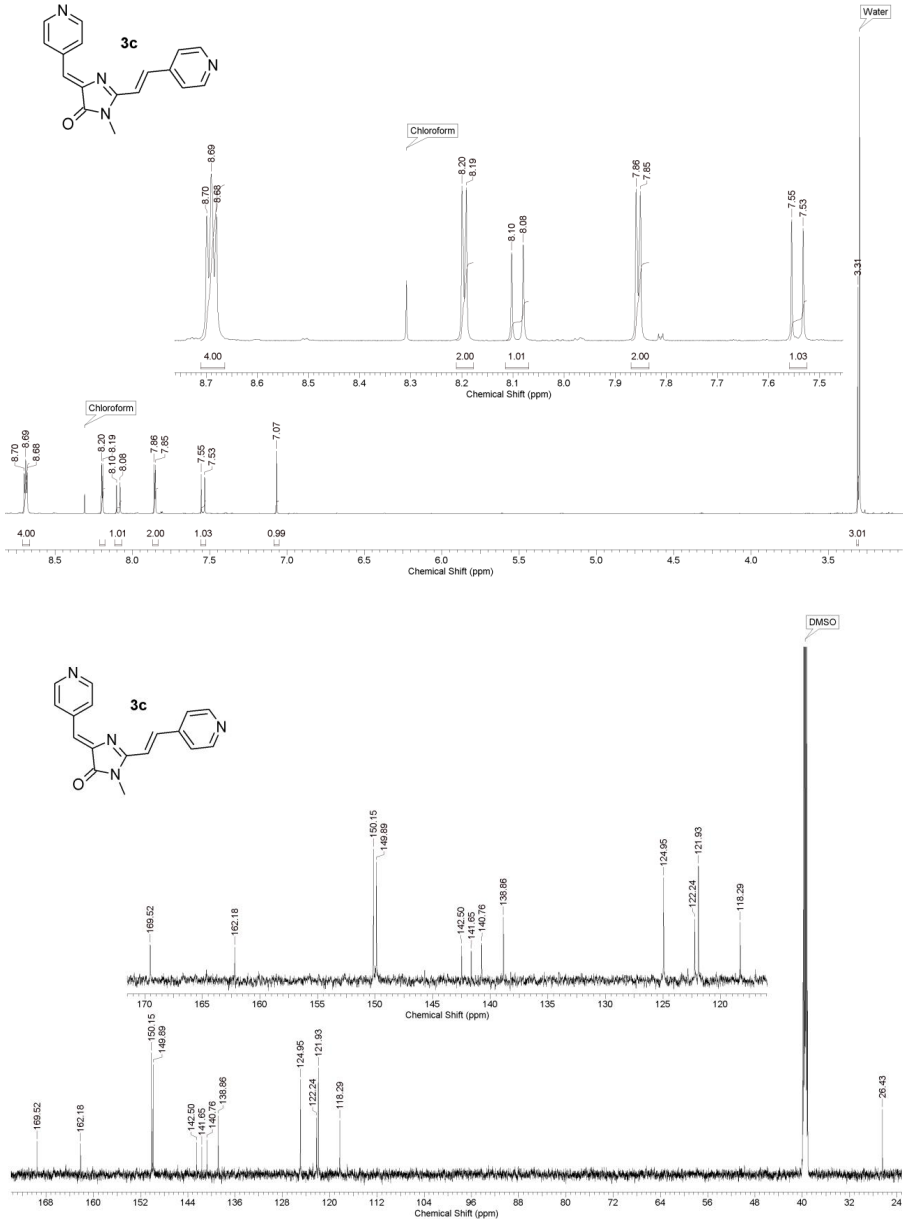


Figure 3.15: Compound 3c: ^{1}H NMR (700 MHz, DMSO-d_6) δ=8.66 - 8.71 (m, 4 H, Ar), 8.20 (d, J=5.9 Hz, 2 H, Ar), 8.09 (d, J=15.8 Hz, 1 H, CH=CH), 7.86 (d, J=5.9 Hz, 2 H, Ar), 7.54 (d, J=15.8 Hz, 1 H, CH=CH), 7.07 (s, 1 H, Ar-CH), 3.31 (s, 3 H, CH_3); ^{13}C NMR (176 MHz, DMSO-d_6) δ=26.4, 118.3, 121.9, 122.2, 124.9, 138.9, 140.8, 141.7, 142.5, 149.9, 150.2, 162.2, 169.5; HRMS (m/z) calc-d. $C_{17}H_{15}N_4O$ for $[M+H]^+$ 291.1240, found 291.1244.

3.6 Appendix C: Theoretical methods and computational details

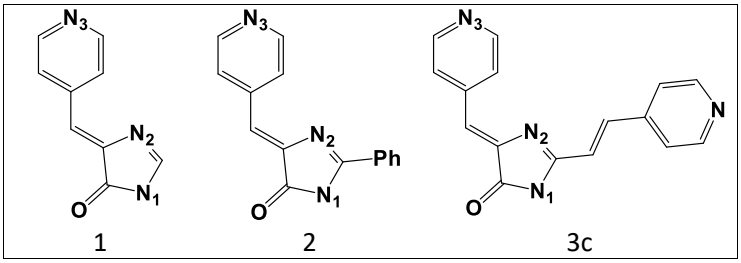


Figure 3.16: Model systems representing compounds 1, 2, and 3c.

We investigated electronic properties of the synthesized chromophores computationally. The calculations were carried out for model structures of **1**, **2**, and **3c** in which the terminal methyl groups were replaced with hydrogens. Fig. 3.16 shows the model systems. Although terminal methyl groups can affect excited-state lifetimes in conjugated GFP-like dyes, their effect on the energetics of excited states or photoacidic properties is known to be small[37]. The comparison (shown below) of gas-phase excitation energies of methylated and unmethylated model structures of **1**, **2**, and **3c** confirms that the computed excitation energies are not sensitive to substituting methyls by hydrogens.

The structures of the gas-phase chromophores were optimized by DFT and TDDFT using long-range corrected functional, ωB97X-D (Ref. 46) and the aug-cc-pVDZ basis set. The structures of the chromophores in solvents were optimized by DFT/TDDFT with ωPBEPBE and aug-cc-pVDZ. All relevant Cartesian coordinates are given below.

Absorption and emission energies were computed by TDDFT/ωPBEPBE (Ref. 47) with aug-cc-pVDZ as vertical energy differences at the S_0 and S_1 optimized geometries, respectively. To validate TDDFT protocol, we carried out additional calculations using another long-range

corrected functional, ωB97X-D as well as wave function methods, SOS-CIS(D)[48] and EOM-EE-CCSD[49]. To analyze excited-state wave functions, we computed natural transition orbitals (NTOs) using libwa module[50–52]. Cross sections for two-photon absorption (2PA) were computed at the EOM-CCSD level of theory[43].

To account for solvent effects, we used non-equilibrium polarizable continuum model (PCM)[53,54]. We employed both linear response (LR) and state-specific (SS) approaches. We also included perturbative corrections, ptLR and ptSS[53,54].

In excited-state calculations of solvated chromophores, we considered two different protocols: (i) single-point energy calculation using gas-phase optimized geometries of S_0 and S_1 and (ii) single-point energy calculation using the geometries of S_0 and S_1 optimized in a particular solvent. We found that the results obtained with protocol (ii) agree much better with the experimental trends; thus, the results reported below follow protocol (ii). Interestingly, the two protocols yield significantly different results only for emission energies because the S_0 structures are not sensitive to the solvent, in contrast to the S_1 structures.

All calculations were carried out using Q-Chem[55,56].

3.7 Appendix D: Computational results

3.7.1 Appendix D1: Excitation energies of 1, 2, and 3c in gas phase

Fig. 3.17 shows energy levels and NTOs corresponding to the two lowest excited states of **1** at the ground-state and relaxed excited-state geometries. Table 3.4 presents excitation energies at the S_0 and S_1 geometries and the corresponding Stokes shifts for **1**, **2**, and **3c** in gas phase computed at different levels of theory; Table 3.5 shows excitation energies for methylated model structures. In agreement with the experimental findings[37], methyl groups have negligible effect on the excitation energies. The results for the key electronic properties are consistent across

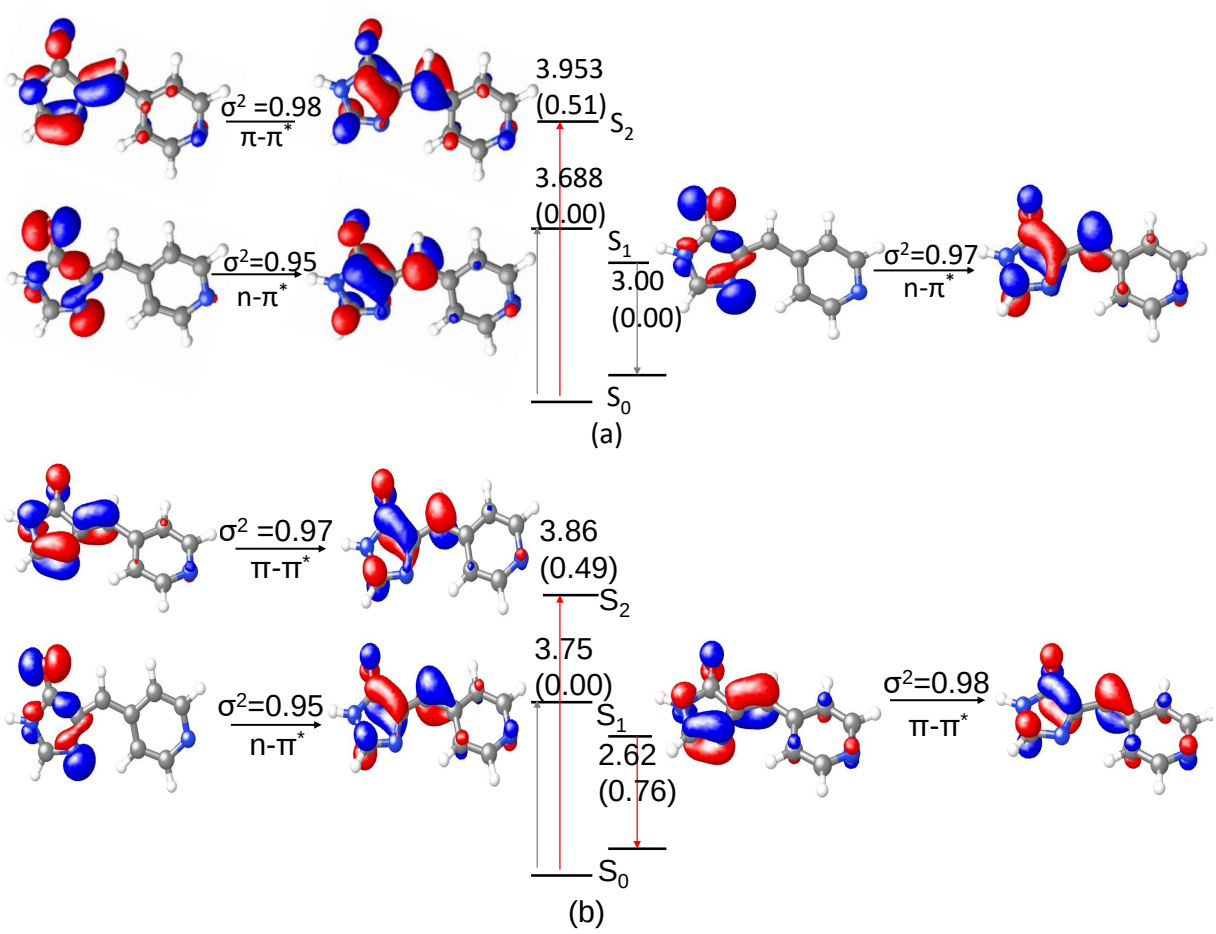


Figure 3.17: Excited states and NTOs for 1 in (a) gas phase and (b) water. Left and right panels show the states at the S_0 and S_1 optimized geometries, respectively.

all methods. The results in Table 3.4 reveal important differences between **1**, **2**, and **3c**. The lowest bright transition at the ground-state geometry in **1** corresponds to the $S_0 \rightarrow S_2$ transition, whereas in **2** and **3c** it corresponds to the S_0-S_1 transition. In **1**, the S_1 state is dark at the ground-state geometry. The oscillator strengths of S_1 and S_2 can be explained by NTOs, which correspond to $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ type transitions, respectively. Because the emission usually happens from S_1, by virtue of Kasha's rule, **1** is non-fluorescent. In **2** and **3c**, the NTOs for the $S_0 \rightarrow S_1$ transition correspond to $\pi \rightarrow \pi^*$ excitation. Participation ratios for all transitions are close to one.

Table 3.4: Excitation energies of 1, 2, and 3c in gas phase. All energies are in eV; oscillator strength is given in parenthesis. aug-cc-pVDZ basis set.

Method	**1**			**2**			**3c**		
	E_{ex} (S_0-S_2)	E_{ex} (S_1-S_0)	Stokes shift	E_{ex} (S_0-S_1)	E_{ex} (S_1-S_0)	Stokes shift	E_{ex} (S_0-S_1)	E_{ex} (S_1-S_0)	Stokes shift
ωPBEPBE	3.953 (0.51)	3.003 (0.00)	0.95	3.660 (0.69)	2.935 (0.60)	0.725	3.512 (1.24)	2.881 (1.25)	0.631
ωB97X-D	3.923 (0.53)	3.128 (0.00)	0.805	3.595 (0.69)	2.914 (0.65)	0.681	3.428 (1.24)	2.829 (1.27)	0.599
SOS-CIS(D)	4.238 (0.77)	3.152 (0.00)	1.086	3.170 (0.71)	2.331 (0.69)	0.839	3.382 (1.04)	2.639 (1.08)	0.743
EOM-CCSD	4.256 (0.55)	3.374 (0.00)	0.882	3.281 (0.80)	2.399 (0.88)	0.882	3.311 (1.09)	2.604 (1.08)	0.707

Using ωB97X-D optimized geometries.

Table 3.5: Excitation energies of methylated analogues of 1, 2, and 3c in gas phase. All energies are in eV; oscillator strength is given in parenthesis. ωB97X-D/aug-cc-pVDZ.

Molecule	E_{ex}^{abs} (f_l)	E_{ex}^{em} (f_l)	Stokes shift
1	3.910 (0.51)	2.976 (0.12)	0.934
2	3.566 (0.57)	2.787 (0.48)	0.779
3c	3.367 (1.00)	2.737 (1.01)	0.630

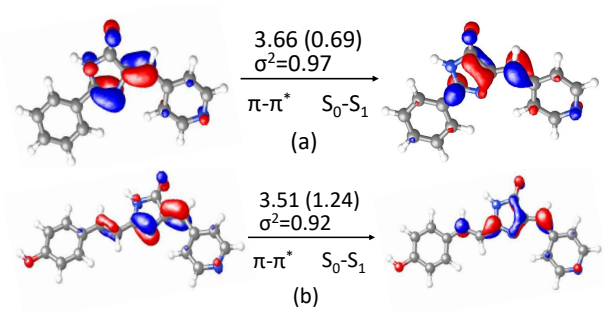


Figure 3.18: NTOs for the $S_0 \rightarrow S_1$ transition in 2 (left) and 3c (right) in the gas phase.

3.7.2 Appendix D2: Solvatochromic properties of molecule 1

Table 3.6 shows electronic properties of **1**. Mulliken's charges on nitrogens in the ground and excited states are shown in Table 3.7. Ground- and excited-state structures are optimized in each solvent, as described in Section 3.6. Fig. 3.19 shows the key bondlengths in S_0- and S_1-optimized structures of **1** in gas phase and in water (only bonds that show significant change

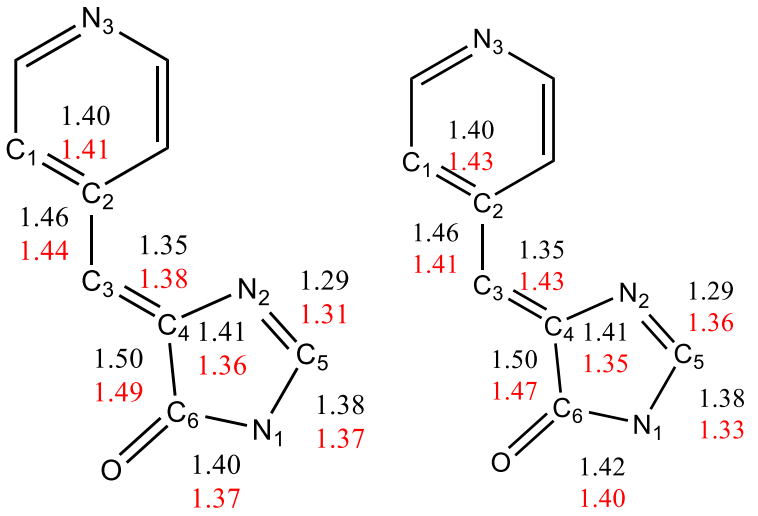


Figure 3.19: Ground- and excited-state structures of 1 in the gas phase (left) and in water (right). Black and red numbers denote selected bondlengths in S_0 and S_1, respectively.

Table 3.6: Electronic properties of 1 in various solvents. Energies are in eV; dipole moments in debye.

Solvent	E_{ex} (S_0-S_2)	μ_{tr} (S_0-S_2)	μ_{gs} (S_0)	μ_{ex} (S_0)	E_{ex} (S_1-S_0)	μ_{tr} (S_1-S_0)	μ_{ex} (S_1)	μ_{gs} (S_1)	ΔE_{ss}	$\Delta\mu_{tr}$	$\Delta\mu_{ge}$
gas phase	3.95	2.30	4.03	5.88	3.00	0.05	6.66	4.08	0.95	2.25	2.63
water	3.86	2.26	5.78	8.04	2.62	3.39	8.69	6.22	1.24	1.13	2.91
acetonitrile	3.86	2.28	5.73	7.98	2.63	3.24	8.44	6.06	1.23	0.96	2.71
methanol	3.86	2.25	5.72	7.96	2.63	3.23	8.43	6.05	1.22	0.98	2.70
acetate	3.86	2.27	5.34	7.42	2.71	3.02	7.82	5.59	1.15	0.75	2.48
dioxane	3.87	2.31	4.75	6.57	2.86	2.37	7.12	4.92	1.02	0.06	2.37
hexane	3.88	2.32	4.62	6.38	2.98	0.07	7.16	4.85	0.90	2.31	3.54

upon excitation are shown). The theoretical values of excitation energies are systematically blue-shifted relative to the experiment. Tables 3.6 and 3.8 show that absorption energy is not affected by solvent polarity (see also Fig. 3.20). Because the variations in absorption energy are small (0.01-0.02 eV), the correlation between calculated and experimental values appears to be poor. Yet, theory and experiment are in qualitative agreement that absorption maximum is not very sensitive to the solvent polarity.

The two protocols described in Section 3.6 yield very different results. When emission energies are computed without re-optimization of excited-state structures in each solvent, there

is no solvatochromic shift, which contradicts the experimental observation. However, upon re-optimization of excited-state structures in each solvent, we obtain solvent-dependent shifts which correlate well with the experimental values. This result illustrates the sensitivity of excited-state geometries to solvent polarity (Fig. 3.20). The analysis of the optimized structures (Fig. 3.19) reveals that changes in bondlengths upon photoexcitation increase in polar solvent. Tables 3.6 and 3.8 show that emission energies decrease in polar solvents, giving rise to the increased Stokes shifts. Thus, in agreement with the experiment, the calculations confirm that the variations in Stokes shifts in **1** and **2** are driven by the variations in emission energy. Although the computed excitation energies are blue-shifted relative to the experimental peak maxima, the differences appear to be systematic and the computed and theoretical Stokes shifts are in good agreement. Moreover, the computed and experimental solvatochromic trends correlate well, confirming that the blue shift of theoretical values relative to the experiment is systematic.

Table 3.7: Mulliken charges on nitrogen atoms in 1 (see Fig. 3.16 for atom numbering).

solvent	$N_1(S_0)$ (S_0)	$N_2(S_0)$ (S_0)	$N_3(S_0)$ (S_0)	$N_1(S_2)$ (S_0)	$N_2(S_2)$ (S_0)	$N_3(S_2)$ (S_0)	$N_1(S_1)$ (S_1)	$N_2(S_1)$ (S_1)	$N_3(S_1)$ (S_1)
gas phase	-0.09	-0.10	-0.44	-0.09	0.04	-0.44	-0.16	0.08	-0.45
water	-0.24	-0.13	-0.47	-0.20	-0.08	-0.48	-0.21	-0.17	-0.52
acetonitrile	-0.24	-0.13	-0.47	-0.20	-0.08	-0.48	-0.22	-0.36	-0.49
methanol	-0.24	-0.13	-0.47	-0.20	-0.08	-0.48	-0.20	-0.31	-0.51
acetate	-0.24	-0.12	-0.46	-0.20	-0.07	-0.47	-0.20	-0.31	-0.49
dioxane	-0.23	-0.11	-0.45	-0.19	-0.07	-0.46	-0.22	-0.28	-0.46
hexane	-0.23	-0.11	-0.45	-0.19	-0.07	-0.46	-0.29	0.00	-0.48

Solvent-dependent variations in excitation energy and Stokes shifts can be explained by comparing the trends in energies with transition and permanent dipole moments (Table 3.6 and Figs. 3.20 and 3.21). Electronic excitation leads to an increase of the dipole moment in the bright state, which is consistent with its $\pi\pi^*$ character. Similarly to the excitation energies, ground-state dipole moment and transition dipole moment at the S_0 geometry are not strongly affected by variations in solvent polarity. In contrast, at the S_1 geometry both permanent and

transition dipole moments show large variations; both values increase in polar solvents. We observe good correlation in the trends in Stokes shifts with the transition dipole moment (Fig. 3.21) and the change in permanent dipole moment ($\Delta\mu_{ge}$, Fig. 4 of main text). This correlation confirms that the solvatochromism of the Stokes shifts in these chromophores originates in increased charge separation in the $\pi\pi^*$ state, which leads to significant and solvent-dependent structural relaxation of the excited state. Fig. 3.21 (bottom) compares the trend in FQY versus the transition dipole moment (in this figure, hexane is excluded because of the change in electronic state character). Interestingly, there is a reasonable correlation between the two quantities — the FQY decreases with the increase of μ_{tr}.

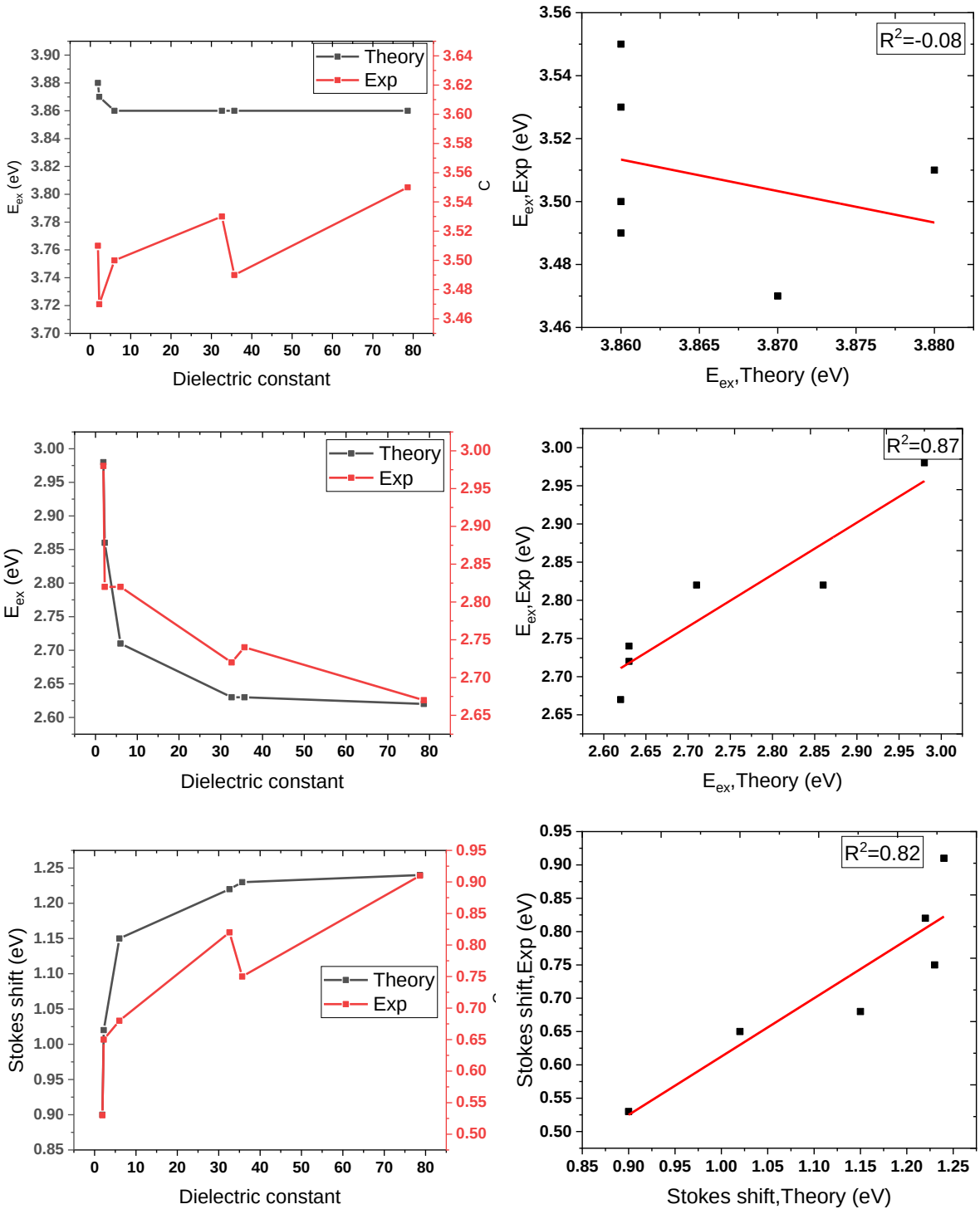


Figure 3.20: Variation in absorption (top) and emission (middle) energies and Stokes shifts (bottom) of 1 in different solvents (left) and correlation between theory and experiment (right).

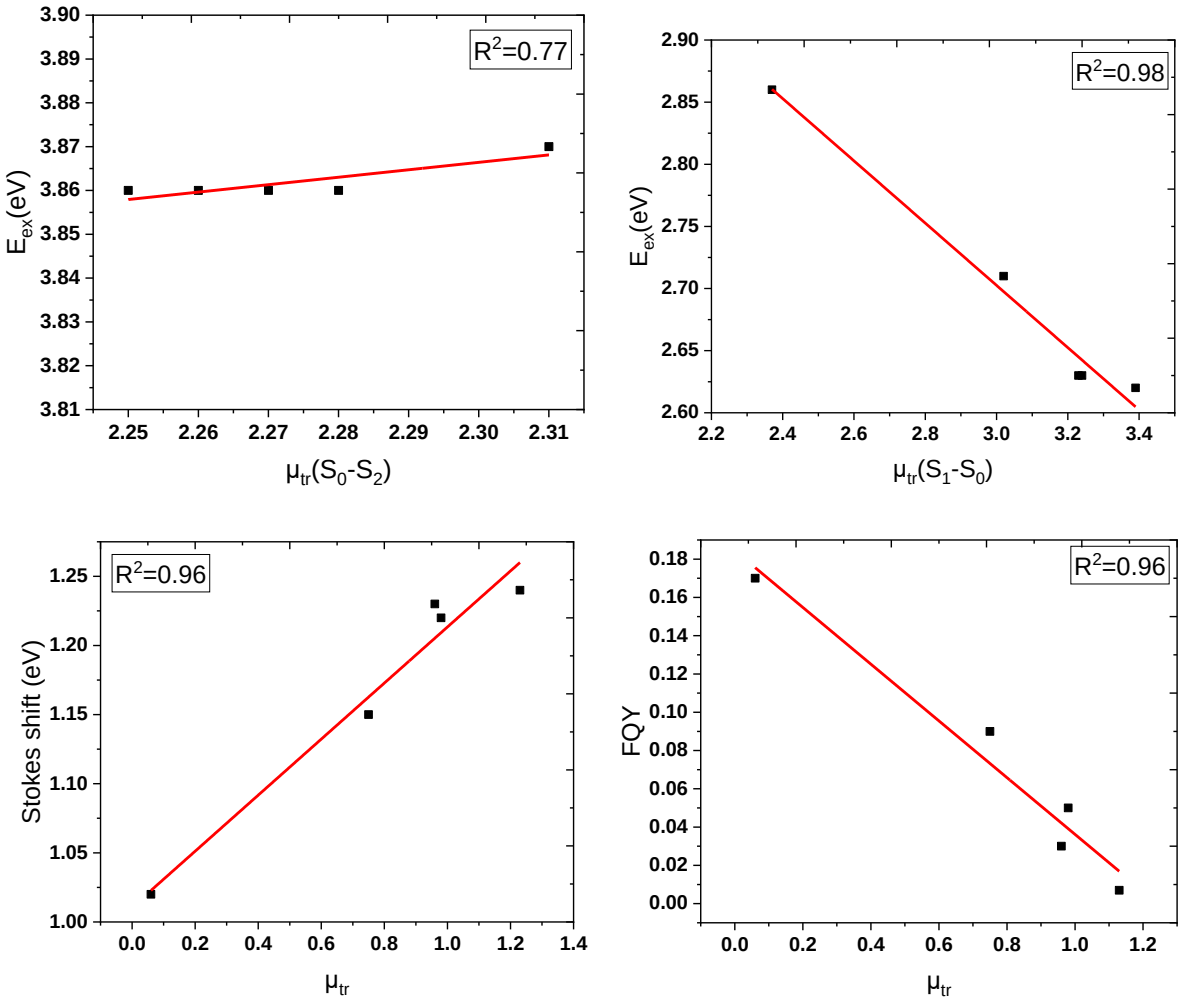


Figure 3.21: Top: Absorption (left) and emission energies (right) in different solvents versus transition dipole moment for 1. Bottom: Stokes shift (left) and FQY (right) in different solvents versus transition dipole moment.

3.7.3 Appendix D3: Solvatochromic properties of molecule 2

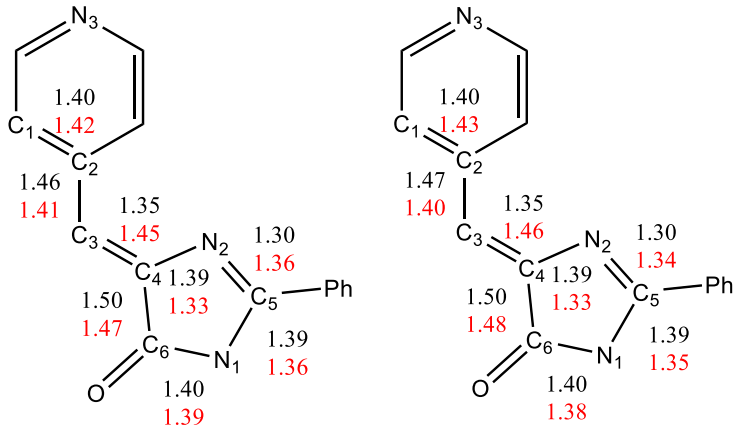


Figure 3.22: Ground- and excited-state structures of 2 in the gas phase (left) and in water (right). Black and red numbers denote selected bondlengths in S_0 and S_1, respectively.

Table 3.8: Electronic properties of 2 in various solvents. Energies are in in eV, dipole moments in debye.

Solvent	E_{ex} (S_0-S_1)	μ_{tr} (S_0-S_1)	μ_{gs} (S_0)	μ_{ex} (S_0)	E_{ex} (S_1-S_0)	μ_{tr} (S_1-S_0)	μ_{ex} (S_1)	μ_{gs} (S_1)	ΔE_{ss}	μ_{tr}	$\Delta\mu_{ge}$
gas phase	3.66	2.77	5.31	5.91	2.93	2.88	6.09	5.84	0.73	0.11	0.78
water	3.57	2.76	7.33	9.44	2.47	3.94	10.79	8.71	1.11	1.19	3.46
acetonitrile	3.58	2.76	7.28	9.36	2.48	3.77	10.68	8.64	1.10	1.02	3.40
methanol	3.57	2.76	7.27	9.35	2.48	3.92	10.67	8.63	1.09	1.16	3.39
acetate	3.58	2.77	6.84	8.73	2.57	3.78	9.80	8.04	1.00	1.01	2.96
dioxane	3.59	2.77	6.17	7.69	2.75	3.51	8.39	7.09	0.84	0.78	2.22
hexane	3.59	2.77	6.02	7.44	2.79	3.44	8.06	6.87	0.81	0.67	2.05

Table 3.9: Mulliken charges on nitrogen atoms in 2 (see Fig. 3.16 for atom numbering).

solvent	$N_1(S_0)$ (S_0)	$N_2(S_0)$ (S_0)	$N_3(S_0)$ (S_0)	$N_1(S_1)$ (S_0)	$N_2(S_1)$ (S_0)	$N_3(S_1)$ (S_0)	$N_1(S_1)$ (S_1)	$N_2(S_1)$ (S_1)	$N_3(S_1)$ (S_1)
gas phase	-0.29	-0.30	-0.47	-0.21	-0.13	-0.48	-0.23	-0.17	-0.49
water	-0.34	-0.37	-0.52	-0.31	-0.34	-0.53	-0.33	-0.40	-0.55
acetonitrile	-0.34	-0.37	-0.52	-0.31	-0.34	-0.53	-0.33	-0.40	-0.55
methanol	-0.34	-0.37	-0.52	-0.31	-0.34	-0.53	-0.33	-0.40	-0.55
acetate	-0.33	-0.35	-0.51	-0.30	-0.33	-0.52	-0.32	-0.40	-0.52
dioxane	-0.31	-0.34	-0.50	-0.29	-0.32	-0.50	-0.31	-0.38	-0.52
hexane	-0.31	-0.33	-0.49	-0.28	-0.31	-0.50	-0.30	-0.38	-0.52

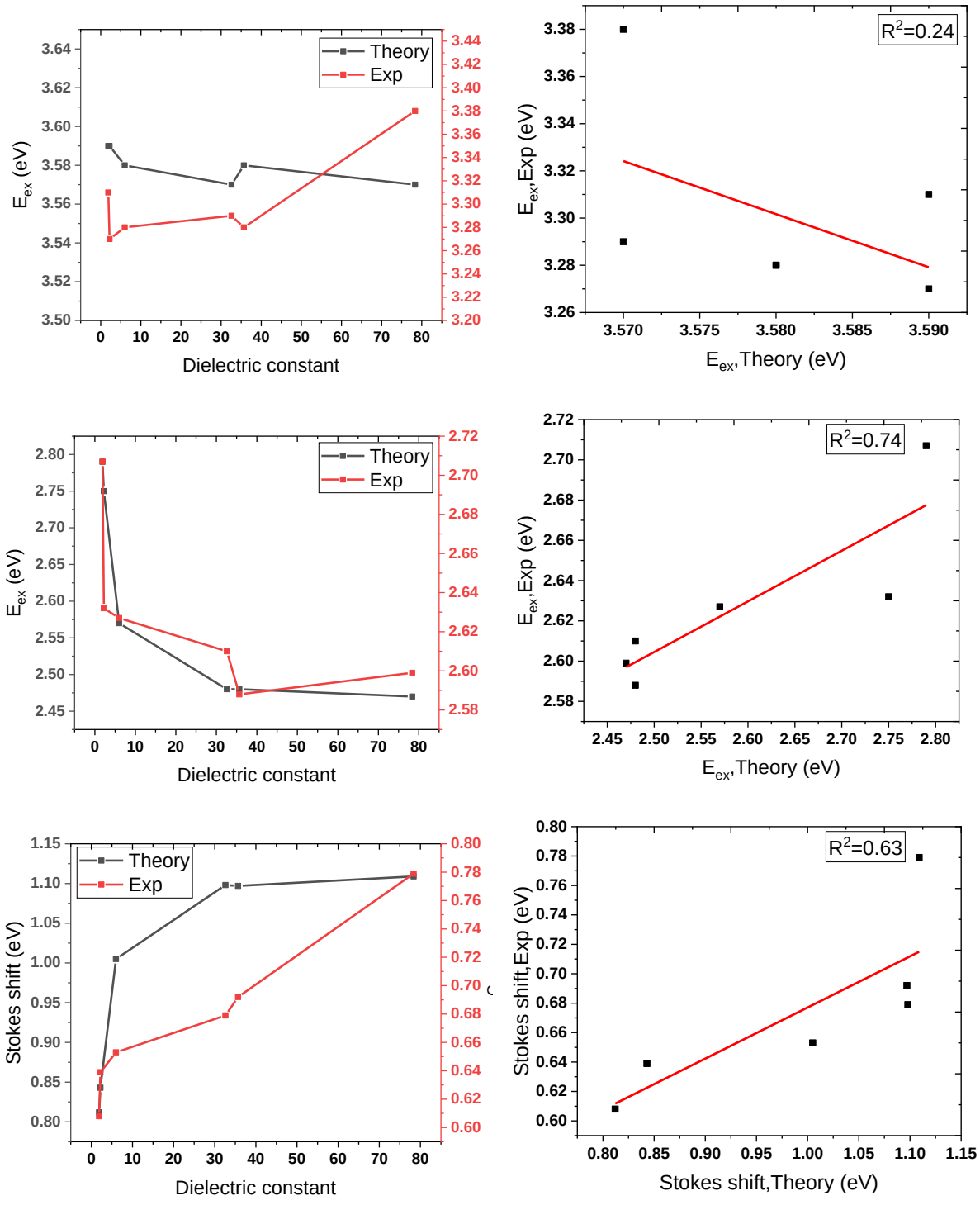


Figure 3.23: Variations in absorption (top) and emission (middle) energies and Stokes shifts (bottom) of 2 in different solvents (left) and correlation between theory and experiment (right).

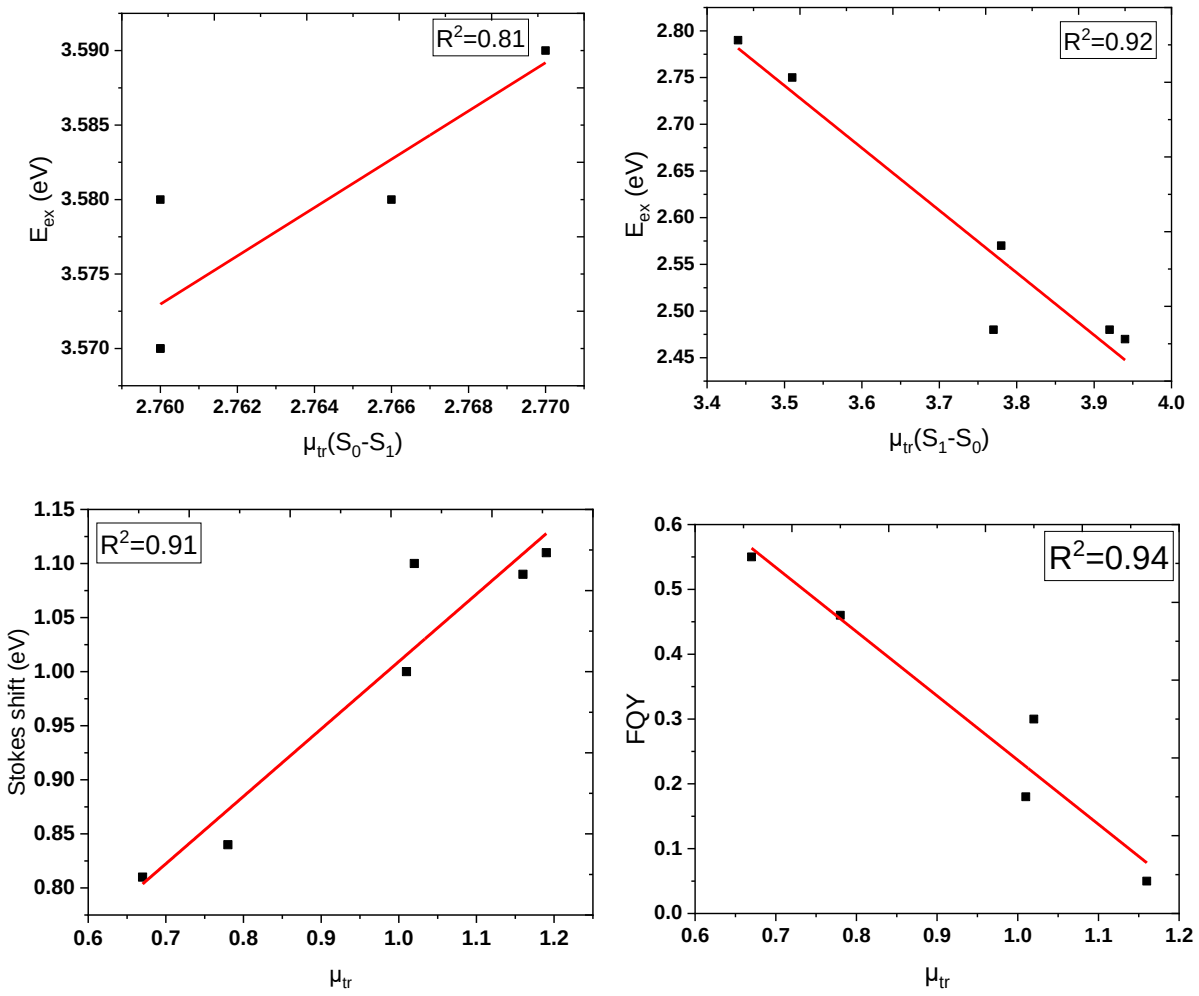


Figure 3.24: Top: Absorption (left) and emission energies (right) in different solvents versus transition dipole moment for 2. Bottom: Stokes shift (left) and FQY (right) in different solvents versus transition dipole moment.

3.7.4 Appendix D4: Solvatochromic properties of molecule 3

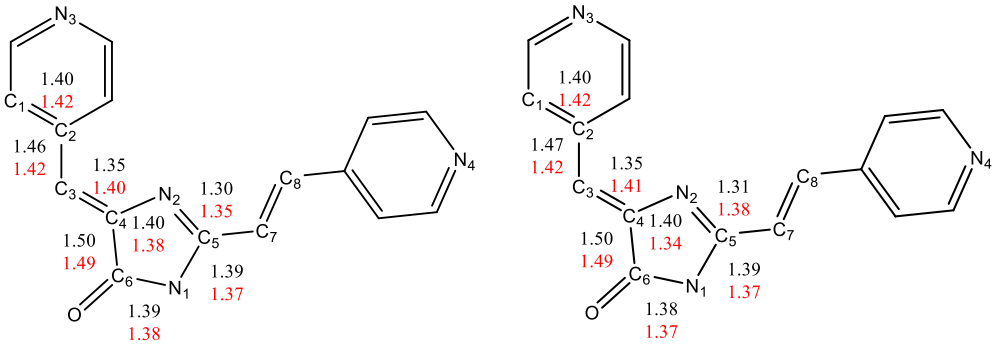


Figure 3.25: Ground- and excited-state structures of 3 in the gas phase (left) and in water (right). Black and red numbers denote selected bondlengths in S_0 and S_1, respectively.

Table 3.10: Electronic properties of 3c in various solvents. Energies are in in eV, dipole moments in debye.

Solvent	E_{ex} (S_0-S_1)	μ_{tr} (S_0-S_1)	μ_{gs} (S_0)	μ_{ex} (S_0)	E_{ex} (S_1-S_0)	μ_{tr} (S_1-S_0)	μ_{ex} (S_1)	μ_{gs} (S_1)	ΔE_{ss}	μ_{tr}	$\Delta\mu_{ge}$
gas phase	3.51	3.79	7.63	9.61	2.88	4.21	9.39	8.81	0.63	0.42	1.76
water	3.37	3.86	10.50	14.99	2.26	5.39	19.25	14.28	1.11	1.53	8.75
acetonitrile	3.37	3.86	10.42	14.87	2.27	5.35	18.89	14.09	1.10	1.49	8.47
methanol	3.37	3.87	10.40	14.85	2.28	5.35	18.84	14.05	1.09	1.48	8.44
acetate	3.39	3.85	9.75	13.82	2.45	4.61	14.63	11.31	0.94	0.76	4.88
dioxane	3.42	3.82	8.76	12.02	2.44	4.67	12.60	11.08	0.98	0.85	3.84
hexane	3.43	3.82	8.54	11.60	2.56	4.31	10.97	10.01	0.87	0.49	2.43

Table 3.11: Mulliken charges on nitrogen atoms in 3c (see Fig. 3.16 for atom numbering).

solvent	N_1(S_0) (S_0)	N_2(S_0) (S_0)	N_3(S_0) (S_0)	N_1(S_1) (S_0)	N_2(S_1) (S_0)	N_3(S_1) (S_0)	N_1(S_1) (S_1)	N_2(S_1) (S_1)	N_3(S_1) (S_1)
gas phase	-0.31	-0.19	-0.49	-0.27	-0.18	-0.50	-0.29	-0.33	-0.51
water	-0.36	-0.28	-0.53	-0.34	-0.26	-0.54	-0.38	-0.42	-0.57
acetonitrile	-0.36	-0.28	-0.53	-0.34	-0.26	-0.54	-0.37	-0.42	-0.57
methanol	-0.36	-0.28	-0.53	-0.34	-0.26	-0.54	-0.37	-0.42	-0.57
acetate	-0.35	-0.26	-0.52	-0.33	-0.24	-0.53	-0.37	-0.41	-0.56
dioxane	-0.34	-0.24	-0.51	-0.32	-0.22	-0.51	-0.34	-0.38	-0.54
hexane	-0.33	-0.24	-0.50	-0.31	-0.21	-0.51	-0.34	-0.37	-0.53

3.7.5 Appendix D5: Analysis of ground- and excited-state structures of 1, 2, and 3c

Table 3.12: Key structural parameters of 1, 2, and 3c in S_0 and S_1 and changes in bondlengths (ΔBL). All bondlength are in Å.

System	Solvent	State	C_1C_2	C_2C_3	C_3C_4	C_4C_6	C_6N_1	N_1C_5	C_5N_2	C_4N_2
1	gas phase	S_0	1.40	1.46	1.35	1.50	1.40	1.38	1.29	1.41
		S_1	1.41	1.44	1.38	1.49	1.37	1.37	1.31	1.36
		ΔBL	0.01	-0.02	0.03	-0.01	-0.03	-0.01	0.02	-0.05
	water	S_0	1.40	1.46	1.35	1.50	1.42	1.38	1.29	1.41
		S_1	1.43	1.41	1.43	1.47	1.40	1.33	1.36	1.35
		ΔBL	0.03	-0.05	0.08	-0.03	-0.02	-0.05	0.07	-0.06
2	gas phase	S_0	1.40	1.46	1.35	1.50	1.40	1.39	1.30	1.39
		S_1	1.42	1.41	1.45	1.47	1.39	1.36	1.36	1.33
		ΔBL	0.02	-0.05	0.10	-0.03	-0.01	-0.03	0.06	-0.06
	water	S_0	1.40	1.47	1.35	1.50	1.40	1.39	1.30	1.39
		S_1	1.43	1.40	1.46	1.48	1.38	1.35	1.34	1.33
		ΔBL	0.03	-0.07	0.11	-0.02	-0.02	-0.04	0.04	-0.06
3	gas phase	S_0	1.40	1.46	1.35	1.50	1.39	1.39	1.30	1.40
		S_1	1.42	1.42	1.40	1.49	1.38	1.37	1.35	1.38
		ΔBL	0.02	-0.04	0.05	-0.01	-0.01	-0.02	0.05	-0.02
	water	S_0	1.40	1.47	1.35	1.50	1.38	1.39	1.31	1.40
		S_1	1.42	1.42	1.41	1.49	1.37	1.37	1.38	1.34
		ΔBL	0.02	-0.05	0.06	-0.01	-0.01	-0.02	0.07	-0.06

Optimized ground- and excited-state structures of **1**, **2**, and **3c** are shown in Figs. 3.19, 3.22, and 3.25 and summarized in Table 3.12. In agreement with previous studies of GFP-like chromophores[39–41], photoexcitation results in significant changes in the bondlength pattern, which can be explained by the simple Hückel model[?,?,36]. The most affected bonds are those of the methyne bridge (C_2C_3 and C_3C_4) and around N_2 (C_5N_2 and C_4N_2). Formally double bonds (C_3C_4 and C_5N_2) elongate and formally single bonds (C_2C_3 and C_4N_2) contract. The changes are consistently larger in polar solvents. The changes in bondlengths alternation reflect changes in relative weights of leading resonance structures and are, therefore, related to changes in charge redistribution (as discussed, for example, in Refs.[36,41]). Thus, they can be related to trends in dipole moments (i.e., larger change in bondlengths in S_1 corresponds to larger dipole

moment). The changes in bondlengths can also be related to the shape of the S_1 PES. As was shown in Ref.[40], partial flipping of bond orders in excited state leads to a flatter (along torsional coordinate) PES, which increases the efficiency of internal conversion. Thus, the trend in equilibrium structures suggest a possible explanation of the observed anti-correlation between FQY and solvent polarity.

3.7.6 Appendix D6: Photoacidity/photobasicity of 1, 2, and 3c

The NTO analysis reveals the origin of photobasicity — in all three compounds, electronic excitation results in the redistribution of electronic density on the imidazolone cycle. Tables 3.7, 3.9, and 3.11 show partial charges of the nitrogen atoms in the ground and excited states for **1**, **2**, and **3c**. As one can see, the most significant charge redistribution occurs on N_2, which becomes more negative (thus, more basic) in S_1. The change in charge is large in **1** and **3c** (about 0.2e) and is relatively small in **2**. In all three molecules, the charge on N_2 shows the largest solvent-dependent variations.

3.7.7 Appendix D7: 2PA cross sections of 1 and 2

Table 3.13 shows excitation energies, oscillator strengths, and 2PA cross-sections (for parallel polarization) for HBDI and chromophores **1** and **2**. 2PA excitation wavelengths λ correspond to $E_{ex}/2$. We note that the computed excitation energies are systematically overestimated, meaning that the actual wavelength can be longer. As one can see, the 2PA cross-section in **2** is about 8-10 times smaller than in HBDI, which suggest that it still might be used in a two-photon excitation regime. The lowest transitions in **1** are also dim, but it also has a very bright 2PA transition at 458 nm. The intensity of this band can spill over to longer wavelengths[44], so **1** might actually be brighter than HBDI.

Table 3.13: Excitation energies, oscillator strength (f_f), and 2PA cross-section for degenerate resonant photons ($\lambda_1 = \lambda_2 = 2 \times \hbar/\omega_{ex}$), aug-cc-pVDZ.

Compound	State	E_{ex}, eV	f_l	λ, nm	δ^{2PA} (a.u.)	σ (GM)
HBDI	1A′	3.97	0.77	625	977.762	5.64
1	2A′	4.26	0.55	582	64.997	0.43
	5A′	5.42	0.20	458	4321.8	46.39
2	1A′	3.28	0.80	756	137.18	0.54
	2A′	3.91	0.08	634	131.93	0.74
	3A′	4.09	0.02	606	98.7	0.60

3.8 Chapter 3 references

[1] Klymchenko, A. S. Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. *Acc. Chem. Res.* **2017**, *50*, 366–375.

[2] Hori, Y.; Norinobu, T.; Sato, M.; Arita, K.; Shirakawa, M.; Kikuchi, K. Development of Fluorogenic Probes for Quick No-Wash Live-Cell Imaging of Intracellular Proteins. *J. Am. Chem. Soc.* **2013**, *135*, 12360–12365.

[3] Schoen, I.; Ries, J.; Klotzsch, E.; Ewers, H.; Vogel, V. Binding-Activated Localization Microscopy of DNA Structures. *Nano Lett.* **2011**, *11*, 4008–4011.

[4] Collot, M.; Kreder, R.; Tatarets, A. L.; Patsenker, L.; Mely, Y.; Klymchenko, A. S. Bright Fluorogenic Squaraines with Tuned Cell Entry for Selective Imaging of Plasma Membrane vs. Endoplasmic Reticulum. *Chem. Comm.* **2015**, *51*, 17136–17139.

[5] Lukinavicius, G. et al. Fluorogenic Probes for Live-Cell Imaging of the Cytoskeleton. *Nat. Methods* **2014**, *11*, 731–733.

[6] Hu, F.; Liu, B. Organelle-Specific Bioprobes Based on Fluorogens with Aggregation-Induced Emission (AIE) Characteristics. *Organic & biomolecular chemistry* **2016**, *14*, 9931–9944.

[7] Tsien, R. The Green Fluorescent Protein. *Annu. Rev. Biochem.* **1998**, *67*, 509–544.

[8] Meech, S. Excited State Reactions in Fluorescent Proteins. *Chem. Soc. Rev.* **2009**, *38*, 2922–2934.

[9] Acharya, A.; Bogdanov, A. M.; Bravaya, K. B.; Grigorenko, B. L.; Nemukhin, A. V.; Lukyanov, K. A.; Krylov, A. I. Photoinduced Chemistry in Fluorescent Proteins: Curse or Blessing? *Chem. Rev.* **2017**, *117*, 758–795.

[10] Baleeva, N. S.; Baranov, M. S. Synthesis and Properties of 5-Methylidene-3, 5-dihydro-4H-imidazol-4-ones (Microreview). *Chem. Heterocycl. Comp.* **2016**, *52*, 444–446.

[11] Deng, H.; Su, Y.; Hu, M.; Jin, X.; He, L.; Pang, Y.; Dong, R.; Zhu, X. Multicolor Fluorescent Polymers Inspired from Green Fluorescent Protein. *Macromol.* **2015**, *48*, 59695979.

[12] Deng, H.; Zhu, X. Emission Enhancement and Application of Synthetic Green Fluorescent Protein Chromophore Analogs. *Mater. Chem. Front.* **2017**, *1*, 619–629.

[13] Walker, C. L.; Lukyanov, K. A.; Yampolsky, I. V.; Mishin, A. S.; Bommarius, A. S.; Duraj-Thatte, A. M.; Azizi, B.; Tolbert, L. M.; Solntsev, K. M. Fluorescence Imaging Using Synthetic GFP Chromophores. *Curr. Opinion in Chem. Biol* **2015**, *27*, 64–74.

[14] Baranov, M. S.; Lukyanov, K. A.; Yampolsky, I. V. Synthesis of the Chromophores of Fluorescent Proteins and their Analogs. *Russ. J. Bioorg. Chem.* **2013**, *39*, 223–244.

[15] Baldridge, A.; Samanta, S.; Jayaraj, N.; Ramamurthy, V.; Tolbert, L. Steric and Electronic Effects in Capsule-Confined Green Fluorescent Protein Chromophores. *J. Am. Chem. Soc.* **2011**, *133*, 712–715.

[16] Baldridge, A.; Feng, S.; Chang, Y.-T.; Tolbert, L. Recapture of GFP Chromophore Fluorescence in a Protein Host. *ACS Comb. Sci.* **2011**, *13*, 214–217.

[17] Williams, D. E.; Dolgopolova, E. A.; Pellechia, P. J.; Palukoshka, A.; Wilson, T. J.; Tan, R.; Maier, J. M.; Greytak, A. B.; Smith, M. D.; Krause, J. A.; Shustova, N. B. Mimic of the Green Fluorescent Protein β-Barrel: Photophysics and Dynamics of Confined Chromophores Defined by a Rigid Porous Scaffold. *J. Am. Chem. Soc.* **2014**, *137*, 2223–2226.

[18] Dolgopolova, E. A.; Rice, A. M.; Smith, M. D.; Shustova, N. . Photophysics, Dynamics, and Energy Transfer in Rigid Mimics of GFP-Based Systems. *Inorg. Chem.* **2016**, 7257–7264.

[19] Warner, K. D.; Chen, M. C.; Song, W.; Strack, R. L.; Thorn, A.; Jaffrey, S. R.; Ferré-D'Amaré, A. R. Structural Basis for Activity of Highly Efficient RNA Mimics of Green Fluorescent Protein. *Nat. Struct. Mol. Biol.* **2014**, *21*, 658–663.

[20] Paige, J. S.; Wu, K. Y.; Jaffrey, S. R. RNA Mimics of Green Fluorescent Protein. *Science* **2011**, *333*, 642–646.

[21] Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R. Broccoli: Rapid Selection of an RNA Mimic of Green Fluorescent Protein by Fluorescence-Based Selection and Directed Evolution. *J. Am. Chem. Soc.* **2014**, *136*, 16299–16308.

[22] Feng, G.; Luo, C.; Yi, H.; Yuan, L.; Lin, B.; Luo, X.; Hu, X.; Wang, H.; Lei, C.; Nie, Z.; Yao, S. DNA Mimics of Red Fluorescent Proteins (RFP) Based on G-Quadruplex-Confined Synthetic RFP Chromophores. *Nucl. Acids Res.* **2017**, *45*, 10380–10392.

[23] Song, W.; Filonov, G. S.; Kim, H.; Hirsch, M.; Li, X.; Moon, J. D.; Jaffrey, S. R. Imaging RNA Polymerase III Transcription Using a Photostable RNA–Fluorophore Complex. *Nat. Chem. Biol.* **2017**, *13*, 1187.

[24] Bozhanova, N. G.; Baranov, M. S.; Klementieva, N. V.; Sarkisyan, K. S.; Gavrikov, A. S.; Yampolsky, I. V.; Zagaynova, E. V.; Lukyanov, S. A.; Lukyanov, K. A.; Mishin, A. S. Protein Labeling for Live Cell Fluorescence Microscopy with a Highly Photostable Renewable Signal. *Chem. Sci.* **2017**, *8*, 7138–7142.

[25] Povarova, N. V.; Bozhanova, N. G.; Sarkisyan, K. S.; Gritcenko, R.; Baranov, M. S.; Yampolsky, I. V.; Lukyanov, K. A.; Mishin, A. S. Docking-Guided Identification of Protein Hosts for GFP Chromophore-Like Ligands. *J. Mater. Chem. C* **2016**, *4*, 3036–3040.

[26] Baranov, M. S.; Solntsev, K. M.; Baleeva, N. S.; Mishin, A. S.; Lukyanov, S. A.; Lukyanov, K. A.; Yampolsky, I. V. Red-Shifted Fluorescent Aminated Derivatives of a Conformationally Locked GFP Chromophore. *Chem. Eur. J.* **2014**, *20*, 13234–13241.

[27] Terasaki, M.; Loew, L.; Lippincott-Schwartz, J.; Zaal, K. Fluorescent Staining of Subcellular Organelles: ER, Golgi Complex, and Mitochondria. *Curr. Prot. Cell Biol.* **2001**, 4.4.1–4.4.18.

[28] Shim, S.-H.; Xi, C.; Zhong, G.; Babcock, H. P.; Vaughan, J. C.; Huang, B.; Wang, X.; Xu, C.; Bi, G.-Q.; Zhuang, X. Super-Resolution Fluorescence Imaging of Organelles in Live Cells with Photoswitchable Membrane Probes. *Proc. Nat. Acad. Sci.* **2012**, *109*, 13978–13983.

[29] Carlini, L.; Manley, S. Live Intracellular Super-Resolution Imaging Using Site-Specific Stains. *ACS Chem. Biol.* **2013**, *8*, 2643–2648.

[30] Chatterjee, T.; Mandal, M.; Gude, V.; Bag, P. P.; Mandal, P. K. Strong Electron Donation Induced Differential Nonradiative Decay Pathways for para and meta GFP Chromophore Analogues. *Phys. Chem. Chem. Phys.* **2015**, *17*, 20515–20521.

[31] Deng, H.; Yu, C.; Gong, L.; Zhu, X. Self-Restricted Green Fluorescent Protein Chromophore Analogues: Dramatic Emission Enhancement and Remarkable Solvatofluorochromism. *J. Phys. Chem. Lett.* **2016**, *7*, 2935–2944.

[32] Tsai, M.-S.; Ou, C.-L.; Tsai, C.-J.; Huang, Y.-C.; Sun, Y.-C. C. S.-S.; Yang, J.-S. Fluorescence Enhancement of Unconstrained GFP Chromophore Analogues Based on the Push–Pull Substituent Effect. *J. Org. Chem.* **2017**, *82*, 8031–8039.

[33] Baldridge, A.; Kowalik, J.; Tolbert, L. M. Efficient Synthesis of New 4-Arylideneimidazolin-5-ones Related to the GFP Chromophore by 2+3 Cyclocondensation of Arylideneimines with Imidate Ylides. *Synthesis* **2010**, *2010*, 2424–2436.

[34] Baleeva, N. S.; Myannik, K. A.; Yampolsky, I. V.; Baranov, M. S. Bioinspired Fluorescent Dyes Based on a Conformationally Locked Chromophore of the Fluorescent Protein Kaede. *Eur. J. Org. Chem.* **2015**, *2015*, 5716–5721.

[35] Kamlet, M. J.; Abboud, J. L. M.; Abraham, M. H.; Taft, R. W. Linear Solvation Energy Relationships. 23. A Comprehensive Collection of the Solvatochromic Parameters, π^*, α, and β, and Some Methods for Simplifying the Generalized Solvatochromic Equation. *J. Org. Chem.* **1983**, *48*, 2877–2887.

[36] 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.

[37] Baranov, M. S.; Solntsev, K. M.; Lukyanov, K. A.; Yampolsky, I. V. A Synthetic Approach to GFP Chromophore Analogs from 3-Azidocinnamates. Role of Methyl Rotors in Chromophore Photophysics. *Chem. Comm.* **2013**, *49*, 5778–5780.

[38] Carlotti, B.; Benassi, E.; Spalletti, A.; Fortuna, C. G.; Elisei, F.; Barone, V. Photoinduced Symmetry-Breaking Intramolecular Charge Transfer in a Quadrupolar Pyridinium Derivative. *Phys. Chem. Chem. Phys.* **2014**, *16*, 13984–13994.

[39] Epifanovsky, E.; Polyakov, I.; Grigorenko, B. L.; Nemukhin, A. V.; Krylov, A. I. Quantum Chemical Benchmark Studies of the Electronic Properties of the Green Fluorescent Protein Chromophore: I. Electronically Excited and Ionized States of the Anionic Chromophore in the Gas Phase. *J. Chem. Theory Comput.* **2009**, *5*, 1895–1906.

[40] Altoé, P.; Bernardi, F.; Garavelli, M.; Orlandi, G.; Negri, F. Solvent Effects on the Vibrational Activity and Photodynamics of the Green Fluorescent Protein Chromophore: A Quantum-Chemical Study. *J. Am. Chem. Soc.* **2005**, *127*, 3952–3963.

[41] 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.

[42] Dong, J.; Solntsev, K. M.; Tolbert, L. M. Solvatochromism of the Green Fluorescence Protein Chromophore and its Derivatives. *J. Am. Chem. Soc.* **2006**, *128*, 12038–12039.

[43] 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.

[44] de Wergifosse, M.; Houk, A. L.; Krylov, A. I.; Elles, C. G. Two-Photon Absorption Spectroscopy of trans-Stilbene, cis-Stilbene, and Phenanthrene: Theory and Experiment. *J. Chem. Phys.* **2017**, *146*, 144305.

[45] Chai, J. D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008**, *128*, *8*, 084106.

[46] Chai, J. D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *J. Chem. Phys.* **2008**, *10*, *44*, 6615-6620.

[47] Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M. A long-range-corrected density functional that performs well for both ground-state properties and time-dependent density functional theory excitation energies, including charge-transfer excited states. *J. Chem. Phys.* **2009**, *130*, *5*, 054112.

[48] Rhee, Y. M.; Head-Gordon, M. Scaled second-order perturbation corrections to configuration interaction singles: Efficient and reliable excitation energy methods. *J. Phys. Chem. A.* **2007**, *111*, *24*, 5314-5326.

[49] Krylov, A. I, Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: The hitchhiker's guide to Fock space. *Annual review phys. Chem.* **2008**, *59*.

[50] Leszczynski, J.; Shukla, M. Practical Aspects of Computational Chemistry II: An Overview of the Last Two Decades and Current Trends. *Springer Science & Business Media* **2012**.

[51] Plasser, F.; Wormit, F.; Dreuw, A. New tools for the systematic analysis and visualization of electronic excitations. I. Formalism. *J. Chem. Phys.* **2014**, *141*, *2*, 024106.

[52] Bäppler, S. A.; Plasser, F.; Wormit, M.; Dreuw, A. Exciton analysis of many-body wave functions: Bridging the gap between the quasiparticle and molecular orbital pictures. *Phys. Review A.* **2014**, *90*, *5*, 052521.

[53] You, Z. Q.; Mewes, J. M.; Dreuw, A.; Herbert, J. M. Comparison of the Marcus and Pekar partitions in the context of non-equilibrium, polarizable-continuum solvation models. *J. Chem. Phys.* **2015**, *143*, *20*, 204104.

[54] Mewes, J. M.; You, Z. Q.; Wormit, M.; Kriesche, T.; Herbert, J. M.; Dreuw, A. Experimental benchmark data and systematic evaluation of two a posteriori, polarizable-continuum corrections for vertical excitation energies in solution. *J. Phys. Chem. A.* **2015**, *119*, *21*, 5446-5464.

[55] Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. TB.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; others. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, *2*, 184-215.

[56] Krylov, A. I.; Gill, P. MW. Q-Chem: an engine for innovation. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, *3*, *3*, 317-326.

Chapter 4: Influence of the first chromophore-forming residue on photobleaching and oxidative photoconversion of EGFP and EYFP

4.1 Introduction

Fluorescent proteins (FPs) constitute a unique group of the genetically encoded fluorescence probes with the chromophore formed from their own amino acid residues. Genetic encodability and self-sufficient chromophore maturation determine the high value of FPs as the multipurpose imaging tools. Protein engineering has played an essential role in the development of the available FPs pallette, which currently includes dozens of spectral variants. Introduction of only two mutations (F64L and S65T) to the first described wild-type FP - avGFP - has resulted in EGFP[1], as yet most popular fluorescent protein. In EGFP, the chromophore dwells almost exclusively in the bright anionic state (fluoresces at ex=490 nm/2.53 eV, em=510 nm/2.43 eV), whereas the wild-type avGFP chromophore exists mostly in the protonated form (abs=395 nm/3.14 eV) and is weakly fluorescent. One more representative of the classic FPs - EYFP, in which Ser65Gly/Thr203Tyr substitutions lead to 35 nm fluorescence excitation bathochromic shift relative to avGFP, was also derived from avGFP[2].

The bicyclic chromophore of avGFP and most of its derivatives (including EGFP and EYFP) is formed from the -X65-Tyr66-Gly67- tripeptide motif by autocatalytic post-translational modification that involves consecutive cyclization, dehydration, and oxidation[3,4]. Tyr66 and

Gly66 are highly conservative in the native FPs.

Despite the variability of the first residue in a chromophore triad (X), proteins with different amino acids in this position form very similar chromophores. Thus, a typical GFP-like chromophore can be found in avGFP with SYG triad, EGFP with TYG, and EYFP with GYG. Although amino acid at the first position only weakly affects the structure of the chromophore core, it is essential for the interaction of the chromophore with its protein environment which, in turn, dramatically influences properties of the fluorescent protein, in particular those relevant for applications, such as fluorescence brightness and lifetime, Stokes shifts, and photostability[5–7]. The importance of the 65th position can be illustrated by the fact that Ser65 substitution by Gly, Ala, Cys, Val, or Thr led suppresses the shortwave neutral chromophore absorbance peak (395 nm/3.14 eV) in favor of the anionic chromophores peak at 470-490 nm (2.64-2.53 eV)[2,8,9]. In EGFP, the S65T mutation causes a significant rearrangement of the hydrogen-bond network in the chromophore region[2,10,11]: the threonine residue forms a new hydrogen bond with Val61[10]. Also, Thr65 induces Glu222 protonation and accelerates chromophore maturation (maturation time constant is 0.45 h in GFP-S65T versus 2 h in wild-type avGFP) speeding up the rate-limiting oxidation reaction rate[8,12]. In EYFP, S65G and V68L substitutions result in a 0.9 Åshift of the chromophore towards the barrel surface relative to its position in GFP-S65T and wild-type avGFP[13]. Both mutations also improve brightness of the cells expressing respective mutants relative to the avGFP-expressing cells, probably due to their effect on the protein folding or chromophore maturation[1]. Remarkably, the shift of the chromophore observed in EYFP (and connected particularly with the S65G substitution) leads to the appearance of the fluorescence sensitivity to halide and nitrate anions in this protein[14].

Mutational analysis of the first chromophore-forming amino acid position (Ser65 in avGFP) had been carried out in several studies and aimed primarily at determining the influence of this position on the maturation of the chromophore[10] and its basic spectral characteristics[2].

However, a systematic analysis of the influence of this position on the less obvious physico-chemical characteristics of fluorescent proteins, such as photostability, fluorescence lifetime, blinking, excited-state reactions, mediated by the molecular interactions of the chromophore with the nearest protein environment, to the best of our knowledge, has not yet been carried out. Existing data indicate that the amino acid in the 65th position has a significant impact on the GFP photophysics. For example, the EGFP mutants carrying the T65G substitution show significantly reduced quantum yield, shorter fluorescence lifetime, and an increased extinction coefficient[15,16]; an increased photostability of such proteins was also reported[16]. At the same time, in EYFP carrying the same GYG chromophore, the quantum yield is even higher and the lifetime is longer than in the EGFP with the TYG chromophore.

The effect of the 65th position on the ability of GFP-like proteins to undergo light-induced oxidative green-to-red photoconversion (called oxidative redding) is also of interest. Redding was described for green proteins of different taxonomic origin with Thr, Ser, Cys, Asn, Lys, and Gly in the first position of the chromophore, but the efficiency of the red spectral form appearance was maximal in EGFP (Thr65)[17]. As for EYFP, it is capable of redding only in the presence of halide anions, and even if they are present, it is much less effective than in EGFP[18]. The current mechanistic hypothesis[18] states that the redding is initiated by the electron transfer from the electronically excited chromophore to a nearby residue. Consequently, the effectiveness of this gateway step determines the ultimate yield of the red form. Under the same hypothesis, the yield of bleaching is also correlated with the effectiveness of the photoinduced electron transfer. The calculations of the energetics of one-electron oxidation and possible electron transfer pathways suggested that excited-state electron transfer proceeds through a hopping mechanism via Tyr145; the role of Tyr145 in redding has been confirmed by mutagenesis[18]. In YFPs, the -stacking of the chromophore with Tyr203 reduces its electron-donating ability, which can be restored by halide binding, due to its effect on the -stacking[18]. However, a

possible role of Gly65 was not investigated.

In this contribution, we examine the mutants of EGFP and EYFP proteins with reciprocal substitutions at the 65th position, EGFP-T65G and EYFP-G65T, focusing on their brightness, photostability, fluorescence lifetime, and redding ability compared to parental proteins. To rationalize the observed differences, we carried out quantum chemical and molecular dynamics simulations to estimate radiative and radiationless decay rates. On the basis of these calculations, we developed a kinetic model of the photocycle, which provides a unified picture of how the chromophores structure affects the photophysical properties of fluorescent proteins. The simulations revealed that the main effect of the T65G mutation is the reduced excited-state lifetime of the GYG chromophore, resulting in its increased photostability. The effect of the residue in position 65 on the brightness and quantum yield is explained by an interplay between the radiative and radiationless relaxation channels. The effect of the mutation 65 in EYFP is modulated by the -stacking interactions between the chromophore and Tyr203.

4.2 Results

4.2.1 Mutants general description (spectral characteristics)

EYFP-G65T and EGFP-T65G mutants generally showed spectral similarity to their parental proteins (Fig. 4.1). Like the original EGFP, EGFP-T65G has a single main absorption maximum, peaking at approximately 488 nm (2.54 eV) and corresponding to the anionic chromophore with fluorescence emission maximum at 510 nm (2.43 eV). The neutral (protonated) state of the chromophore in EGFP-T65G (absorption maximum 395 nm/3.14 eV) is minor, although it is more expressed than in EGFP, which is consistent with literature data on the role of Thr65 in maintaining the neutral state of Glu222 and the hydrogen-bond network favoring

chromophores deprotonation. The absorption spectra of EYFP-G65T, which have two pronounced maxima 410 and 513 nm (3.02 and 2.42 eV, respectively), corresponding to the neutral and anionic chromophores, is distinctly different from both the parent protein (EYFP) and from EGFP-T65G.

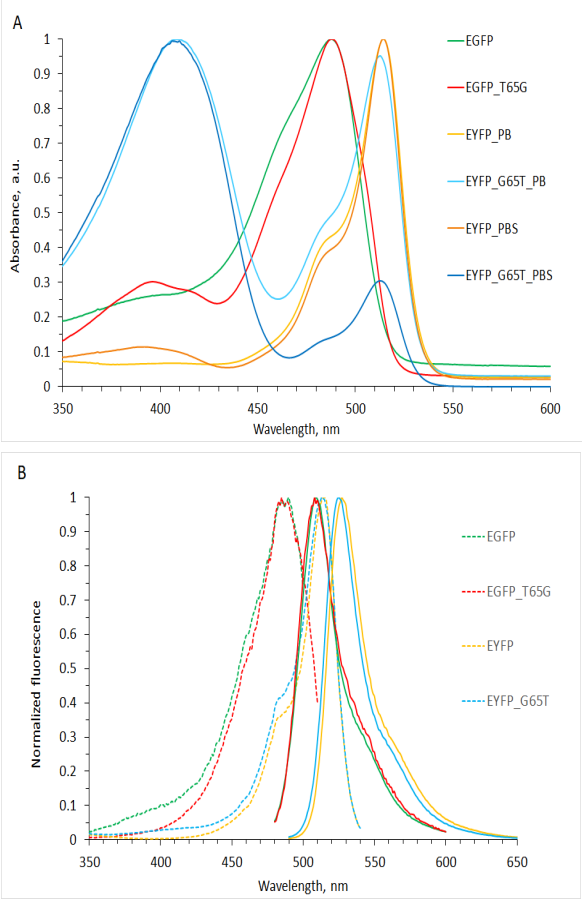


Figure 4.1: Absorption (A) and fluorescence (B) spectra of EGFP, EYFP, and mutants. In the fluorescence graph, dashed lines show fluorescence excitation, solid lines fluorescence emission. PB denotes phosphate buffer and PBS denotes phosphate buffered saline containing sodium chloride (see text).

A small (about 2 nm) blue shift in the anion and a significant (about 15 nm) red shift in the neutral chromophore absorption, which is unusual for proteins with the chromophore -stacked

with Tyr203 including EYFP[2], are noteworthy in the spectral comparison of EYFP-G65T with EYFP. Even more remarkable, however, is the observed dramatic dependence of the relative amplitudes of the peaks at 410 and 513 nm (3.02 and 2.42 eV) on the composition of the external environment. For example, in the hydrophosphate-dihydrophosphate buffer (PB, pH 7.4) the amplitudes ratio is about 1:1, while in the PBS buffer (pH 7.4, app. 140 mM Cl^-) the ratio becomes approximately 3:1 in favor of the neutral chromophore. Therefore, the protonation state of the EYFP-G65T chromophore seems to exhibit an enhanced sensitivity to the electrostatic interactions with the solvated ions. This property makes it a promising candidate for the sensitive core of the ratiometric halide ion sensor.

Parental EYFP also shows spectral sensitivity to the buffer content. That is, having essentially no absorbance around 400 nm (3.10 eV) in PB, in PBS it absorbs at 395 nm (3.14 eV) (as the classic protonated GFP-chromophore), while decreasing its main absorbance at 515 nm (2.41 eV) peak by circa 6%. EYFPs halide sensitivity, which is attributed to the shift of chromophores pKa induced by electrostatic interactions, is well-known; and fluorescence intensity decrease by about 40% at pH 7.0 was reported for this protein[14]. However, the contrast of the optical response to halides addition in the case of G65T mutant appears to be significantly higher.

4.2.2 Photostability

We measured the photostabilities of mutants versus parental proteins with and without electron acceptors in media, aiming to reveal the influence of the T65G/G65T substitutions on the primary excited-state electron transfer process that is believed to result in a permanent bleaching. Also, for EYFP/EYFP-G65T we introduced an additional variable - halide presence - to photostability measurements testing their possible role in the excited-state chemistry. The photostability is quantified by the bleaching half-times (the time it takes for the fluorescence to drop by a factor of two), i.e., longer half-times correspond to more photostable proteins. The

EGFP-T65G mutant (with enhanced photostability relative to EGFP in vitro and in cellulo[18]) showed approximately twofold higher photostability (relative photostability is defined as the ratio of the photobleaching rates) in PBS and almost 20-fold higher in PBS with 200 μM of potassium ferricyanide relative to those of EGFP (Fig. 4.2A, 4.2B, Table 4.1).

Table 4.1: Fluorescent properties of EGFP, EYFP, and their mutants, EGFP-T65G and EYFP-G65T.

FP	$\lambda_{ex}/\lambda_{em}$	EC	FQY	RB, %	FL PB	FL PBS	PB, S PB	PB, S PBS	PB, S PB+Ox	PB, S PBS+Ox	Redding PBS
EGFP	489/509	55000	0.60	100	n/d	2.8	n/d	80±10	n/d	5±2.5	strong
T65G	488/508	70000	0.06	13	n/d	1.3	n/d	170±25	n/d	85±15	weak
EYFP	514/526	67000	0.61	124	3.18	3.0	21±2	35±2	3±2	2±1	moderate
G65T	510/525	n/d	0.78	n/d	3.7	3.5	25±8	180±8	10±2	32±7	moderate
					4.0	0.5					
						3.9					

n/d = not determined.Relative brightness is calculated as a product of the molar extinction coefficient and the fluorescence quantum yield, and reported relative to the brightness of EGFP. For EGFP and EYFP, the absolute quantum yields are shown, for the mutants the quantum yields measured relative to the equally absorbing EGFP or EYFP are shown. For EYFP the single fluorescence lifetime value measured under 450 nm excitation is shown. For EYFP-G65T a pair of values measured under 400 and 490 nm excitation (value400/value490) is shown. In PBS, at 400 nm excitation EYFP-G65T showed fluorescence decay better fitted by bi-exponential function (τ_1=3.5 ns, τ_2=0.5 ns). Photobleaching is reported as the bleaching haf-time for each fluorescent protein, i.e., larger values correspond to the slower photobleaching rate and higher photostability.

Taking into account that EGFP-T65G (EC=70000 $M^{-1}cm^{-1}$; FQY=0.06) is 8 times dimmer than EGFP, one could explain the increase of photostability in PBS by its shorter excited-state lifetime (which is also responsible for its reduced emitter efficiency). However, the degree of the photostability increase in EGFP-T65G in the presence of oxidant does not match the degree of the proteins brightness decrease, and this probably indicates a less effective oxidative bleaching channel in this protein. The shapes of the bleaching curves with oxidant may favor this hypothesis: EGFP-T65G has a bimodal curve, but its fast component, probably related to excited-state electron transfer, is relatively short. EYFP and its mutant behavior seems more complex, in part due to the presence of an additional variable (presence or absence of the chloride anions in PBS or PB buffer, respectively) in the experimental conditions. In PB, EYFP-G65T photostability is close to that of the parental EYFP, while in PBS, the mutant shows circa 5-fold decreased photobleaching rate relative to EYFP under the same conditions and around 7-fold relative to itself in PB (Fig. 4.2C, Table 4.1). This observation is in accord with the absorption spectra behavior of two proteins (Fig. 4.1A): in the presence of 140 mM chloride only around 25% of EYFP G65T chromophore is in anion state and absorbs excitation light. In fact, EYFPs photostability also somewhat increases in the presence of chloride, probably for the same reason.

Oxidant addition (200 μM potassium ferricyanide) has significantly accelerated bleaching in all cases (Fig. 4.2D, Table 4.1). EYFPs photostability in PB with oxidant is reduced 6-7-fold. However, as in the case of EGFP/EGFP-T65G, the effect of ferricyanide on EYFP-G65T and EYFP behavior varies under different conditions. For EYFP with oxidant, the rate of photobleaching in PB and PBS is almost the same. One can suppose that the bleaching channel, which is dominant when electron acceptor is added, is less sensitive to chromophores pKa than the one that functions under normal conditions. In PB+oxidant, EYFP-G65T shows a

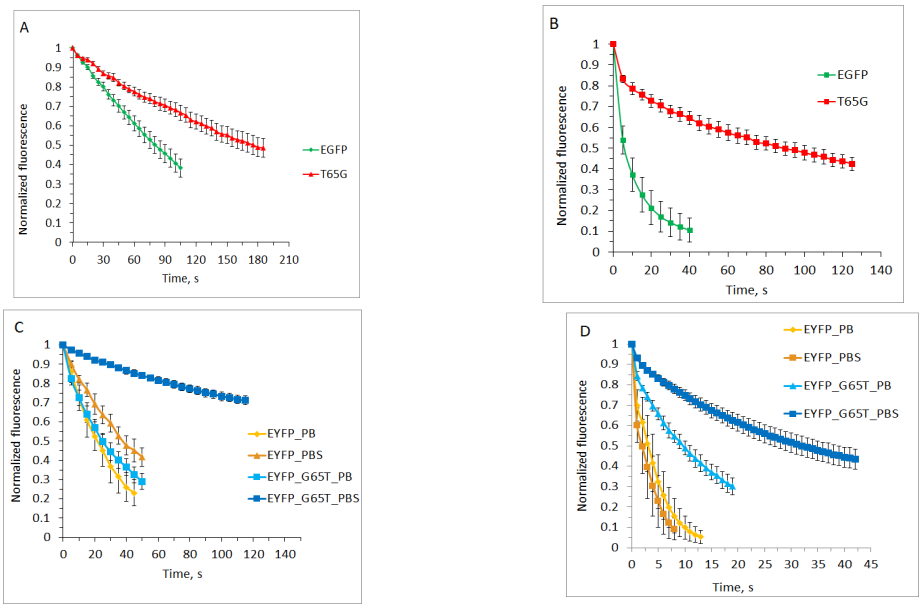


Figure 4.2: Bleaching kinetics in the immobilized proteins EGFP, EYFP, and their mutants in vitro. (A) Photoconversion of EGFP and EGFP-T65G in PBS; (B) Photoconversion of EGFP and EGFP-T65G in PBS in the presence of 0.2 mM potassium ferricyanide; (C) Photoconversion of EYFP and EYFP-G65T in PB and PBS (PBS contains potassium chloride); (D) Photoconversion of EYFP and EYFP-G65T in PB and PBS in the presence of 0.2 mM potassium ferricyanide. Green/yellow fluorescence intensities were background-subtracted and normalized to the maximum values. Standard deviation values (n = 1520 measurements in a representative experiment out of five independent experiments) are shown.

3-fold decrease in the photobleaching rate relative to EYFP (whereas without oxidant the rates are almost equal). In PBS+oxidant, the mutant demonstrates a 10-fold photostability increase relative to EYFP (versus 5-fold increase without oxidant), and a 3-fold increase relative to itself in PB+oxidant (versus a 6-fold increase without ferricyanide). Taken together, these ratios suggest that the oxidant reduces the dependence of the bleaching efficiency on the chromophore protonation state (see G65T-PB-ox versus G65T-PBS-ox, which changes in the presence of chloride), while the replacement of G65T generally disfavors the oxidative bleaching channel (see G65T-PB-ox versus EYFP-PB-ox).

4.2.3 Redding

We also tested redding efficiency among the mutants irradiated in presence of ferricyanide, our hypothesis being that the red form appearance rate should be inversely related to the photostability. This appears to be true in a pair of EGFP/EGFP-T65G where the parental protein demonstrated 20-30-fold more efficient redding (and 20-fold lower photostability) (Fig. 4.3A).

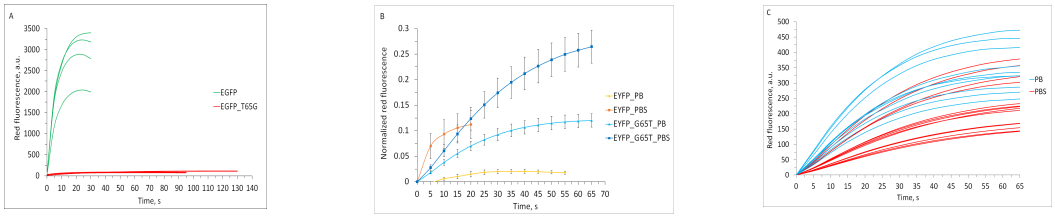


Figure 4.3: Redding kinetics in the EGFP, EYFP, and their mutants. (A) Appearance of red fluorescence in EGFP and EGFP-T65G. Non-normalized data for several measurements are shown. (B) Appearance of red fluorescence in EYFP and EYFP-G65T in PB and PBS (PBS contains potassium chloride). Averaged curves are shown. Red fluorescence intensities were background-subtracted and normalized to the maximum values. Standard deviation values (n = 1520 measurements in a representative experiment out of five independent experiments) are shown. (C) Appearance of red fluorescence in EYFP-G65T in PB and PBS (PBS contains potassium chloride). Non-normalized data for several measurements are shown.

As in the case of bleaching, the reduced rate of redding in EGFP-T65G cannot be explained only by the 8-fold lower relative brightness of this mutant, especially since it absorbs light even more effectively than the original protein (EC = 70000 versus 55000 in EGFP). When comparing the redding rates in different proteins, care should be taken to normalize the appearing red signal to the initial intensity of the green fluorescence. We suggest a normalization method adequate when studying redding of the same protein under slightly different conditions (for example, in cell culture). However, when comparing different proteins, this method can lead to artifacts because it does not take into account the chromophore's ability to absorb light and its quantum efficiency. We evaluated redding in EYFP-G65T and EYFP in PB (without chloride) and PBS (140 mM Cl^-), both supplemented with 200 μM of ferricyanide.

For EYFP in PB, we observed almost no detectable appearance of the red form (Fig. 4.3B). We do not consider the weak growth of the red signal visible on the graph to be reliable and attribute it to the imperfectness of the procedure of the subtraction of the red component of the main spectral form leaking through the RFP filter set (see Subtraction and normalization procedure in Supplementary materials). In PBS, we detected well-expressed (both in rate and absolute value) redding, in agreement with the observations reported earlier. EYFP-G65T undergoes redding both with and without chloride, demonstrating similar kinetics/rate but different yield (red signal plateau) under these two conditions (Fig. 3B). It is, however, possible that the seeming quantitative difference in the redding yields of EYFP-G65T in PB and PBS represents an artifact originating from an inadequate normalization procedure. To address this issue, we also compared non-normalized datasets for EYFP-G65T redding (Fig. 4.3C); this comparison did not confirm the trend exhibited by the normalized/averaged curves. Generally, the G65T mutation in EYFP enables the ability to undergo redding in the standard regime, i.e., independently on the halide binding. To compare quantitatively redding in PB and PBS, one should take into account an extreme sensitivity of the EYFP-G65T brightness to halide presence, which leads to a 2.5-3-fold difference in the green signal intensity at zero time.

4.2.4 Lifetime

Fluorescence lifetime of EGFP was measured to be 2.8 ns[13]; other studies estimated it in the range from 2.3 to 2.8 ns[16,19–22]. The spread in the reported values reflects the sensitivity towards the instrument and measurement conditions[16,19–22]. EGFP-T65G fluorescence lifetime (1.3 ns) is twice shorter than in EGFP[16].

For EYFP, fluorescence lifetime weakly depends on the halide presence (3.18±0.07 ns in PB (without Cl^-) and 3±0.08 ns in PBS (with Cl^-)), which is in rough agreement with the relevant data reported for the near homologs of EYFP[23,24]. Similarly to our observations

in the spectral domain, EYFP-G65T demonstrated a complex behavior in lifetime domain (Table 1). Thus, the mutant shows two clearly distinguishable lifetime values under 400 and 490 nm excitation wavelengths in the PB and PBS environment. In PB, both fluorescence decay kinetics can be fitted by the single-exponential functions with τ_{400}=3.7 ns, τ_{490}=4 ns. In PBS, excitation at 400 nm leads to a bi-exponential decay (τ_1=3.5 ns, τ_2=0.5 ns), where the faster component might be attributed to excited-state proton transfer (ESPT), although the ESPT kinetics is usually much faster than hundreds of picoseconds[25]. Excitation of the anionic form does not significantly change its fluorescent lifetime compared to the PB value.

4.2.5 Computational results

To rationalize observed differences in photophysical behavior of the four proteins due to the residue in position 65, we carried out the following quantum chemical and molecular dynamics calculations:

- Quantum-chemical calculations of the isolated model chromophores (structures, excitation energies, and oscillator strengths);

- Molecular dynamics simulations of the model proteins in the ground and electronically excited states;

- Hybrid QM/MM (quantum mechanics/molecular mechanics) calculations of the spectral properties of the model proteins (excitation energies and oscillator strengths for the structures taken from the ground-state molecular dynamics simulations).

The results from these calculations were used to estimate radiative and radiationless lifetimes, as described below. We considered four model systems, representing EGFP, EGFP-T65G, EYFP, and EYFP-G65T. For EYFP, we carried out calculations with and without chloride anions, as in Ref. 18. In all simulations, we considered only the deprotonated (anionic) chromophore. The protonation states of the protein residues were determined using Propka software[25–27] and verified by comparing the results of the molecular dynamics simulations with the available crystal structures (2Y0G for EGFP[9], 1F0B for EYFP), following the same protocols as in our earlier work[17] . Specifically, we determined that Glu222 is protonated (neutral) and His148 is neutral (HSD form, protonated at δN atom). The protonation state of Glu222 agrees with the conclusions of the experimental study. These protonation states give rise to a robust hydrogen-bond network around the chromophore, as shown in Fig. 4.4. The structures of the model chromophores and the definition of the QM/MM partitioning are shown in Fig. 4.5. Full details of the computational protocols are provided in Chapter 2.

We begin with characterization of the bare model chromophores. Figure 4.5 shows the structures of the isolated model chromophores and defines important structural parameters; it also shows how the chromophores are connected to the protein backbone. As one can see, the conjugated core of the TYG and GYG chromophores is the same; it comprises the phenolate and imidazolinone rings connected via the methine bridge. However, whereas the GYG chromophore is directly attached to the protein backbone through the exocyclic imidozalinones carbon, the TYG chromophore has an additional -CH(OH)CH3 tail attached to it. The presence of this tail has a relatively small effect on the excitation energy (red shift of about 0.02 eV), but leads to a 4% decrease in the oscillator strength of the bare chromophore. The -CH3 and -CH(OH)CH3 groups differ by their electron-donating ability - the presence of electronegative OH makes the latter a less effective electron donor. Thus, we attribute

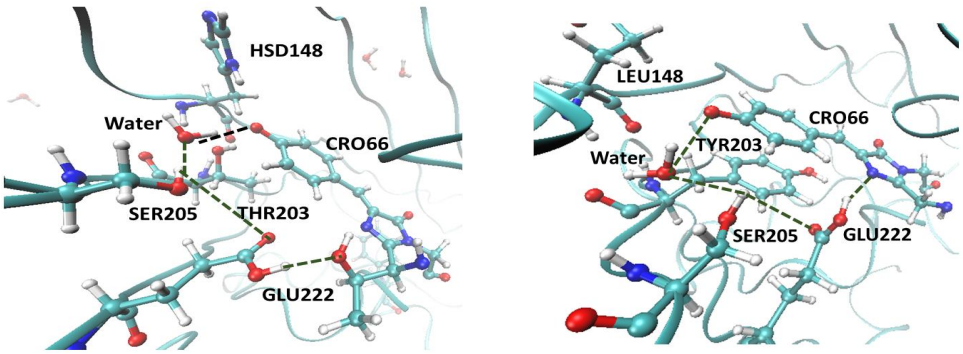


Figure 4.4: Hydrogen-bond network around the chromophore (CRO) in EGFP (left) and EYFP (right). The network includes CRO:O-water314-SER205-GLU222-CRO:O (Thr65, in EGFP). Glu222 is protonated and His148 is neutral in EGFP (protonated at $delta$N atom). Also shown is π-stacking of the chromophore and Tyr203 in EYFP.

the larger oscillator strength in GYG relative to TYG to an increased electron density in the conjugated part of the chromophore due to stronger electron-donating ability of -CH3. To test this hypothesis, we carried out calculations for a fluorinated GYG chromophore in which one -CH3 group was replaced by -CF3. The fluorinated GYG chromophore shows significant reduction (7.3%) of the oscillator strength relative to the GYG chromophore, consistently with strong electron-withdrawing ability of fluorine (see Fig. 4.6).

As discussed below, larger oscillator strength in GYG chromophore contributes to its increased brightness and reduced radiative lifetime (i.e., faster fluorescence). Importantly, the tail has a major effect on the hydrogen-bond network around the chromophore, its planarity and conformational flexibility. Figure 4 shows the hydrogen-bond networks around the chromophore in EGFP and EYFP. As one can see, in both proteins there are 4 hydrogen bonds around the chromophore. However, the Ser205-Glu222 distance (Ser205:O - Glu222:OE1) is much larger

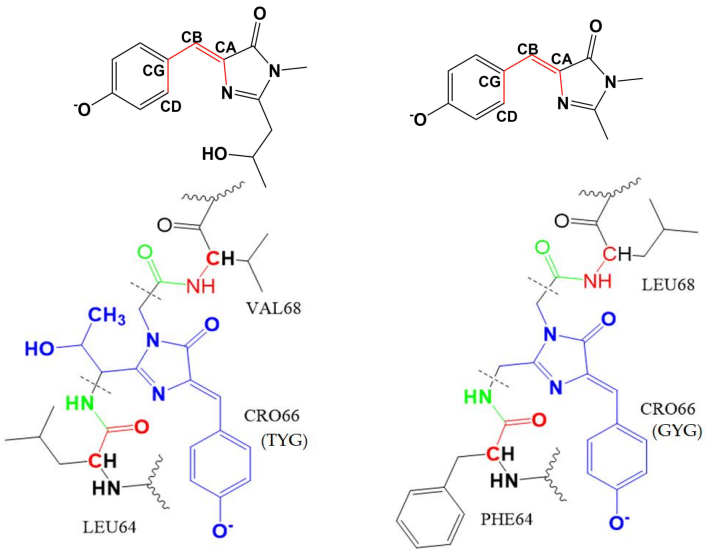


Figure 4.5: Top: Structures of the model TYG (EGFP, YFP-G65T) (left) and GYG (YFP, EGFP-T65G) (right) chromophores. Torsional angles ϕ and τ are defined as CD-CG-CB-CA and CG-CB-CA-N, respectively. The difference between the two angles Δ=ϕ-τ quantifies whether the chromophore is planar (Δ=0) or not. Bottom: the QM/MM partitioning for EGFP (left) and EYFP (right). Blue color denotes the QM region and the black dotted lines denote the QM-MM boundary. Charges of red and green atoms were set to zero in the MM region. In EGFP-T65G, the chromophore is GYG and the neighboring residues are the same as in EGFP. Likewise, in EYFP-G65T, the chromophore is TYG and the neighboring residues are the same as in EYFP.

(3.74 Åand 4.27 Åin EGFP and EYFP, respectively) in EYFP. By allowing for larger thermal fluctuations causing transient breaking of the hydrogen bonds, a larger Ser205-Glu222 distance signifies a weaker hydrogen-bond network, which is illustrated by the results in Table 4.2.

Table 4.2. summarizes the analysis of the hydrogen-bond pattern and deviations of the chromophore from planarity in the course of the ground-state equilibrium dynamics. The average number of hydrogen bonds is smaller for the GYG chromophores compared to the proteins with TYG chromophores. Interestingly, despite a smaller number of hydrogen bonds, the deviation from planarity is smaller for EGFP-T65G relative to EGFP, both in terms of the average values

Table 4.2: Average number of hydrogen bonds (and standard deviation) formed within 6 Åaround the chromophore computed along the equilibrium trajectories. Distance and angle cut off were set to 3.2 Åand 20°, respectively). Deviation of the chromophore from planarity (Δ, in degrees) is also shown.

Protein/Chro	EGFP/TYG	EGFP-T65G/GYG	EYFP/GYG	EYFP-G65T/TYG	EYFP+Cl^-/GYG
No. H-bond	2.81	2.31	1.34	1.93	1.45
STD(hbond)	1.12	1.03	0.83	1.06	0.87
Δ	7.40	6.44	4.41	5.50	7.02
STD (Δ)	16.7	8.2	7.8	8.0	7.29

of Δ and in terms of standard deviation. The latter indicates a larger dynamic range of chromophore motions in EGFP. A reduced range of thermal torsional motions in EGFP-T65G and smaller deviations from planarity are probably due to the bulkier size of the TYG chromophore. The EYFP chromophore shows smaller deviations from planarity, because of the stabilizing effect of the π-stacking with Tyr203 (this is consistent with the observations in Ref. 18). The average number of hydrogen bonds around the chromophore is larger in EYFP-G65T than in EYFP because the OH group of the threonine participates in the hydrogen-bond network.

Torsional motions of the chromophore modulate the oscillator strength of the S_0-S_1 transition, as illustrated in Fig. 4.6. In EGFP, the standard deviation for Δ (which quantifies the twisting motion of the chromophore) is 17 degrees. In this range of motion, the oscillator strength can be reduced by several percent (Fig. 4.6 shows that the oscillator strength depends quadratically on ϕ). These results explain the variations in the average oscillator strengths for the S_0-S_1 transition for the four systems discussed below.

On the basis of the QM/MM calculations of the transition energies and oscillator strengths, we estimate intrinsic fluorescence lifetime, τ_{fl}. Intrinsic radiative lifetime is inversely proportional to the oscillator strength of the transition (f_l) and to the square of corresponding excitation energy (E_{ex}). In atomic units, intrinsic radiative lifetime τ_{fl}[28] is:

$$\frac{1}{\tau_r} = \frac{\omega_0^2 f_{abs}}{2\pi(c')^3\epsilon}, \tag{4.1}$$

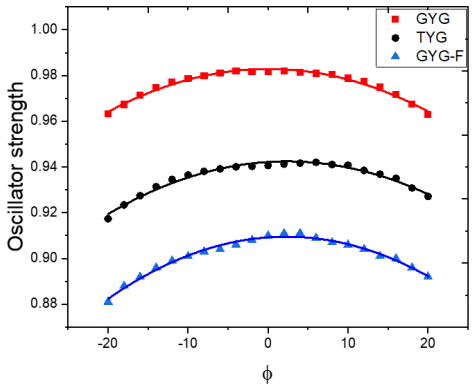


Figure 4.6: Oscillator strength for the S_0-S_1 transition in the isolated TYG, GYG, and fluorinated GYG (GYG-F in which one -CH3 is replaced with -CF3) chromophores along torsional angle ϕ (all other degrees of freedom are relaxed) computed with ωB97X-D/aug-cc-pVDZ.

where c' is the speed of light in the medium ($c' = c/n$; c is the speed of light in vacuum and n is the index of refraction) and ϵ is the dielectric constant. For vacuum, ϵ=1 and c=137. Dielectric constant in proteins is small (i.e., 2-8). The index of refraction of water is 1.33; the refractivity of protein solutions is generally larger, around 1.6[29,30].

Table 4.3: Theoretical estimates of radiative lifetime for different mutants. Computed excitation energies and oscillator strengths are also shown. QM/MM absorption energies and oscillator strengths are averaged over 21 snapshots taken from ground-state equilibrium molecular dynamics simulations. τ_{fl},rel values are relative lifetimes calculated with respect to τ_{fl} in EGFP.

Mutant	E_{ex}, eV (f_l) (gas)	E_{ex},eV (f_l) (QM/MM)	τ_{fl}, ns gas, n=1	τ_{fl}, ns QM/MM, n=1	τ_{fl}, ns QM/MM, n=1.6	τ_{fl} rel, ns gas	τ_{fl} rel, ns QM/MM
EGFP	3.101 (1.02)	3.081 (0.97)	29.50	31.24	7.63	1.00	1.00
EGFP-T65G	3.123 (1.05)	3.142 (1.04)	28.25	28.18	6.88	0.95	0.90
EYFP	3.123 (1.05)	3.097 (1.05)	28.25	28.71	7.02	0.95	0.92
EYFP-G65T	3.123 (1.05)	3.015 (0.98)	28.25	32.49	7.94	1.00	1.04
EYFP+Cl^-	3.123 (1.05)	3.077 (1.07)	28.25	28.57	6.97	0.95	0.91

Table 3 shows excitation energies of the isolated chromophores and average excitation energies and oscillator strengths computed for 21 QM/MM snapshots taken from the ground-state equilibrium trajectories. These values are used to compute radiative lifetimes by Eq. (1) with

n=1 and ϵ=1. The absolute values of the computed lifetimes are almost 10 times longer than the experimentally observed fluorescence lifetimes, which is expected given the uncertainties in the computed values and the key constants (i.e., n and ϵ). Using n=1.6 brings the computed values down, to the range of 7-8 ns. Moreover, we note that Eq. (1) provides only an upper bound of τ_{fl} and does not account for other decay channels available to such complex polyatomic systems as fluorescent proteins. However, we expect that these calculations capture the essential trend of variations in the intrinsic fluorescence lifetime due to the variations in the oscillator strength induced by thermal motions and differences in the chromophores structure. To zoom into this trend, the last two columns of Table 4.3 show relative values of the computed fluorescence lifetimes with respect to that of EGFP. As one can see, the proteins with the GYG chromophore are expected to have intrinsic fluorescence lifetimes shorter by 8-12% than their counterparts with TYG. This difference is due to slight red shifts, dynamically reduced oscillator strengths in TYG, and the electronic effect of OH, all caused by the bulkier and more electronegative threonine group.

In contrast to a relatively modest effect of the residue in position 65 on the chromophore structure in the ground state, it has a dramatic effect on the excited-state potential energy surface indicates a significant effect on subsequent dynamics of the chromophore following photoexcitation. The origin of this strong effect of hydrogen bonding is a much flatter torsional potential of the chromophore in the excited state[31,32]. Fig. 4.7 shows the scans of potential energy surfaces in the ground and the first excited state of the isolated GYG chromophore along the two torsional angles. As one can see, the chromophore in its ground state is rather rigid due to its π-conjugated system: the barriers for the ϕ (phenolate flip) and τ (imidozalinone flip) rotations are about 31.61 and 34.47 kcal/mol, respectively. However, in the S_1 state (which has π-π^* character), the bond order is reduced, giving rise to rel4tively flat potential energy profiles along the twisting coordinates (the computed barriers for the ϕ and τ rotations are 3.59

and 4.52 kcal/mol, respectively). These flat profiles are responsible for low FQY of isolated chromophores[31–34]. The hydrogen-bond network around protein-bound chromophores plays a crucial role by stabilizing the otherwise floppy structure in a planar configuration, thus preventing the chromophores trapping in dark twisted states and suppressing the radiationless relaxation via conical intersections[31–33]. That is why different hydrogen-bond patterns around GYG and TYG chromophores have a profound effect on their excited-state dynamics. Specifically, as illustrated by excited-state molecular dynamics simulations, GYG chromophores are much more likely to twist in the excited state than the TYG chromophores.

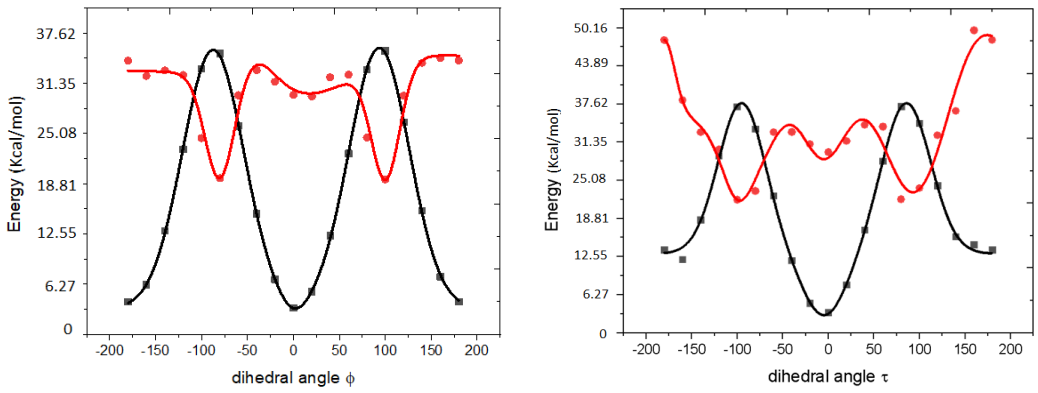


Figure 4.7: PES scans (relative energies) for the isolated GYG chromophore along the dihedral angles ϕ (left) and τ (right) in the ground (black) and electronically excited (red) states. All other degrees of freedom are frozen. The dots represent ab initio calculations (ωB97X-D/aug-cc-pvDZ) and the solid lines are fits to the force-field torsional potential used in molecular dynamics simulations (see Chapter 2). In contrast to the isolated chromophores, the protein-bound excited chromophores can only undergo phenolate flip (ϕ twist) because the imidozalinone ring is covalently bound to the protein backbone.

To quantify excited-state evolution, we carried out molecular dynamics simulations using the modified force-field parameters (see Appendix B and Fig. 4.11, 4.12) designed for the S_1 state. Starting from 101 snapshots harvested from the ground-state trajectories for each protein, we propagated excited-state trajectories for 3 ns; the results were saved each 2.5 ps. To estimate the rate of radiationless relaxation, we monitored the dihedral angle along simulation trajectories and defined two populations: A (planar chromophore, defined as $\phi < 50°$) and B (twisted chromophore, $\phi > 50°$). The dihedral angle τ does not fluctuate significantly (20°) in the course of dynamics, because of the covalent bond between the imidazolinone ring and the proteins backbone. Once the value of ϕ reached the critical value of 50°, we stopped the trajectory assuming that strongly twisted structures undergo fast and irreversible non-adiabatic transitions to the ground state. Fig. 9 shows the evolution of the two populations (A and B) in the studied proteins. As one can see, in the EGFP-T65G mutant all chromophores eventually undergo twisting in the course of excited-state dynamics. The twisting dynamics can be used to estimate the rate of the radiationless relaxation using 1st order kinetics fit of A(t):

$$A(t) = e^{-kt}\tau_{nr} = \frac{ln2}{k} \tag{4.2}$$

where radiationless (non-radiative) half-life is τ_{nr}.

Fitting the decay of the planar population (shown in Fig. 4.8) by a first-order kinetics yields half-lives of 5.92 ns and 0.25 ns for EGFP and EGFP-T65G, respectively. In EYFP and EYFP-G65T the computed half-lives are 1.73 ns and 10.8 ns, respectively. These numbers roughly correspond to the excited-state decay via radiationless relaxation. As one can see, the T65G mutation in EGFP leads to a 23-fold drop in the non-radiative lifetime, which is in a semi-quantitative agreement with the 10-fold drop in FQY. The effect of the mutation of residue 65 in EYFP is slightly smaller, only 6-fold, which is qualitatively in agreement with 30% larger FQY in EYFP-G65T relative to EYFP. Addition of halide to EYFP leads to faster twisting by a

factor of three, because halide binding upsets -stacking of the chromophore with Tyr203.

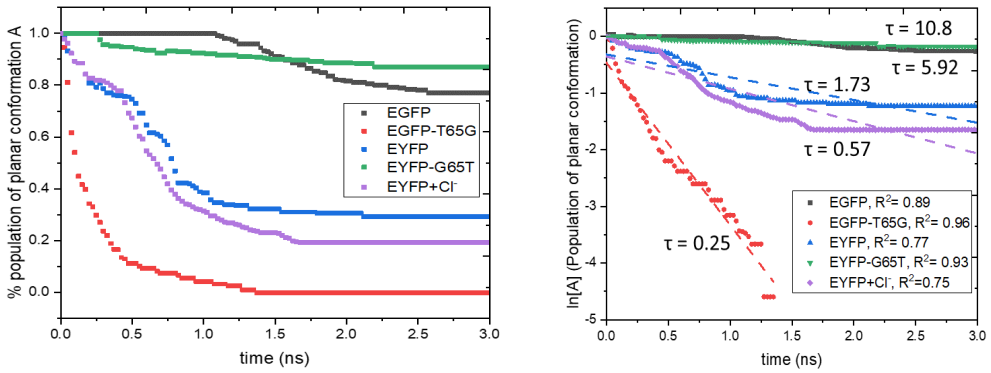


Figure 4.8: Left: Evolution of planar (A) population in excited-state molecular dynamics simulations of EGFP, EGFP-T65G, EYFP, EYFP-G65T, and EYFP+Cl⁻. Right: Linear fit for ln[A].

4.3 Discussion

Photophysical properties of the fluorescent proteins are determined by an interplay between chromophores intrinsic electronic structure, its interactions with the surrounding residues, and several competing excited-state processes. We begin by outlining the connection between the macroscopic observables (extinction coefficients, brightness, and photostability) with the microscopic properties of the chromophores. The extinction coefficient is proportional to the intrinsic

brightness of the chromophore as characterized by the oscillator strength of the S_0-S_1 transition. Apparent excited-state lifetime (τ) is a result of the intrinsic fluorescence lifetime, τ_{fl}, and various non-radiative decay channels (τ_{nr}):

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}} \tag{4.3}$$

$$\tag{4.4}$$

The non-radiative channels include radiationless relaxation and bleaching. However, given the small quantum yield of bleaching in typical fluorescent proteins ($<10^{-5}$)[38], the second term in Eq. (4) is dominated by the radiationless relaxation lifetime. FQY is determined by the ratio of the radiative and radiationless lifetimes:

$$FQY = \frac{\tau_{nr}}{\tau_{fl} + \tau_{nr}} \tag{4.5}$$

That is, for a given τ_{fl} , FQY is larger when radiationless decay is slow (longer τ_{nr}). Conversely, for a fixed τ_{nr} , FQY is larger for systems with shorter radiative lifetime. Intrinsic radiative lifetime is related to the chromophores excitation energy and oscillator strength by Eq. (1), i.e., larger oscillator strength leads to shorter radiative lifetimes (i.e., faster fluorescence rate). The photostability of the fluorophores depends on the ratio of excited-state lifetime and the rate of the bleaching process, i.e., within the first-order kinetics, the yield of bleaching Y_{bl} can be estimated as:

$$Y_{bl} = \frac{\tau}{\tau_{bl}} \tag{4.6}$$

This means that for a given rate of the bleaching process (via photo-oxidation or other photochemical processes), the yield of bleached forms is smaller for systems with shorter apparent excited-state lifetimes. As photostability is inversely proportional to Y_{bl}, the ratios of $1/Y_{bl}$ can be interpreted as relative photostabilities. Of course, bleaching rates can vary significantly among different proteins, because electron-transfer pathways and the rates are

sensitive to mutations[17]. Because of the high cost of such calculations, the effects of mutations on the rates of electron transfer are not investigated in the present study.

Our simulations suggest that the principal effect of the T65G mutation is two-fold: (i) it increases the oscillator strength, leading to shorter fluorescence lifetimes; and (ii) it increases chromophores flexibility in the excited state, leading to faster radiationless relaxation. These two trends qualitatively explain all results from Table 1: EGFP-T65G has a larger extinction coefficient than EGFP because of its larger oscillator strength; Because of the faster radiationless relaxation, EGFP-T65G has lower FQY than EGFP; likewise, EYFP has lower FQY than EYFP-G65T; Having glycine in position 65 leads to faster radiationless relaxation (shorter half-life of A), thus suppressing the bleaching and leading to an increased photostability; Larger FQY in EYFP relative to EGFP-T65G arises due to the suppression of torsional motions by the π-stacking interactions, which is reflected in longer radiationless half-life. To relate these calculations to the photophysical properties of the proteins, we collect the estimated half-lives due to radiationless relaxation and estimated radiative half-lives in Table 4. Fig. 9 shows the comparisons between theory and experiment graphically. We estimate FQY and the relative rates of bleaching using Eqns. (5) and (6). Although the computed FQY is not in quantitative agreement with the experimental values (which is not surprising, given that several there are several approximations in the model), the differences between the mutants are described rather well: compare, for example, the ratio of FQY in EGFP-T65G and EGFP: 0.09 (computed) versus 0.1 (experimental). Likewise, the computed ratio of FQY in EYFP and EYFP-G65T is 0.34, to be compared with the experimental ratio of 0.78. The trend in the apparent fluorescence lifetime (τ) is also captured reasonably well: the computed relative lifetimes are 1:0.1:0.4:1.4, to be compared with the experimental ratios of 1:0.4:1.1:1.4 (estimated from the fluorescence lifetimes from Table 4.1).

Finally, the estimated bleaching yield (assuming the same rate of bleaching process) ratios are

1:0.1:0.4:1.4. The relative photostabilities can be estimated as the ratios of the inverse of Y_{bl}: 1:10:2.5:0.71. The experimental macroscopic bleaching half-lives, which are inversely proportional to the rate of bleaching and can, therefore, be interpreted as relative photostabilities, are 1:2.1:0.4:2.2 (see Table 1, PBS). As anticipated above, the agreement here is worse, because the present study does not account for the variations in electron-transfer rates due to the mutations. The rate of bleaching is expected to vary between different systems, because the rate of electron transfer is sensitive to structural variations, especially between EGFP and EYFP[17]. In EYFP, the rate of bleaching is strongly affected by the halide binding, for example, the calculated electron-transfer rate to Tyr145 are five orders of magnitude faster in EYFP+Cl^- (kEYFP+Cl^-: kEYFP = 13300 :1), which overshadows small variations in radiationless relaxation rates.

Table 4.4: Computed radiative and radiationless lifetimes of EGFP, EGFP-T65G, EYFP, EYFP-G65T, and EYFP+Cl^- (in parenthesis, the values relative to EGFP are shown) and estimated photophysical parameters.

Protein	τ_{fl}, ns ($\tau_{fl,rel}$	τ_{nr}, ns $\tau_{nr,rel}$	τ, ns (τ_{rel})	FQY	$Y_{bl,rel}$
EGFP	7.63 (1.00)	5.92 (1.00)	3.33 (1.0)	0.44	1.0
EGFP-T65G	6.88 (0.90)	0.25 (0.04)	0.24 (0.1)	0.04	0.1
EYFP	7.02 (0.92)	1.73 (0.29)	1.39 (0.4)	0.20	0.4
EYFP-G65T	7.94 (1.04)	10.8 (1.82)	4.58 (1.4)	0.58	1.4
EYFP+Cl^-	6.97 (0.91)	0.57 (0.10)	0.53 (0.2)	0.07	0.2

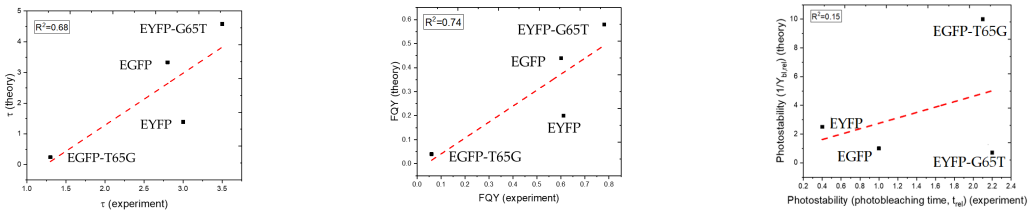


Figure 4.9: Correlation between theoretical and experimental apparent fluorescence lifetimes (left), FQY (middle), and the rate of bleaching (right).

We conclude this section by considering the implications of the above findings for imaging applications. Faster radiationless relaxation of EGFP-T65G may result in the suppression of

not only the bleaching but also other excited-state processes. This might be valuable from the practical point of view, for example, in the context of live-cell imaging where fluorophores with the decreased photoreactivity would potentially demonstrate such advantages as the decreased phototoxicity and fewer artifacts due to the redox-photoreactions with the intracellular compounds.

The high spectral sensitivity of EYFP-G65T to chloride (and probably to other anions of similar size) may open up new avenues for the design of FP-based molecular indicators, including those functioning in the lifetime domain. This is of special interest since EYFP and its circularly permuted variants have been utilized in several popular indicators[35–37].

4.4 Materials and Methods

The experimental measurements were carried out as follows. His-tagged proteins were expressed in Escherichia coli and purified by a metal-affinity resin. The resin beads with immobilized proteins were placed into phosphate buffer (PB, pH 7.4), or phosphate buffered saline (PBS, pH 7.4), or PB/PBS with 0.2 mM potassium ferricyanide as an oxidant and illuminated with strong blue light using a fluorescence microscope. Changes of fluorescence in green/yellow and red channels were monitored during illumination.

4.4.1 Spectroscopy and fluorescence brightness evaluation

For absorbance and fluorescence excitation-emission spectra measurements, Cary 100 UV/VIS spectrophotometer and Cary Eclipse fluorescence spectrophotometer (Varian) were used. Fluorescence brightness was evaluated as a product of molar extinction coefficient by quantum yield multiplication. Measurements on all native proteins were carried out in

phosphate buffered saline (PBS, pH 7.4, Gibco). For molar extinction coefficient determination, we relied on measuring mature chromophore concentrationn[38]. EYFP and its mutants were alkali-denatured in 1 M NaOH. Under these conditions GFP-like chromophore is known to absorb at 447 nm with extinction coefficient of 44,000 M^1 cm^1. Based on the absorption of the native and alkali-denatured proteins, molar extinction coefficients for the native states were calculated. For determination of the quantum yield, the areas under fluorescence emission spectra of the mutants were compared with equally absorbing EYFP (quantum yield 0.61)[38] and EGFP (quantum yield 0.60)[38].

4.4.2 Microscopy

For wide-field fluorescence microscopy, a Leica AF6000 LX imaging system with Photometrics CoolSNAP HQ CCD camera was used. Green and red fluorescence images were acquired using 63×1.4NA oil-immersion objective and standard filter sets: GFP (excitation BP470/40, emission BP525/50) and TX2 (excitation BP560/40, emission BP645/75). Photobleaching and redding were monitored in time-lapse imaging in the green and red channels at low light intensity combined with exposures to blue light of maximum intensity (GFP filter set, light power density of 2-3 W/cm^2). Images were acquired and quantified using Leica LAS AF software.

4.4.3 Protein Expression and Purification

EYFP and EGFP as well as EGFP-T65G and EYFP-G65T mutants were cloned into the pQE30 vector (Qiagen) with a 6His tag at the N terminus, expressed in E. coli XL1 Blue strain (Invitrogen) and purified using TALON metal-affinity resin (Clontech).

4.4.4 Site-Directed Mutagenesis

The EGFP-T65G, and EYFP-G65T mutants were generated using overlap-extension PCR technique with the following oligonucleotide set containing the appropriate substitutions: forward 5-ATGCGGATCCATGGTGAGCAAGGGCGAG-3, reverse 5-ATGCAAGCTTTTACTTGTACAGCTCGTC-3 and forward 5-ACCACCTTCACCTACGGCCTG-3 and reverse 5-CAGGCCGTAGGTGAAGGTGGT-3 for EYFP G65T; forward 5-ATGCGGATCCATGGTGAGCAAGGGCGAG-3, reverse 5-ATGCAAGCTTTTACTTGTACAGCTCGTC-3 and forward 5-ACCACCTTCGGCTACGGCCTG-3, reverse 5-CAGGCCGTAGCCGAAGGTGGT-3 for EGFP-T65G. For bacterial expression, a PCR-amplified BamHI/HindIII fragment encoding an FP variant was cloned into the pQE30 vector (Qiagen).

4.4.5 Fluorescence lifetime imaging microscopy of the purified proteins upon single-photon excitation.

Femtosecond laser pulses (80 MHz repetition rate, up to 100 fs, up to 25 nJ per pulse) were generated by a Ti:Sapphire oscillator (Tsunami, Spectra-Physics) pumped by a green Nd:YVO4 CW laser (532 nm, Millennia Prime 6sJ, Spectra-Physics) and frequency doubled in an LBO nonlinear crystal (Spectra-Physics). Second harmonic laser beam was coupled to an inverted optical microscope Olympus IX71 by a Thorlabs FESH0750 dielectric filter mounted at 45° and then focused by objective lens (40 $\times$ 0.75NA, UPlanFLN, Olympus) on a sample, which was placed on a 3-axis stage. The samples were prepared as droplets of the purified fluorescent proteins dissolved in phosphate buffered saline (PBS, pH 7.4, GIBCO) applied onto a standard 24 24 mm cover glass (Heinz Herenz, Germany). The average laser power was tuned with a polarizing attenuator and further attenuated with a glass neutral filter. A

typical laser power coupled to the microscope was about 3 μW. The central wavelength of the fundamental harmonic pulses was either 800 or 980 nm, and of the second harmonic pulses 400 or 490 nm respectively. The SF10 prism compressor was used to compensate for the group velocity dispersion in the objective lens and other optical elements. Fluorescence was excited by one-photon absorption of femtosecond laser, passed back through the objective lens and laser coupling filter, was filtered by a long-pass dielectric filter (FELH0500, Thorlabs), and then was directed to the input of Acton SP300i monochromator with two separate outputs. PI-MAX 2 CCD camera (Princeton Instruments) at the first output was employed for the fluorescence spectra registration. Photomultiplier tube of the time-correlated single photon counting system SPC-730 (Becker and Hickl GmbH) at the second output detected the fluorescence decay kinetics in the 510 nm530 nm band. Fluorescence decay data were primarily acquired using SPCImage software (Becker and Hickl, Germany) and then exported in ASCII format and analyzed using Origin Pro 9 software (OriginLab, USA).

4.4.6 Computational details

Protonation state and crystal structures.

Structures of EGFP and EYFP were taken from the protein data bank (PDB) with ids: 2Y0G and 1F0B respectively. The protonation states of titratable residues were determined using *PropKa* software[25–27]. Particularly important are the protonation states of the residues around the chromophore: the Glu 222 and His 148. *PropKa*[25] predicts the glutamate to be neutral (GLUP 222) for both EGFP and EYFP. To validate this prediction, we carried out separate molecular dynamics simulations of EGFP and EYFP with protonated and deprotonated Glu222 and analyzed the key interatomic distances from the equilibrium trajectories. Direct comparison of these calculated distances with the crystal structures confirmed dominant population of the protonated Glu222 residue. This conclusion is also in accord with the prior experimental

findings. In EGFP, we visually inspected the local environment around the His148 residue and concluded that it exists in HSD (neutral, protonated at δN atom) form because of hydrogen bonding with the phenolate oxygen of the chromophore. Additional confirmation of HSD protonation state was obtained from similar His148-chromophore distances from equilibrium molecular dynamics simulation and the crystal structure. The resulting hydrogen-bond patterns around the chromophore are shown in Fig. 4.4.

Molecular dynamics setup.

We used CHARMM27 force-field parameters for protein residues[39] and the ground-state anionic chromophores force-field parameters were obtained from an old work[40]. Charged amino acids on the surface were locally neutralized by adding counter ions close (4.5 Å) to them. Charged residues that do not form a salt bridge inside the protein barrel were also neutralized by adding appropriate counter ions at the surface. This neutralization protocol resulted in the addition of 21 Na^+ and 14 Cl^- in the EGFP, and 20 Na^+ and 14 Cl^- in the EYFP. The proteins were solvated in water boxes producing a buffer of 15 Åwith the box size of approximately 69 Å 77 Å 75Å. The TIP3P[41] water model was used to describe external waters. Molecular dynamics simulations were performed using these solvated neutralized model structures as follows:

- Minimization for 2000 steps with 2 fs time step.

- Equilibration of the solvent (keeping the protein frozen) for 500 ps with 1 fs time step.

- Equilibration of the system for 2 ns (with 1 fs time step) with periodic boundary condition (PBC) under the isobaric-isothermal NPT ensemble.

- Production run for 2 ns with 1 fs time step in an NPT ensemble. Molecular dynamics simulations were performed with NAMD[42] in an NPT ensemble with Langevin dynamics. Pressure and temperature were kept at 1 atm and 298 K during the simulation.

QM/MM setup.

We computed electronic properties (vertical excitation energies, oscillator strengths) using snapshots generated along equilibrium trajectories (production runs of molecular dynamics simulations) using a QM/MM scheme. The chromophore is included in the QM region and the rest of the system is treated as fixed MM point charges (see fig. 4.5) via electrostatic embedding. Hydrogen atoms were added at the QM/MM boundary to saturate the valencies. Point charges on the red and green atoms in Fig. 5 were set to zero and the excess charge was redistributed over the rest of the atoms of the respective residues to avoid over-polarization of the QM atoms at QM/MM boundary. Electronic structure calculations were performed at the ωB97X-D/aug-cc-pVDZ[43,44] level of theory. Benchmark results using different electronic structure methods are presented in the Appendix. All quantum chemistry and QM/MM calculations were carried out using the Q-Chem electronic structure package[45].

4.5 Conclusions

In this contribution, we investigated the effect of residue in position 65 on the photophysical properties of EGFP and EYFP, with an emphasis on photostability and oxidative redding. We compared bleaching and redding kinetics in EGFP, EYFP, and their mutants with reciprocally substituted chromophore residues, EGFP-T65G and EYFP-G65T. Measurements showed that T65G mutation significantly increases EGFP photostability and inhibits its excited-state oxidation efficiency. Remarkably, while EYFP-G65T demonstrated highly increased spectral sensitivity to chloride, it is also able to undergo redding in the absence of chloride.

To shed light on the origin of the observed differences in photophysical behavior of the two seemingly very similar chromophores, TYG (EGFP and EYFP-G65T) and GYG (EYFP and EGFP-T65G), we carried out atomistic simulations of the four model systems. The effect of the residue in position 65 can be explained by a simple kinetic model of the photocycle, which considers the competition between radiative and radiationless relaxation channels and photochemical bleaching. The atomistic simulations reveal that the main effect of the T65G mutation is the reduced excited-state lifetime of the GYG chromophore, resulting in its increased photostability. The effect of the residue in position 65 on the brightness and quantum yield is explained by an interplay between the radiative and radiationless relaxation channels. Directed simulation- and structure-guided tuning of a relative significance of the radiative/radiationless processes can be a basis for the development of the new fluorescent proteins with pre-determined photostability and fluorescence lifetime optimized for application in the next-generation imaging techniques.

4.6 Appendix A: Absorbance and fluorescence data normalization

Absorption and fluorescence excitation/emission spectra (fig. 4.1) were normalized to the maximum mean value. Photobleaching curves (fig. 4.2) were background-subtracted and normalized to the maximum values. EGFP and EGFP-T65G redding kinetics are shown as non-normalized and non-averaged background-subtracted curves (fig. 4.3A). In the case of EYFP and EYFP-G65T redding representation, we had to take into account the fact that fluorescence of the yellow (non-converted) spectral form, routinely detected in GFP channel, has a spectral crosstalk with the red (RFP) channel, thus complicating correct red form appearance registration. Namely, redding-specific signal was masked by the background yellow fluorescence.

To address this issue, we subtracted the background signal acquired in RFP channel considering its kinetics strictly proportional to that observed in GFP channel (according to the following equation RFPcorrected=RFPraw-(GFPraw*RFPzero/GFPzero), where raw is unprocessed value, zero initial value). The products of subtraction described above are represented either as maximum yellow signal-normalized and averaged curves (fig. 4.3B) or as non-normalized single measurements (fig. 4.3C).

4.7 Appendix B: Force field parameters for excited-state classical molecular dynamics simulations

To perform excited-state molecular dynamics, we reparameterized partial charges, key bond lengths, angles, and torsional angles, as well as the corresponding force constants. First, we optimized the structure of the isolated chromophore with B97x-D/aug-cc-pVDZ in its ground and the first excited state. We then computed the NBO charges with the same functional/basis set and used the differences between ground- and excited-state charges to calculate the partial charges in the excited state force-field by using the following equations:

$$q_{ex-charmm} = q_{gs-charmm} + \Delta q_{NBO(ex-gs)}, \tag{4.7}$$

where,

$$\Delta q_{NBO(ex-gs)} = q_{NBO(ex)} - q_{NBO(gs)}. \tag{4.8}$$

Bond lengths, angles, and dihedral angles were computed with the equations

$$E = k(b - b_0)^2, \tag{4.9}$$

$$E = k(A - A_0)^2. \tag{4.10}$$

Where B_0 is equilibrium bond length and A_0 is equilibrium bond angle.

$$E = k[1 + cos(n\phi - \delta)]. \tag{4.11}$$

where n is periodicity, δ is phase and ϕ is dihedral angle. The force constants were computed by tweaking the parameters from the equilibrium value in ground and excited states by constructing the PES. The second derivative of the parabolic fit gives the force constant. We took the ratio of the computed force constant and multiplied that with those in ground state force constants:

$$k_{ex-charmm} = \frac{k_{ex-computed}}{k_{gs-computed}} \times k_{gs-charmm}. \tag{4.12}$$

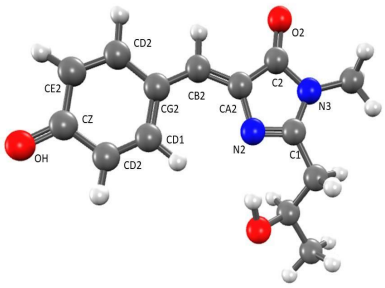


Figure 4.10: EGFP chromophore with atom types consistent with CHARMM 27 forcefield notations.

NOTE: Largest changes in partial charge happen on the methine bridge (marked by red).

The most important parameter was the torsional angle ϕ. The PES scans show that the chromophore is planar in the ground state and twisted in the excited state. We fitted the excited-state potential with a fitting potential with the calculated force constant, which enables the flip around ϕ. Partial charges and other force field parameters are listed at forcefield section. Fitting the potential for excited state PES (right) with respect to ϕ:

$$E = k[1 + cos(n\phi - 180)]; ground - state, n = 2, \tag{4.13}$$

Table 4.5: Partial charges in Charmm27, and in the ground and excited states of the EGFP chromophore (ωB97X-D/aug-cc-pVDZ). The last column shows adjusted partial charges used in excited-state molecular dynamics (see fig. 4.10).

Atom	Charmm27(gs)	NBO (gs)	NBO (ex)	$\Delta q_{NBO(ex-gs)}$	Charmm 27(ex)
C1	0.50	0.43	0.41	-0.02	0.48
N2	-0.60	-0.56	-0.64	-0.08	-0.68
N3	-0.57	-0.54	-0.54	0.00	-0.57
C2	0.57	0.66	0.61	-0.05	0.52
O2	-0.57	-0.75	-0.68	0.07	-0.50
CA2	0.10	-0.08	0.17	0.25	0.35
CB2	-0.14	-0.09	-0.47	-0.38	-0.52
HB2	0.21	0.25	0.25	0.00	0.21
CG2	-0.09	-0.21	0.06	0.27	0.18
CD1	-0.08	-0.17	-0.25	-0.08	-0.16
HD11	0.14	0.21	0.21	0.00	0.14
CD2	-0.08	-0.17	-0.25	-0.08	-0.16
HD21	0.14	0.24	0.24	0.00	0.14
CE1	-0.28	-0.31	-0.22	0.09	-0.19
HE11	0.10	0.21	0.21	0.00	0.10
CE2	-0.28	-0.31	-0.22	0.09	-0.19
HE21	0.10	0.23	0.23	0.00	0.10
CZ	0.45	0.46	0.39	-0.07	0.38
OH	-0.62	-0.72	-0.67	0.05	-0.57
CA3	-0.18	-0.41	-0.41	0.00	-0.18
HA31	0.09	0.26	0.26	0.00	0.09
HA32	0.09	0.20	0.20	0.00	0.09
C	0.51 (H)	0.21	(H)0.21	0.00	0.51
O	-0.51	—	—	–	-0.51
N	-0.47	(H)0.24	(H)0.24	-0.00	-0.47
HN	0.31	—	—	–	0.31
CA	0.07	-0.51	-0.51	0.00	0.07
HA	0.09	0.24	0.24	0.00	0.09
CB1	0.14	0.09	0.10	0.01	0.15
HB1	0.09	0.20	0.20	0.00	0.09
OG1	-0.66	-0.77	-0.82	-0.05	-0.71
HG1	0.43	0.50	0.52	0.02	0.41
CG1	-0.27	-0.65	-0.65	0.00	-0.27
HG11	0.09	0.21	0.21	0.00	0.09
HG12	0.09	0.20	0.20	0.00	0.09
HG13	0.09	0.23	0.23	0.00	0.09

Table 4.6: Bond lengths in Charmm27 forcefield and computed with ωB97X-D/aug-cc-pVDZ. The last column shows adjusted partial charges used in excited-state molecular dynamics.

Bond	$k_{gs,charmm}$	$b_{0(gs,charmm)}$	$\Delta b_{0(computed)}$	$b_{0(ex,charmm)}$	$k_{gs,computed}$	$k_{ex,computed}$	$k_{ex,charmm}$
CG2Q-CB2Q	437	1.410	0.035	1.445	564.301	327.936	253.96
CB2Q-CA2Q	500	1.390	0.000	1.390	623.96	430.6	345.05

Table 4.7: Bond angles in Charmm27 forcefield and computed with ωB97X-D/aug-cc-pVDZ. The last column shows adjusted partial charges used in excited-state molecular dynamics.

Bond	$k_{gs,charmm}$	$A_{0(gs,charmm)}$	$\Delta A_{0(computed)}$	$A_{0(ex,charmm)}$	$k_{gs,computed}$	$k_{ex,computed}$	$k_{ex,charmm}$
CG2Q-CB2Q-CA2Q	130.0	133.2	-5.50	127.7	201.96	195.058	123.1

$$E = k[1 + cos(n\phi - 180)]; excited - state, n = 4. \tag{4.14}$$

The major difference in the ground and excited state PES other than force constants is the change in periodicity (n) of the fitting potentials with much lower value of force constant for the torsional angle ϕ.

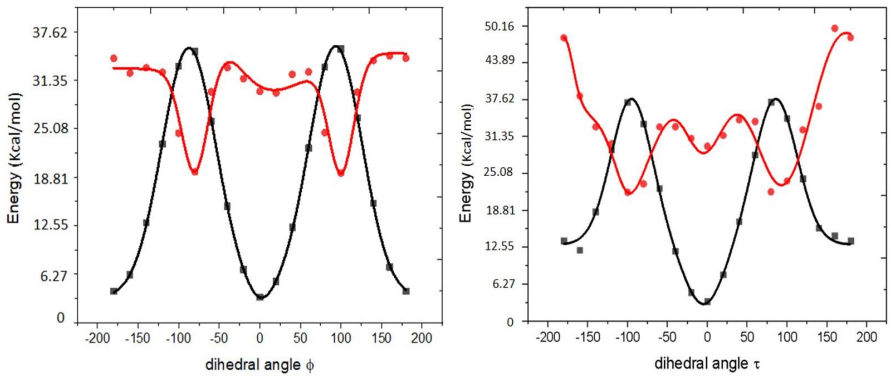


Figure 4.11: Ground- and excited-state torsional potentials for ϕ (twisting of the phenolic ring) and τ (twisting of the imidazolinone ring) of the bare HBDI chromophore. Black dots are *ab initio* calculations whereas red and black lines mark *ab initio* force-field. The barrier heights for twisting along ϕ and τ in the excited state are 3.5 kcal/mol and 3.2 kcal/mol, respectively. The respective ground-state barriers are 32.1 and 34.9 kcal/mol. Reproduced from Ref. 10.

Table 4.8: Parameterized force constant and periodicity (n) for torsional potentials for angles ϕ and τ.

Dihedral	$k_{gs,charmm}$	n (gs)	n (ex)	$k_{gs,qm}$	$k_{ex,qm}$	$k_{ex,charmm}$
ϕ	2.7	2	4	15.05	3.79	0.68
τ	3.9	2	4	14.99	4.90	1.27

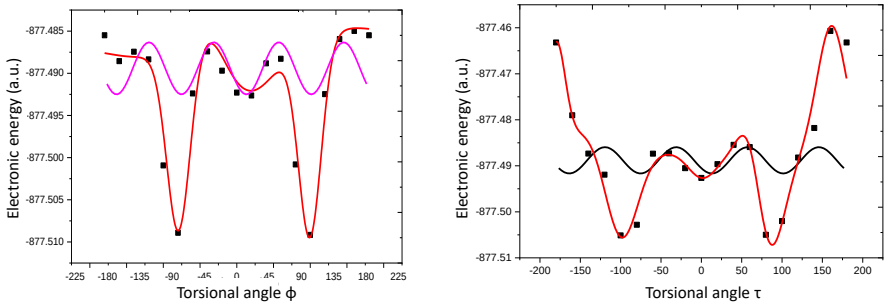


Figure 4.12: Excited-state torsional potentials for ϕ (left) and τ (right) of the bare HBDI chromophore. Red curves: fit to *ab initio* calculations (from which the parameters were extracted). Pink and black curves: torsional potentials computed with the modified force-field.

As one can see, our fit reproduces the barriers for twisting reasonably well, but does not reproduce the depth of the well of the twisted structures (the fitted potential is too shallow). Hence, to prevent the trajectories from re-crossing, in the excited-state MD simulations we simply stop the trajectories once they twist by more than a specified threshold value (50°).

4.8 Appendix C: Excitation energies

Table 4.9: Computed excitation energy (eV), oscillator strength, and transition dipole moment (TDM, a.u.) at the ground-state optimized geometry of the isolated TYG and GYG chromophores.

Method	TYG E_{ex}	TYG f_l	TYG TDM (a.u.)	GYG E_{ex}	GYG f_l	GYG TDM (a.u.)
ωB97X-D/aug-cc-pVDZ (TDA)	3.367 (S_0-S_1)	1.425	4.156	3.387 (S_0-S_1)	1.467	4.204
ωB97X-D/aug-cc-pVDZ (RPA)	3.101 (S_0-S_1)	1.016	3.657	3.123 (S_0-S_1)	1.052	3.708
ωB97X-D/aug-cc-pVTZ (TDA)	3.363 (S_0-S_1)	1.420	4.151	3.382 (S_0-S_1)	1.460	4.198
ωB97X-D/aug-cc-pVTZ (RPA)	3.097 (S_0-S_1)	1.014	3.657	3.119 (S_0-S_1)	1.052	3.710
EOM-CCSD/ aug-cc-pVDZ	2.947 (S_0-S_1)	1.153	3.785	2.965 (S_0-S_2)	1.188	3.839

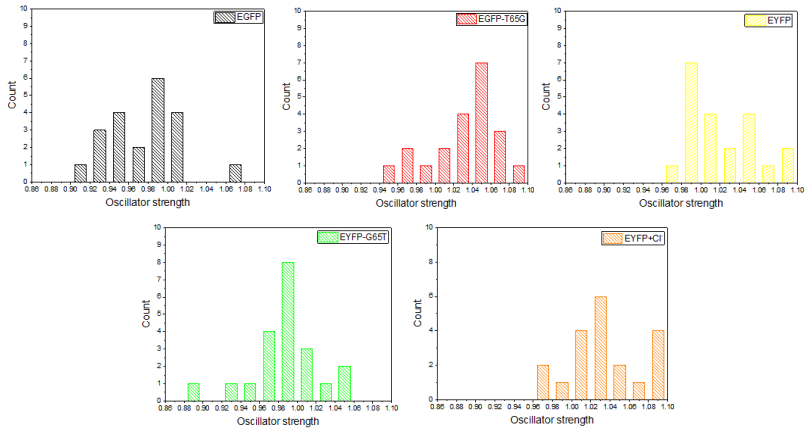


Figure 4.13: Distribution of oscillator strengths (ωB97x-D/aug-cc-pVDZ) computed for 21 QM/MM snapshots from the ground-state molecular dynamics.

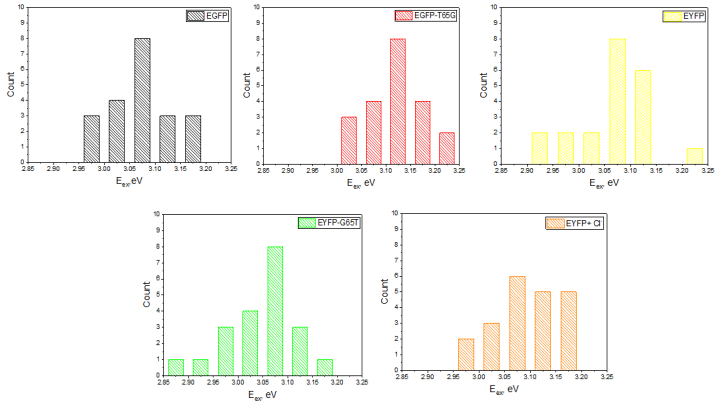


Figure 4.14: Distribution of excitation energ (ωB97x-D/aug-cc-pVDZ) computed for 21 QM/MM snapshots from the ground-state molecular dynamics.

4.9 Chapter 4 references

[1] B. P. Cormack, R. H. Valdivia, S. Falkow, FACS-optimized mutants of the green fluorescent protein (GFP), Gene **173**, 33 (1996).

[2] M. Ormö, A.B. Cubitt, K. Kallio, L.A. Gross, R.Y. Tsien, and S.J. Remington, Crystal structure of the aequorea victoria green fluorescent protein, Science **273**, 1392 (1996).

[3] D. P. Barondeau, C. J. Kassmann, J. A. Tainer, E. D. Getzoff, Understanding GFP post-translational chemistry: structures of designed variants that achieve backbone fragmentation, hydrolysis, and decarboxylation, J. Am. Chem. Soc. **128**, 4685 (2006).

[4] N. V. Pletneva, V. Z. Pletnev, K. A. Lukyanov, N. G. Gurskaya, E. A. Goryacheva, V. I. Martynov, A. Wlodawer, Z. Dauter, S. Pletnev, Structural evidence for a dehydrated intermediate in green fluorescent protein chromophore biosynthesis, J. Biol. Chem. **285**, 15978 (2010).

[5] D. M. Chudakov, M. V. Matz, S. Lukyanov, K. A. Lukyanov, Fluorescent proteins and their applications in imaging living cells and tissues, Phys. Rev. **90**, 1103 (2010).

[6] A. Acharya, A. M. Bogdanov, K. B. Bravaya, B. L. Grigorenko, A. V. Nemukhin, K. A. Lukyanov, and A. I. Krylov, Photoinduced chemistry in fluorescent proteins: Curse or blessing?, Chem. Rev. **117**, 758 (2017).

[7] R.Y. Tsien, The green fluorescent protein, Annu. Rev. Biochem. **67**, 509 (1998).

[8] S. Delagrave, R. E. Hawtin, C. M. Silva, M. M. Yang, D. C. Youvan, Red-shifted excitation mutants of the green fluorescent protein, Biotechnology **13**, 151(1995).

[9] A. Royant, M. Noirclerc-Savoye, Stabilizing role of glutamic acid 222 in the structure of Enhanced Green Fluorescent Protein, J. Str. Biol. **174**, 385 (2011).

[10] J. AJ. Arpino, P. J. Rizkallah, D. D. Jones, Crystal structure of enhanced green fluorescent protein to 1.35 Å resolution reveals alternative conformations for Glu222, PloS one **7**, e47132 (2012).

[11] D. P. Barondeau, C. J. Kassmann, J. A. Tainer, E. D. Getzoff, The case of the missing ring: radical cleavage of a carbon- carbon bond and implications for GFP chromophore biosynthesis, J. Am. Chem. Soc. **129**, 3118 (2007).

[12] R. M. Wachter, M-A. Elsliger, K. Kallio, G. T. Hanson, S. J. Remington, Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein, Structure **6**, 1267 (1998).

[13] R. M. Wachter, S. J. Remington, Sensitivity of the yellow variant of green fluorescent protein to halides and nitrate, Curr. Biol. **9**, R628 (1999).

[14] G-J. Kremers, J. Goedhart, D. J. van den Heuvel, H. C. Gerritsen, T. WJ. Theodorus, Improved green and blue fluorescent proteins for expression in bacteria and mammalian cells, Biochemistry **46**, 3775 (2007).

[15] A. V. Mamontova, I. D. Solovyev, A. P. Savitsky, A. M. Shakhov, K. A. Lukyanov, A. M. Bogdanov, Bright GFP with subnanosecond fluorescence lifetime, Scientific reports **8**, 1 (2018).

[16] A. M. Bogdanov, A. S. Mishin, I. V. Yampolsky, V. V. Belousov, D. M. Chudakov, F. V. Subach, V. V.Verkhusha,S. Lukyanov, K. A. Kukyanov, Green fluorescent proteins are light-induced electron donors, Nat. Chem. Biol. **5**, 459 (2009).

[17] A. M. Bogdanov, A. Acharya, A. V. Titelmayer, A. V. Mamontova, K. B. Bravaya, A. B. Kolomeisky, K. A. Lukyanov, A. I. Krylov, Turning on and off photoinduced electron transfer in fluorescent proteins by π-stacking, halide binding, and Tyr145 mutations, J. Am. Chem. Soc. **138**, 4807 (2016).

[18] K. S. Sarkisyan, A. S. Goryashchenko, P. V. Lidsky, D. A. Gorbachev, N. G. Bozhanova, A. Y. Gorokhovatsky, A. R. Pereverzeva, A. P. Ryumina, V. V. Zherdeva, A. P. Savitsky and others, Green fluorescent protein with anionic tryptophan-based chromophore and long fluorescence lifetime, Biophys. Jour. **109**, 380 (2015).

[19] P. Herman, A. Holoubek, B. Brodska, Lifetime-based photoconversion of EGFP as a tool for FLIM, Biochim. Biophys. Acta (BBA)-Gen. Sub. **1863**, 266 (2019).

[20] K. Suhling, J. Siegel, D. Phillips, P. MW. French, S. Leveque-Fort, S. ED. Webb, D. M. Davis, Imaging the environment of green fluorescent protein, Biophys. Jour. **83**, 3589 (2002).

[21] T. Nakabayashi, N. Ohta, Sensing of intracellular environments by fluorescence lifetime imaging of exogenous fluorophores, Anal. Sci. **31**, 275 (2015).

[22] D. Arosio, G. Garau, F. Ricci, L. Marchetti, R. Bizzarri, R. Nifosi, F. Beltram, Spectroscopic and structural study of proton and halide ion cooperative binding to gfp, Biophys. Jour. **93**, 232 (2007).

[23] S. Jayaraman, P. Haggie, R. M. Wachter, S. J. Remington, AS. Verkman, Mechanism and cellular applications of a green fluorescent protein-based halide sensor, J. Biol. Chem. **275** 6047 (2000).

[24] C. Cheng, G. Huang, H. Hsu, C. Prabhakar, Y. Lee, E. W. Diau, J. Yang, Effects of hydrogen bonding on internal conversion of GFP-like chromophores. II. The meta-amino systems, J. Phys. Chem. B. **117**, 2705 (2013).

[25] D. C. Bas, D. M. Rogers, J. H. jensen Very fast prediction and rationalization of pKa values for protein–ligand complexes, Proteins: Structure, Function, and Bioinformatics **73**, 765 (2008).

[26] M. H. M. Olsson, C. R. Sondergaard, M. Rostkowski, and J. H. Jensen, PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions, J. Chem. Theory Comput. **7**, 525 (2011).

[27] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, Nucleic acids research **32**, W665 (2004).

[28] T. R. Gosnell, Fundamentals of Spectroscopy and Laser Physics, Camb. Univ. Press 3 (2002).

[29] D. B. Hand, The refractivity of protein solutions, J. Biol. Chem. **108**, 703 (1935).

[30] T. L. MacMeekin, M. L. Merton, N. J. Hipp, Refractive indices of amino acids, proteins, and related substances, Advances in Chemistry **44**, 54 (1964).

[31] I. V. Polyakov, B. L. Grigorenko, E. M. Epifanovsky, A. I. Krylov, A. V. Nemukhin, Potential energy landscape of the electronic states of the GFP chromophore in different protonation forms: Electronic transition energies and conical intersections, J. Chem. Theory Comput. **6**, 2377 (2010).

[32] M. E. Martin, F. Negri, M. Olivucci, Origin, nature, and fate of the fluorescent state of the green fluorescent protein chromophore at the CASPT2//CASSCF resolution, J. Am. Chem. Soc. **126**, 5452 (2004).

[33] P. Altoe, F. Bernardi, M. Garavelli, G. Orlandi, F. Negri, Solvent effects on the vibrational activity and photodynamics of the green fluorescent protein chromophore: A quantum-chemical study, J. Am. Chem. Soc. **127**, 3952 (2005).

[34] C. Ko, B. Levine, A. Toniolo, L. Manohar, S. Olsen, H. Werner, T. J. Martinez, Ab initio excited-state dynamics of the photoactive yellow protein chromophore, J. Am. Chem. Soc. **125**, 12710 (2003).

[35] V. V. Belousov, A. F. Fradkov, K. A. Lukyanov, D. B. Staroverov, K. S. Shakhbazov, A. V. Terskikh, S. Lukyanov, Genetically encoded fluorescent indicator for intracellular hydrogen peroxide, Nat. Methods **3**, 281 (2006).

[36] D. S. Bilan, M. E. Matlashov, A. Y. Gorokhovatsky, C. Schultz, G. Enikolopov, V. V. Belousov, Genetically encoded fluorescent indicator for imaging NAD+/NADH ratio changes in different cellular compartments, Biochim. Biophys. Acta **1840**, 951 (2014).

[37] M. E. Matlashov, Y. A. Bogdanova, G. V. Ermakova, N. M. Mishina, Y. G. Ermakova, E. S. Nikitin, P. M. Balaban, S. Okabe, S. Lukyanov, G. Enikolopov and others, Fluorescent ratiometric pH indicator SypHer2: Applications in neuroscience and regenerative biology, Biochim. Biophys. Acta **1850**, 2318 (2015).

[38] N. C. Shaner, P. A. Steinbach, R. Y. Tsien, A guide to choosing fluorescent proteins, Nat. Methods **2**, 905 (2005).

[39] A. D. JR MacKerell, N. Banavali, N. Foloppe, Development and current status of the CHARMM force field for nucleic acids, Biopolymers: Original Research on Biomolecules **56**, 257 (2000).

[40] N. Reuter, H. Lin, and W. Thiel, Green fluorescent proteins: Empirical force field for the neutral and deprotonated forms of the chromophore. Molecular dynamics simulations of the wild type and S65T mutant, J. Phys. Chem. B **106**, 6310 (2002).

[41] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. **79**, 926 (1983).

[42] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. **26**, 1781 (2005).

[43] J.-D. Chai and M. Head-Gordon, Systematic optimization of long-range corrected hybrid density functionals, J. Chem. Phys. **128**, 084106 (2008).

[44] J.-D. Chai and M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom-atom dispersion interactions, Phys. Chem. Chem. Phys. **10**, 6615 (2008).

[45] Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A.T.B.; Wormit, M.; Kussmann, J.; Lange, A.W.; Behn, A.; Deng, J.; Feng, X., et al., Advances in molecular quantum chemistry contained in the Q-Chem 4 program package, Mol. Phys. **113**, 184 (2015).

Chapter 5: Interplay between Locally Excited and Charge Transfer States Governs the Photoswitching Mechanism in the Fluorescent Protein Dreiklang

5.1 Introduction

Many fluorescent proteins (FPs) undergo reversible photoswitching upon photoexcitation, which is instrumental in several imaging modalities[1], including super-resolution techniques[2,3]. The most common mechanism is cis-trans photoisomerization of the chromophore, sometimes coupled with changes in its protonation state; notable examples include Dronpa[4], Padron[5], and KFP[6]. However, an entirely different mechanism is operating in Dreiklang[7], where the switching is based on reversible photoinduced hydration/dehydration of the imidazolinone ring of the chromophore (Fig. 5.1).

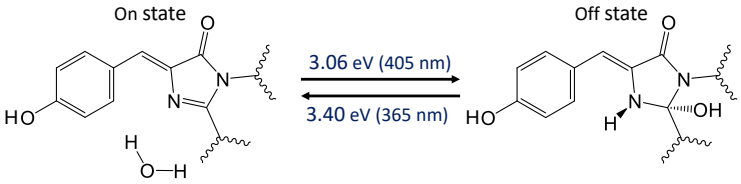


Figure 5.1: On-off photoconversion in Dreiklang is activated by photoexcitation of the neutral form of the chromophore in ON-state. The OFF-form can be turned on by photoexcitation at higher energy.

Dreiklang[7] was derived by random mutagenesis from Citrine, a close relative of EYFP[8]. It has the same chromophore, formed by the glycine-tyrosine-glycine (GYG) tripeptyde π-stacked

with a nearby tyrosine residue (Tyr203). The chromophore's conjugated core (Fig. 5.1) is the same as in EGFP[9], but due to the T65G mutation the connection to the peptide backbone via imidazolinone's carbon is slightly different.

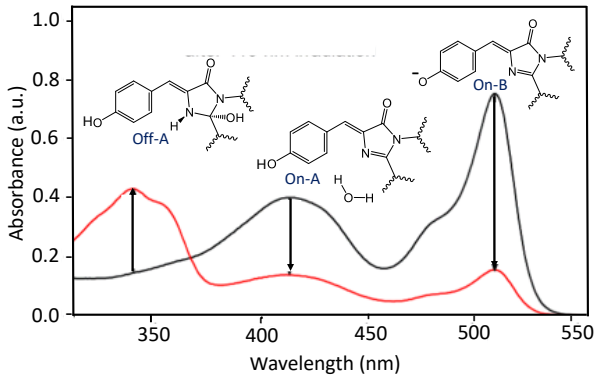


Figure 5.2: Steady-state absorption spectra of the ON-state (black) and following irradiation (red) at 3.02 eV (410 nm) at pH 7.5. The spectra are from Ref. 10. The band maxima are at 3.01 eV and 2.43 eV in the ON-state and at 3.65 eV in the OFF-state.

Fig. 5.2 shows absorption spectra of the ON- and OFF- forms of Dreiklang[7,10]. The absorption spectrum of the ON-form features two bands: one at $\sim$ 3.01 eV (411-413 nm) and a twice-more intense one at 2.43 eV (511 nm), with a shoulder at 2.58 eV (480 nm). These bands are assigned to the neutral and anionic forms of the chromophore, traditionally called form A and form B[11,12]. In many other GFP-like proteins[11], excitation of either band leads to the identical fluorescence spectra with the maximum around 2.3 eV (green or yellow), ascribed to form B. This is explained by ultrafast (picoseconds or shorter) excited-state proton-transfer (ESPT) from the chromophore to a proton acceptor via a proton wire[11,13,14]. In wt-GFP, ESPT proceeds as a sequential proton transfer from the excited neutral chromophore to the Glu222 carboxylate through a water molecule and the hydroxyl group of Ser205[12,15–18]. This pathway can be disrupted by mutations.

In Dreiklang, excitation of peak B leads to fluorescence at 2.34 eV (529 nm), with a quantum yield of 0.41. However, in contrast to many other FPs, the excitation of peak A leads to very

weak (albeit non-negligible[10]) fluorescence. This weak steady-state fluorescence is identical[10] to the fluorescence produced by excitation of peak B. These observations suggest that ESPT in Dreiklang is strongly suppressed and happens with a very small quantum yield. Reduced effectiveness of ESPT is consistent with the observation[7] that the essential difference between the parent system (Citrine) and Dreiklang is the upshift of the pKa of the ON-state of the chromophore (7.2 versus 5.7).

The distinguishing feature of Dreiklang is that irradiation of peak A results in photoconversion to the dark form (OFF-state)[7]. Thus, in imaging applications, the fluorescence of Dreiklang can be excited by using 2.43 eV light, recorded at 2.34 eV, and turned off by 3.01 eV light. This decoupling of the fluorescence excitation from photoswitching makes Dreiklang very attractive and is responsible for its name (Dreiklang is the German word for a three-note chord in music). The OFF-state spontaneously returns to the ON-state in the course of minutes. Alternatively, the dark state can be switched back by irradiation at 3.65 eV (340 nm).

The X-ray structures[7] of the ON- and OFF-states (PDB IDs: 3ST2/3ST4 and 3ST3, respectively) show that the ON-state is indeed similar to EGFP/EYFP, whereas the OFF-state has a hydrated chromophore, similar to an intermediate form of the immature chromophore[19] but with a methyne double bond.

The exact details of the photoswitching mechanism remain unknown. Espagne and coworkers[10] investigated the mechanism using transient absorption and concluded that formation of photoproducts occurs on a nanosecond time scale or slower. They reported spectroscopic evidence of the formation of excited-state (on a ps timescale) and ground-state (picosecond to nanosecond timescale) intermediates and proposed a tentative mechanism; however, the proposed structures of the intermediates have not been validated by theoretical modeling of their spectral properties. On the theoretical side, we investigated[20] thermal (ground-state) recovery of the ON-state. The calculations predicted a reaction barrier of about 27 kcal/mol on the

ground-state potential energy surface and identified Glu222 as the key residue involved in the recovery reaction, while the scan of the excited-state surface suggested a barrierless OFF→ON photoreaction. This work also presented a cursory analysis of the structures of the ON- and OFF-states, including tentative assignment of the protonation states of the key residues around the chromophore, and computed their spectral properties.

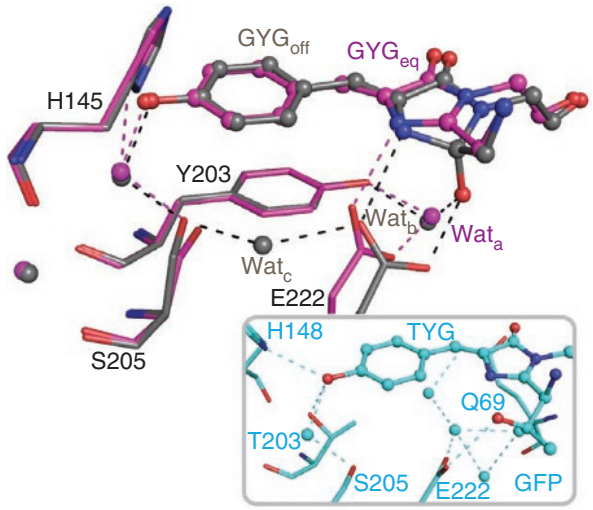


Figure 5.3: Superimposed representations of the hydrogen-bond network around Dreiklang's chromophore in the ON- and OFF-states. Color scheme: ON-state carbons, magenta; OFF-state carbons, gray; oxygen, red; nitrogen, blue. Important water molecules are shown as magenta (ON-state) and gray (OFF-state) spheres. Inset: hydrogen-bond network in EGFP. ***Reproduced from Ref. 7.***

Given the importance of proton wires in the photocycle of FPs, here we revisit the question of protonation states using more advanced computational protocols and assess the effect of different protonation states on the excited states of the chromophore. Dreiklang operates in a wide range of pH (6-9). Figure 6.3 shows superimposed crystal structures[7] of the ON-state (equilibrium structure, PDB ID 3ST2) and OFF-state (PDB ID 3ST3), indicating the hydrogen-bond network around the chromophore. It also compares the network around the Dreiklang chromophore with that in EGFP (PDB ID 1EMA)[9]. Dreiklang's structure clearly shows participation of His145, Glu222, and Ser205, as well as several water molecules. In contrast, in EGFP

the chromophore forms hydrogen bonds with His148 (position 145 is occupied by Tyr145 in EGFP), Thr203, and Glu222.

The two critical residues near the chromophore binding site in Dreiklang are Glu222 and His145. A tentative mechanism proposed in Ref. 10 assumed that both Glu222 and His145 are in the neutral form, at least, in the ON-state. In our study of the thermal recovery reaction[20], we considered Glu222 to be deprotonated and pointed out that the change in its protonation state along the reaction profile plays an essential role. Given the significant differences in the hydrogen-bond network in Dreiklang relative to EGFP, protonation states of the key residues should be carefully re-evaluated. We use the structure as the primary gauge and compare the distances between the selected residues and the chromophore. In some cases (e.g., for structures that have exactly the same atoms in the quantum part), we also consider total energies of the optimized structures.

After obtaining model structures for different forms of the chromophore and for different protonation states of His145 and Glu222, we compute excitation energies and analyze the effect of the protein environment. The key finding is that in the neutral form of the chromophore there is a low-lying state of charge-transfer (CT) character (Tyr203$\rightarrow$Chro), corresponding to electron transfer from Tyr203 to the chromophore. This state is only present in the neutral form and is located within 0.25 eV of the bright locally excited (LE) state of $\pi\pi^*$ character. We further investigate implications of the CT state by dynamical simulations and geometry optimizations. Our results indicate that population of the CT state plays the key role in Dreiklang's photoswitching. On the basis of our calculations, we propose a refined picture of the photoconversion mechanism, summarized in Fig. 5.4. As discussed below, this mechanism is consistent with all available experimental findings[7,10].

The structure of the paper is as follows. We begin by describing computational protocols. We then discuss the results of the simulations using different protocols. We first consider different protonation states and the excited states of the chromophore. We then discuss excited-state dynamics of the chromophore and the role of the CT state in the photoconversion. We conclude by discussing the implications of the revised photocycle.

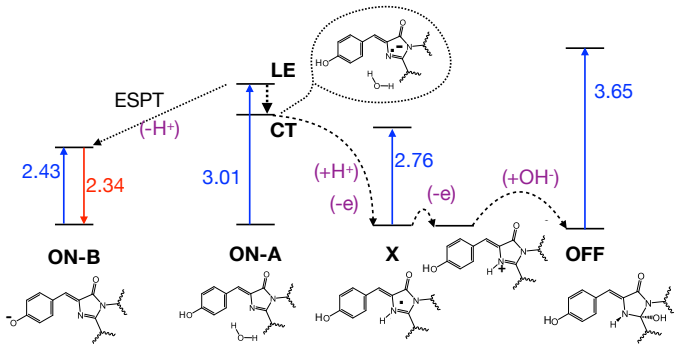


Figure 5.4: Revised Dreiklang's photocycle. Excitation of form A can lead to ESPT and fluorescence, but this channel is suppressed in Dreiklang. Alternatively, the locally excited chromophore can undergo a non-adiabatic transition to the CT state, which is then stabilized by proton transfer. After releasing the electron back to Tyr203, intermediate X undergoes nucleophilic attack by nearby water, forming the hydrated chromophore.

5.2 Computational methods and protocols

We begin with the crystal structure of the recovered ON-state (3ST4), which is nearly exactly superimposable on the equilibrium ON-structure (3ST2)[7]. The structure includes two water molecules: W354 near the phenolate end and W242 near the imidazolinone moiety. We note that in the previous study[20] we used 3ST2 as the starting point for the ON-state and the model structure also included an additional water molecule, which is present in 3ST3 structure (OFF-state) but not seen in 3ST2 and 3ST4.

We consider the following protonation states: Chromophore is anionic or neutral in the ON-state and is neutral in the OFF-state. Depending on the local environment, His145 can have 3

different protonation states: HSD (protonated at N_δ), HSE (protonated at N_ϵ), and HSP (protonated on both N, positively charged). Glu222 can be GLU (anionic) or GLUP (protonated). Figures 5.15 and 5.16 in the Appendix A summarize the names and definitions of different protonation states. *Propka*[21] suggested a neutral state (HSD or HSE) for His145 (pKa 2.2) and GLUP state (pKa 9.2) for the Glu222.

We built the model structures as follows. Starting from the PDB structure, hydrogen atoms were added using the *VMD* plugin and a modified (to include the chromophore) *CHARMM27* topology file. Protonation states were initially assigned by *Propka*[21] and then manually set for the chromophore, His145, and Glu222. Charged amino acids on the surface were locally neutralized by adding counterions close ($\sim$4.5 Å) to them. Charged residues that do not form salt bridges inside the protein barrel were also neutralized by adding appropriate counter ions at the surface. For HSD-GLUP structures, this protocol resulted in the addition of 19 Na^+ and 12 Cl^- in the neutral forms (ON- and OFF-states), and 19 Na^+ and 11 Cl^- in the anionic forms. For other protonation states the number of counter ions was adjusted accordingly. The proteins were solvated in water boxes producing a solvation layer of 15 Å. The TIP3P water model was used to describe water. Molecular dynamics (MD) simulations were performed using these solvated neutralized model structures as follows:

1. Minimization using steepest descent algorithm for 2000 steps (protein, crystal water, counterions).

2. Minimization using steepest descent algorithm for 2000 steps of the fully solvated structure (keeping protein frozen), with the subsequent equilibration of the solvent (keeping the protein frozen) for 500 ps with 1 fs time step using the NPT (isobaric-isothermal) ensemble.

3. Full equilibration of the system for 2 ns (with 1 fs time step) with periodic boundary condition (PBC) using the NPT ensemble (Noose-Hoover barostat with Langevin dynamics).

4. Production run for 2 ns with 1 fs time step using the NPT ensemble. Pressure and temperature were kept at 1 atm and 298 K.

The structures from production-run MD simulations were used to compute average structural parameters. We also used 21 snapshots from MD simulations to compute QM/MM (quantum mechanics/molecular mechanics) excitation energies; in these calculations, the geometry of the QM part was not optimized.

To obtain better structures for more accurate estimate of the excitation energies, we carried out QM/MM optimizations using a mechanical embedding scheme (ONIOM), starting from the final structures from Step 1. To reduce the system size, in these calculations we removed the counterions and pruned the solvation shell, only retaining waters within 4 Å from the surface of the protein. In these calculations, the size of the system was $\sim$5900 atoms and the charge was -7 (for the neutral ON form in HSE-GLUP state).

In the MD and QM/MM simulations we used *CHARMM27* parameters for standard protein residues[22] and the parameters derived by Reuter *et al.* for the anionic GFP chromophore[23]. The parameters for the hydrated form of the chromophore were derived from additional quantum mechanical calculations (optimized structures and natural bond orbital (NBO) charges[24]), as described in the SI. QM/MM optimizations were carried out using ONIOM. The definitions of the QM part used in ONIOM are shown in Fig. 5.5 (large QM). All coordinates were allowed to relax, except for the positions of link atoms (C_α carbons of the amino-acid residues shown in Fig. 5.5), which were pinned to the positions from the MM-relaxed structures.

The QM part was described by ωB97X-D/aug-cc-pVDZ in the QM/MM optimizations and in the AIMD (ab initio MD) simulations. This functional[25,26] belongs to the family of long-range corrected functionals in which the notorious self-interaction error is greatly reduced[27–30];

it also includes dispersion correction[31]. The benchmarks illustrated excellent performance of ωB97X-D for structures and energy differences of a broad range of compounds[25,26]. Using long-range corrected functionals is particularly important for charged systems and for describing CT states.

Excitation energies were computed using a finite cluster approach with slightly larger QM system (extended QM, see Fig. 5.5), which also included Ile64 and Leu68 directly connected to the chromophore. Excitation energies were computed using several electronic structure methods: TD-DFT with ωB97X-D, SOS-CIS(D)[32], EOM-CCSD[33], and XMCQDPT2[34]. In these calculations we used the following basis sets: cc-pVDZ, aug-cc-pVDZ on all atoms, and a mixed basis set, aug-cc-pVDZ on the heavy atoms of the chromophore and Tyr203 and cc-pVDZ on the rest of the atoms. The charge of the the large QM and extended QM is +1 for the neutral forms of the chromophore (due to the positively charged arginine) and zero for the anionic forms for all protonation states of the His148 and Glu222 except HSD-GLU, HSE-GLU, and HSP-GLUP (see Fig. 5.16 in the Appendix A for the definition of protonation states).

The XMCQDPT2 calculations were based on the CASSCF wave functions obtained by distributing 16 electrons over 12 orbitals and using density averaging. The active space included orbitals from the chromophore and Tyr203.

In addition, we computed excitation energies using electrostatic embedding, as in our previous studies[35–38]. To prevent the overpolarization of the QM part, the charges on the boundary atoms were redistributed as follows[37,38]: bonds before -CONH were cut and capped with hydrogen atoms and the charge on CONH was set to zero; the excess charge was then redistributed over other atoms of the residue to maintain the total charge of the amino acid. These calculations were performed using 21 snapshots from the MD trajectories (step 4 above) and the large QM system with the aug-cc-pVDZ basis set.

Fig. 5.5 shows the QM parts used in the ONIOM optimizations (large QM) and in the calculations of excitation energies (large QM and extended QM). We also carried out calculations with minimal QM (chromophore), and with the medium QM (chromophore and Tyr203).

Excited-state AIMD simulations were performed using the same protocol as the geometry optimization (ONIOM embedding, large QM, ωB97X-D/aug-cc-pVDZ, CHARMM27 forcefield), with constant energy (NVE) ensemble and using initial velocities corresponding to 298 K thermal distribution with 1 fs time step for 10 ps (10,000 steps).

All electronic structure calculations were carried out with Q-Chem[39,40], except for XMC-QDPT2 calculations, which were carried out with *Firefly*[41]. MD simulations were performed with NAMD[42].The excited-state analysis was carried out using the *libwfa* library[43]. In the Appendix D, we also present the results for the structures from Ref. 20, which were obtained with a different QM/MM protocol.

5.3 Results and discussion

5.3.1 Protonation states for the ON-state

It is instructive to begin by revisiting the hydrogen-bonding network around the EGFP chromophore. As clearly seen in Fig. 6.3, the EGFP network comprises Glu222, Ser205, Thr203, and His148. In EGFP, position 145 is occupied by tyrosine (not shown in the figure), which does not form a hydrogen bond with the chromophore. The protonation states of the key residues in EGFP (in the anionic form) are well established: Glu222 is protonated (neutral) and His148 is neutral (HSD form, protonated at N_δ)[12,37,38]. In the neutral form, Glu222 is deprotonated[12] and the protonation state of His148 is the same as in the anionic form. We note that alternative protonation states are thermodynamically accessible and can be populated, especially at different pH. A recent study reported a subatomic resolution X-ray structure of GFP in the neutral (T203I mutant) and anionic (S65T and E222Q mutants) forms[44]. For the neutral form, hydrogen atom

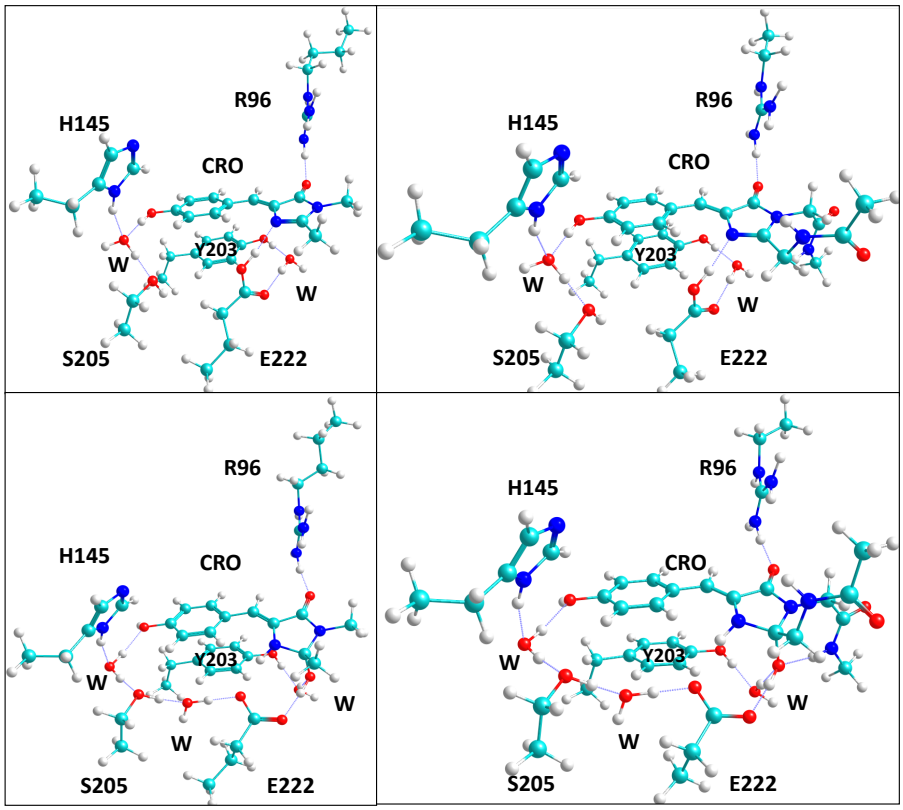


Figure 5.5: Defition of the QM subsystem. The residues numbering corresponds to the crystal structures (3ST3 and 3ST4). Left: Residues included in large QM in the QM/MM calculations of ON- (top) and OFF-states (bottom). Right: Residues included in extended QM in the excited-state calculations of ON- (top) and OFF-states (bottom). Small QM contains only the chromophore and medium QM contains the chromophore and Tyr203. The total charge of the small and medium QM is zero for the A-form (neutral chromophore) and -1 for the B-form (anionic chromophore). For large and extended QM, the total charge of the QM is +1 for the on-A (HSD-GLUP, HSE-GLUP, HSP-GLU), 0 for the on-A (HSD-GLU, HSE-GLU), 0 for the on-B (HSD-GLUP, HSE-GLUP, HSP-GLU), -1 for the on-B (HSD-GLU, HSE-GLU), +1 for the off-A (HSD-GLUP, HSE-GLUP, HSP-GLU), 0 for the off-B (HSD-GLU, HSE-GLU). See Fig. 5.16 in the SI for the definition of the protonation states. For on-A (HSE-GLUP) structure, large and extended QM comprised 113 and 118 atoms, respectively.

densities show that the chromophore is in the neutral form, His148 is in HSD form, and Glu222 is in anionic form, which is consistent with our choices of protonation states in neutral GFP. For the anionic form, the maps confirm that Glu222 is in neutral form (in agreement with the

proton wire picture), but His148 is positively charged (HSP)—this suggests that in the ground state there is an additional proton involved in protonation equilibrium.

In Dreiklang, Thr203 is replaced by tyrosine, which participates in π-stacking instead of hydrogen bonding. This difference has a major effect on the distance between Glu222 and Ser205: compare 4.18 Å in Dreiklang and 3.72 Å in EGFP. Another important difference is that in Dreiklang position 145 is occupied by histidine, which coordinates the water molecule that forms a hydrogen bond with the phenolic oxygen atom of the chromophore. In EGFP, position 145 is occupied by tyrosine (which is not involved in the hydrogen bonding network around the chromophore) and His148 is much closer to the chromophore than in Dreiklang, forming a hydrogen bond. Furthermore, T65G substitution, which, as was shown recently[38], significantly weakens the hydrogen-bonding network around the chromophore, increasing its flexibility in the excited state.

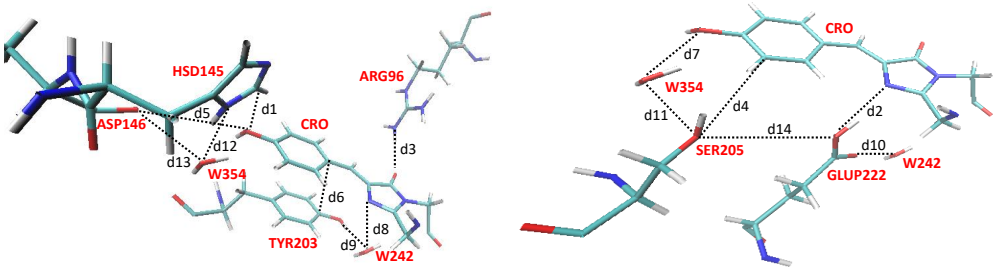


Figure 5.6: Definitions of selected distances used to compare various structures for the ON-form: d1 = CRO:OH-HIS145:CE1, d2 = CRO:N2-GLU222:OE1, d3 = CRO:O2-ARG96:NH2, d4 = CRO:CE2-SER205:OG, d5 = CRO:OH-ASP146:O, d6 = CRO66:CG2-TYR203:CZ, d7 = CRO66:OH-TIP354:OH2, d8= CRO:N2-TIP242:O,d9= TYR203:OH-TIP242:OH2, d10= GLU222:OE2-TIP242:OH2, d11 = SER205:OG-TIP354:OH2, d12 = HIS145:ND1-TIP354:OH2, d13 = ASP146:O-TIP354:OH2, d14= GLU222:OE1-SER205:OG.

Figures 5.6 and 5.7 show the definition of the key distances used to validate the structures of the ON- and OFF-states. Tables 5.8-5.13 in the Appendix D contain the average values computed along the MD trajectories, the values at the QM/MM optimized structures, and compare

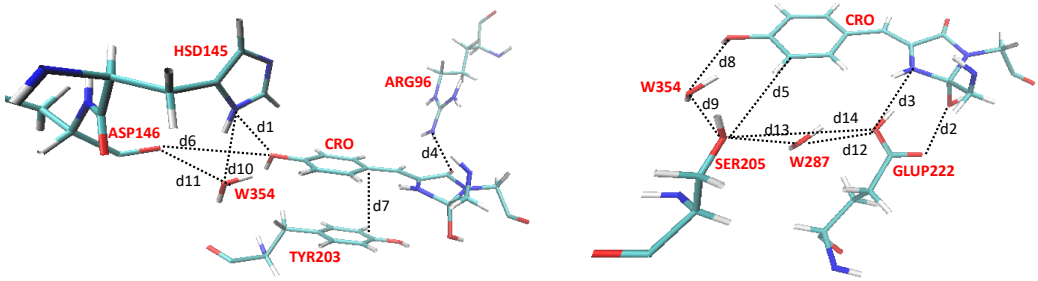


Figure 5.7: Definitions of selected distances used to compare various structures for the OFF-form: d1 = CRO:OH-HIS145:CE1, d2 = CRO:O1-GLU222:OE1, d3 = CRO:N2-GLU222:OE2, d4 = CRO:O2-ARG96:NH2, d5 = CRO:CE2-SER205:OG, d6 = CRO:OH-ASP146:O, d7 = CRO66:CG2-TYR203:CZ, d8 = CRO66:OH-TIP245:OH2, d9 = SER205:OG-TIP245:OH2, d10 = HIS145:ND1-TIP245:OH2, d11 = ASP146:O-TIP245:OH2, d12= GLU222:OE2-TIP287:OH2, d13= SER205:OG-TIP287:OH2, d14= GLU222:OE1-SER205:OG.

them with the respective values from the crystal structures. These values are presented graphically in Figs. 5.8 and 5.9. Fig. 5.20 in the Appendix D shows relative energies of the optimized structures for the model systems where the QM parts contains the same set of atoms, such that the total energies are comparable.

We note that the comparison with crystal structure is complicated by the equilibrium between the anionic and neutral chromophores. The averaged distances from the MD simulations generally agree well with the values from QM/MM optimization, which provides validation of the force-field parameters; the largest differences are observed for d7 and d11 (water position). For the ON-state with the neutral chromophore, we observe the best agreement (as judged from the smallest standard deviations of the QM/MM optimized structures from the X-ray structure) for the HSE-GLUP state (this is in agreement with Ref. 20). The largest variations between different protonation states are observed for d2 (Glu222-imidozalinone) and d14 (Ser205-Glu222). For the latter, the crystal structure value is 4.18 Å and the HSE-GLUP value is 4.72 Å, whereas other protonation states yield shorter distances — e.g., in the structures with GLU d14≈2.5 Å. In Ref. 20, d18=4.87 Å for HSE-GLUP (neutral ON-state), which is close

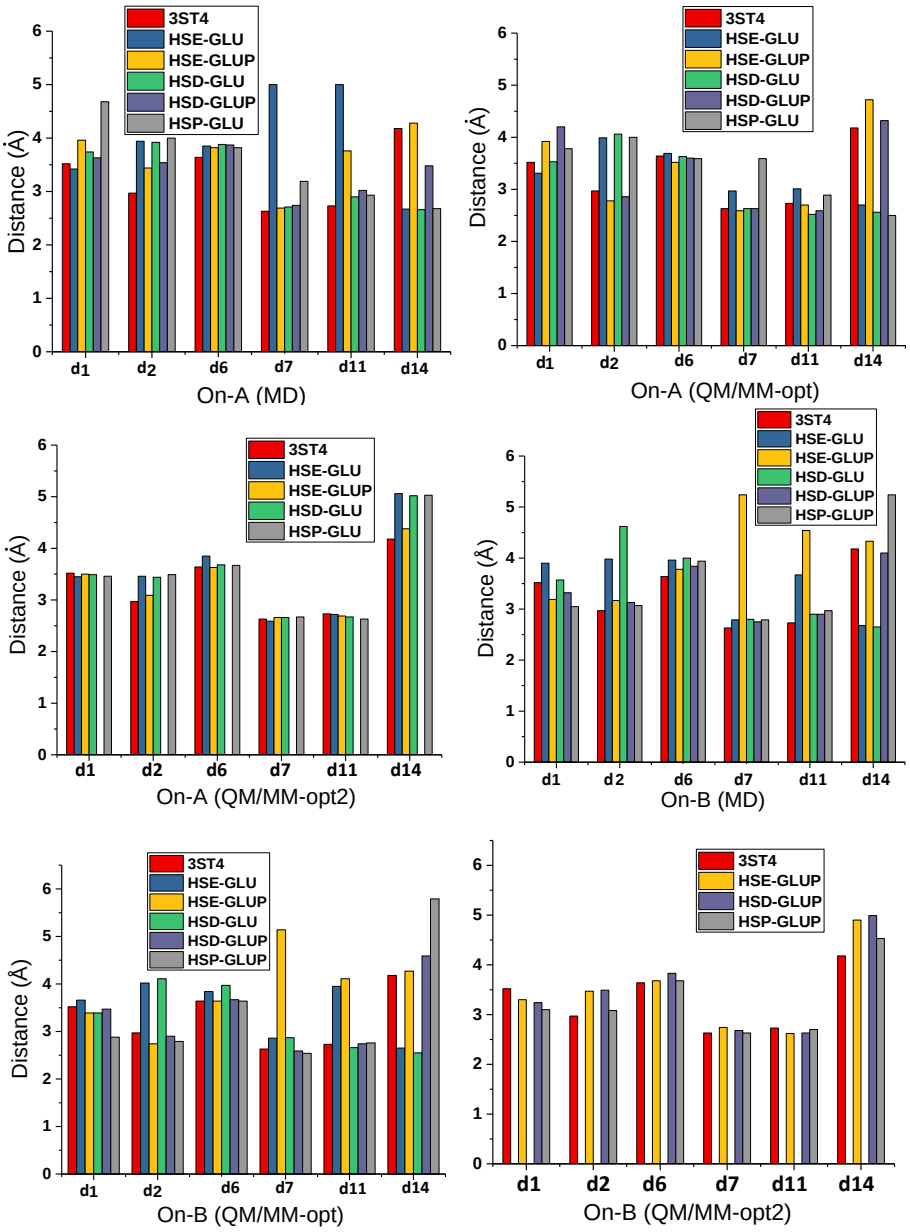


Figure 5.8: Key distances for ON-states: Comparison between crystal structure, average MD values, and QM/MM optimization. See Fig. 5.6 for definitions.

to the present value. Further comparison between the present model structures and those from Ref. 20 is given in the Appendix D (Tables 5.8 and 5.9 and Figure 5.18).

In terms of the total electronic energies, HSE-GLUP is 0.33 eV below HSD-GLUP, which is 0.42 eV lower than HSP-GLU; this energetics are consistent with the HSE-GLUP state being

the most favorable for the ON-state with the neutral chromophore. The gap between HSD-GLU and HSE-GLUP is 1.05 eV.

For the anionic chromophore, we observe the best agreement for HSD-GLUP. Here again d2 and d14 show the largest variations between different protonation states. The HSE-GLUP state is also a viable candidate. In contrast, in Ref. 20 HSP-GLUP was used to describe the anionic ON-state. In terms of the structures, the largest difference between HSD-GLUP and HSP-GLUP is in d14: compare 4.18 Å (X-ray) with 4.59 Å (HSD-GLUP) and 5.79 Å (HSP-GLUP). For HSE-GLUP, the largest differences are observed for d11 (Wat-Asp146) and for Chro-Tyr203: compare 5.14 Å (in HSE-GLUP) versus 2.59 (HSD-GLUP). Here again, MD simulations and QM/MM optimizations are in qualitative agreement.

In terms of the total electronic energies, HSD-GLUP is only 0.35 eV below HSE-GLUP. HSD-GLU and HSE-GLU are nearly isoenergetic (the latter is 0.1 eV lower). Hence, on the basis of structures and energetics, HSD-GLUP appears to be the best match, but other states cannot be ruled out.

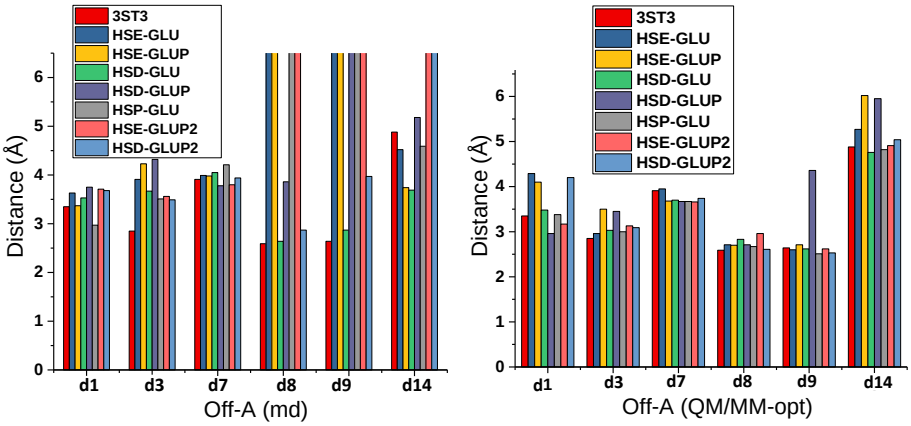


Figure 5.9: Key distances for OFF-states: Comparison between crystal structure, average MD values, and QM/MM optimizations. See Fig. 5.7 for definitions. Note that some MD values for d8, d9, and d14 are off the chart.

Fig. 5.9 shows the key distances for the OFF-state. Here the differences between the MD values and QM/MM optimizations are much larger (some values are off the chart), highlighting

the advantage of using rigorous QM potentials. For the OFF-state, we observe the best agreement in terms of structures for the HSE-GLUP2, HSD-GLU, and HSP-GLU, but the differences are not that large. Comparisons of the total energies favor HSD-GLUP-OE2 (among HSD-GLUP-OE2, HSP-GLU, HSE-GLUP-OE2, HSE-GLUP, and HSD-GLUP series) and HSD-GLU (relative to HSE-GLU); HSP-GLU is slightly more stable compared to HSE-GLUP-OE2 (0.09 eV). In Ref. 20, HSE-GLUP and HSP-GLU were chosen as the best candidates.

Thus, we conclude that HSE-GLUP is the most likely protonation state in the neutral ON state. For the anionic form and for the OFF state, several choices appear to be possible. In the next section, we discuss the effect of the different protonation states on the excited states of the chromophore.

5.3.2 Excited-state analysis

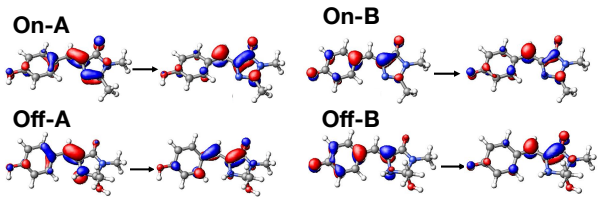


Figure 5.10: NTOs for the lowest bright states of the bare chromophores. Top left: neutral ON-state; top right: anionic ON-state; bottom left: neutral OFF-state; bottom right: anionic OFF-state. ωB97X-D/aug-cc-pVDZ.

We begin by analyzing excited states of the isolated chromophores computed at their equilibrium geometries (see the SI). Table 5.1 shows computed excitation energies and oscillator strengths of the isolated chromophores in ON- and OFF-states and Fig. 5.10 shows the respective natural transition orbitals (NTOs)[43,45]. The excited state of a GFP-like chromophore corresponds to the $\pi \rightarrow \pi^*$ transition, with the main action happening on the methyne bridge[46]. Consistently with previous studies[47,48], we observe that lower-level methods (TD-DFT) overestimate the excitation energies. EOM-CCSD energies are 0.06-0.28 eV below the TD-DFT ones.

Table 5.1: Excitation energies (eV) of the isolated chromophores (ON- and OFF-states, A and B forms) computed at the optimized geometries (ωB97X-D/aug-cc-pVDZ). Oscillator strengths are shown in parenthesis[a].

System	TDDFT aug-cc-pVDZ	SOS-CIS(D) aug-cc-pVDZ	EOM-CCSD aug-cc-pVDZ	XMCQDPT2 cc-pVDZ	XMCQDPT2 aug-cc-pVDZ-mod[b]
on-A	3.75 (0.72) 331 nm	3.88 (1.04) 320 nm	3.69 (0.98) 336 nm	3.54 (0.50) 350 nm	3.26 (0.49) 380 nm
on-B	3.10 (1.00) 400 nm	2.75 (1.07) 451 nm	2.95 (1.14) 420 nm	2.58 (1.11) 481 nm	2.40 (1.02) 517
off-A	4.29 (0.60) 289 nm	4.62 (0.71) 268 nm	4.01 (0.88) 309 nm	4.47 (0.68)[c] 277 nm	4.04 (0.42) 307 nm
off-B	3.39 (0.91) 366 nm	3.07 (1.07)[d] 404 nm	3.23 (1.11) 384 nm	3.02 (0.90) 411 nm	2.82 (0.78) 440 nm

[a] The lowest excited state is the bright state in all cases except when marked otherwise.
[b] aug-cc-pVDZ on heavy atoms and cc-pVDZ on hydrogens.
[c] The lowest bright state corresponds to the S_0-S_2 transition.
[d] The lowest bright state corresponds to the S_0-S_2 transition.

XMCQDPT2 energies are below the EOM-CCSD energies by 0.4-0.5 eV for on-A and on-B, but are nearly he same for off-A. We note a generally good agreement between SOS-CIS(D) and XMCQDPT2 for all four cases: the differences are less than 0.4 eV and XMCQDPT2 values are below SOS-CIS(D). Importantly, all methods capture (qualitatively) the large red shift ($\sim$0.6 eV) between the neutral and anionic chromophores (we note that SOS-CIS(D) overestimates the shift by almost a factor of 2). The shift can be explained in the framework of the Hückel model[46] and is due to the increased delocalization on the methyne bridge in the anionic form. The oscillator strength for the anionic form is higher than that for the neutral, but the values depend on the method, i.e., the ratio is 1.2 for EOM-CCSD, 1.4 for TD-DFT, and 2.2 for XMCQDPT2.

As expected, the excitation energies in the hydrated chromophore (OFF-state) are blue-shifted relative to the ON-state by roughly 0.6 eV due to disrupted conjugation. Here again all methods are in qualitative agreement, although SOS-CIS(D) yields much higher values than TD-DFT, EOM-CCSD, and XMCQDPT2.

As the next step, we consider the effect of the environment on excitation energies. The protein environment is important for quantitative comparison of the theoretical values with experiments[12,35–37]. Here we primarily rely on a finite cluster approach and compute excitation energies using the extended QM system (Fig. 5.5). To assess the effect of the protein beyond the extended QM, we also include the results of the QM/MM calculations using electrostatic embedding computed for 21 snapshots taken from the MD simulations.

The protein environment affects excitation energies through the electrostatic interactions that are sensitive to different charge distributions in the ground and excited states. In addition, protein environment may change the characters of the excited states and even lead to the emergence of new types of states. Orbital analysis[43,49,50] of the transitions provides a clear picture of such qualitative changes. Fig. 5.11 shows NTOs for the two lowest excited states of of the on-A form (HSE-GLUP protonation state; TD-DFT/aug-cc-pVDZ/extended QM). As one can clearly see, the character of the second excited state is the same as in the bare chromophore (Fig. 5.10)—both the hole and particle NTOs are localized on the chromophore and their shapes are not affected by the protein. This state also has a large oscillator strength, consistent with the $\pi\pi^*$ character of the transition. In contrast, the lower state (which only appears in the protein environment) shows a clear charge-transfer character—the hole NTO is a π-type orbital residing on Tyr203 and the particle orbital is the π^* orbital of the chromophore; this state has much lower (but non-negligible) oscillator strength. Below we refer to these two states as locally excited (LE) and charge-transfer (CT) states. The NTOs for all protonation forms of the neutral and anionic forms are shown in the Appendix E (Figures 5.21 and 5.22). As one can see, the two lowest states in the on-A form correspond to the CT state (Tyr203$\rightarrow$Chro) and bright LE state ($\pi\pi^*$) in all protonation states of His145 and Glu222.

In the on-B form, the lowest excited state has the same $\pi\pi^*$ character as in the bare chromophore. In QM-only calculations (extended QM), the second excited state is of CT character (0.3-0.8 eV above), but this state disappears when the rest of the protein is included.

The protonation states of His145 and Glu222 affect the excitation energies, but not the characters of the states. Importantly, the low-lying CT state appears in all protonation forms of on-A.

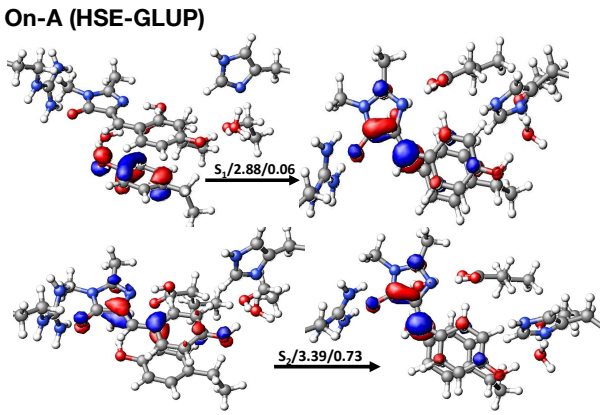


Figure 5.11: NTOs for the two lowest excited states of the protein-bond chromophore (on-A form, HSE-GLUP protonation state). QM/MM/ωB97X-D/aug-cc-pVDZ.

Table 5.17 in the Appendix E shows TD-DFT excitation energies computed for large and extended QM with different basis sets for the on-A form. As one can see by comparing the extended QM with the bare chromophore, the protein environment leads to a red shift of the excitation energy of the LE state by 0.2-0.4 eV. We observe the lowest excitation energy in HSE-GLUP (the most likely protonation state) and the highest in HSP-GLU. The differences between large and extended QM are less than 0.1 eV. The effect of the basis set is small — for all forms, changing the basis from the cc-pVDZ to a mixed basis (aug-cc-pVDZ on the chromophore and Tyr203 and cc-pVDZ on the rest) and to the full aug-cc-pVDZ basis leads to small red shifts for all protonation states; the largest magnitude was 0.06 eV. To estimate the effect of the rest of the protein (beyond extended QM), we compare the results of the QM and QM/MM calculations

using MD snapshots (Table 5.14 in the Appendix E): as one can see, including the rest of the protein leads to a small blue shift of about 0.1 eV for the LE state. The results for the CT state show somewhat stronger dependence on the computational protocol. At the TD-DFT level, the CT state appears 0.3-0.5 eV below the LE state in finite-cluster calculations. Increasing the basis set can blue-shift its energy by up to 0.03 eV. Interestingly, including the effect of the rest of protein (Table 5.14) leads to a larger blue-shift of the CT state than for the LE state ($\sim$0.3 versus $\sim$0.1). The results suggest that the position of the CT state in the QM-only calculations is slightly underestimated. We attribute this effect to the overstabilization of the CT state by the positively charged arginine in finite-cluster calculations; including the rest of the protein and the counterions leads to the partial screening of the arginine field and, therefore, increases the energy of the CT state. We also note that the position of the CT state is sensitive to the counterions and varies among the snapshots; this is similar to the observations reported in Refs. 35 and 51.

Importantly, even including this additional correction, the CT state appears below the LE state at TD-DFT level in the neutral chromophore in all protonation states of His145 and Glu222. To further refine the positions of the LE and CT states, we computed excitation energies using XMCQDPT2; these results are collected in Table 5.20. Similar to the isolated chromophore, the XMCQDPT2 excitation energies of the $\pi\pi^*$ state are red-shifted relative to TD-DFT. The inclusion of the protein environment has the same effect as in TD-DFT — overall red shift relative to the isolated chromophore. In the finite-cluster calculations, the XMCQDPT2 excitation energies of the LE state appear to be red-shifted relative to the experiment by 0.4 eV in HSD-GLUP and by 0.2 in HSE-GLUP protonation states; including the effect of the rest of the protein is expected to reduce this discrepancy. Importantly, XMCQDPT2 calculations confirm the presence of the CT state. At this level of theory, the gap between the LE and CT states is smaller than at the TD-DFT level, which is consistent with the tendency of TD-DFT

to overestimate the positions of valence excited states and to underestimate the position of CT states. In the HSE-GLUP form, our best candidate for the neutral ON-form, the CT state is 0.3 eV below LE state at the XMCQDPT2 level in finite-cluster calculations. Extrapolating to the full protein, we expect this gap to shrink to about 0.15 eV.

These comparisons provide a measure of the uncertainty of the calculations due to the basis set, QM size, and the correlation treatment; they also quantify the variations due to different protonation states. Importantly, although we cannot pinpoint the exact location of the CT state, our results indicate that it is energetically close to the LE state. Taking into account the variations in energies due to different protonation states and uncertainties of computational protocols, we estimate that the CT state is within 0.25 eV of the LE state in the neutral ON-state. We also observe that its position is very sensitive to the hydrogen bond pattern and positions of counterions. Hence, its energy can fluctuate in the course of thermal motions, bringing it in resonance with the LE state. Hence, the CT state can be accessed either via direct excitation (since it has non-zero oscillator strength) or via non-adiabatic transition from the LE state. Since CT states are known to be involved in bleaching and some photochonversions[37,51–53], the appearance of this state in Dreiklang is highly suggestive of its role in photoconversion. Below we further investigate this question.

Tables 5.18, 5.15, and 5.21 show the results for the anionic chromophore (on-B form). In this case, all methods (TD-DFT, SOS-CIS(D), and XMCQDPT2, both finite-cluster and QM/MM calculations) agree that the lowest state is LE of $\pi\pi^*$ character. In finite-cluster calculations, TD-DFT shows the CT state about 0.3-0.6 eV above the LE state, but when the rest of the protein is included, this state disappears. The effect of the protein leads to a small shift of the LE state (-0.2/+0.02 eV). The effect of the protein environment beyond the extended QM is of similar magnitude as for the LE state in ON-state (-0.03/+0.1 eV). The differences between

the cc-pVDZ and aug-cc-pVDZ bases do not exceed 0.1 eV. Better treatment of electron correlation leads to substantial red-shift, up to 0.6 eV. Comparing to the experimental value (2.43 eV), the XMCQDPT2 values are within 0.1-0.2 eV, depending on the protonation state. At the XMCQDPT2 level, the best agreement is observed for HSE-GLUP and HSD-GLU structures.

The results for the OFF-state (shown in Tables 5.19, 5.16, and **??** in the SI) reveal similar trends. Regardless of the protonation state, there are no low-lying CT states. In this case, the protein environment leads to larger red shifts of 0.4-0.8 eV, depending on the protonation state. As for the LE states in the ON-states, the effect of the protein beyond the extended QM is small (0.01-0.2 eV). At the XMCQDPT2 level, the best agreement with experiment is observed for HSE-GLUP and HSE-GLUP2 structures (and the largest deviation—for HSE-GLU and HSD-GLUP2).

We note that the results of the SOS-CIS(D) calculations show rather non-systematic behavior. Whereas TD-DFT systematically overestimates excitation energies of the LE state relatively to XMCQDPT2 in all forms and protonation states, the SOS-CIS(D) results are in between TD-DFT and XMCQDPT2 for the neutral and anionic forms of the ON-state, but not in the OFF-state, where they are above TDDFT for some protonation states. Likewise, SOS-CIS(D) results for the CT state show large discrepancy relative to the XMCQDPT2, which can be traced to the systematic overestimation of the CT states by the CIS method. These type of errors are expected for a low-level method relying on perturbative account of the correlation on top of the CIS wave functions.

To graphically summarize these results, we show the computed excitation energies (with extrapolation correction) versus the experimental band maxima in Fig. 5.12 (raw QM/MM energies are plotted in Fig. 5.23). Whereas the absorption bands corresponding to the LE states are unambiguous, the position of the CT state is not known. In Fig. 5.12, we show the CT excitation energy against the shoulder of the main peak (2.58 eV, see Fig. 5.2), in order to see if

there is a correlation between the computed position of the CT state and the shoulder that might suggest that the shoulder is due the absorption to the CT state.

The extrapolated excitation energies for our best candidates (selected on the basis of the structural analysis) are as follows: on-A/HSE-GLUP — TD-DFT is 3.46 eV and XMCQDPT2 is 2.93 eV, to be compared with the 3.01 experimental value; on-B/HSD-GLUP — TD-DFT is 2.88 eV and XMCQDPT2 is 2.09 eV, to be compared with 2.43 eV experimental value; on-B/HSE-GLUP — TD-DFT is 3.01 eV and XMCQDPT2 is 2.36 eV, to be compared with 2.43 eV experimental value; and for the off-A/HSE-GLUP2 form — 3.90 eV/3.62 eV; HSD-GLU: 3.78/3.44 eV, HSP-GLU: 3.89/3.50 eV; all these numbers are reasonably close to the experimental value of 3.65 eV.

Based on these results, the excitation energies in different protonation states are close and cannot be used to confidently rule out some protonation states (in contrast to other cases[36]). Moreover, given the small energy differences between the respective optimized structures, different protonation states can be populated simultaneously. Overall, the extrapolated XMCQDPT2 results suggest that the least likely protonantion states are HSD-GLUP for the ON-state (both neutral and anionic) and HSE-GLU and HSD-GLUP2 for the OFF-state.

The results also suggest that the shoulder at 2.58 eV in the absorption spectrum of the ON-state (Fig. 5.2) may be due to either the presence of another major protonation state or the CT state of the neutral chromophore; a vibronic nature of the shoulder cannot be ruled out. One way to experimentally pinpoint the location of the CT state and to assess whether the shoulder is due to the CT state would be to measure the dependence of the quantum yield of the photoconversion as a function of the excitation wavelength. One of the implications of the revised mechanism is that direct excitation of the CT state would lead to increased yield of the OFF-form.

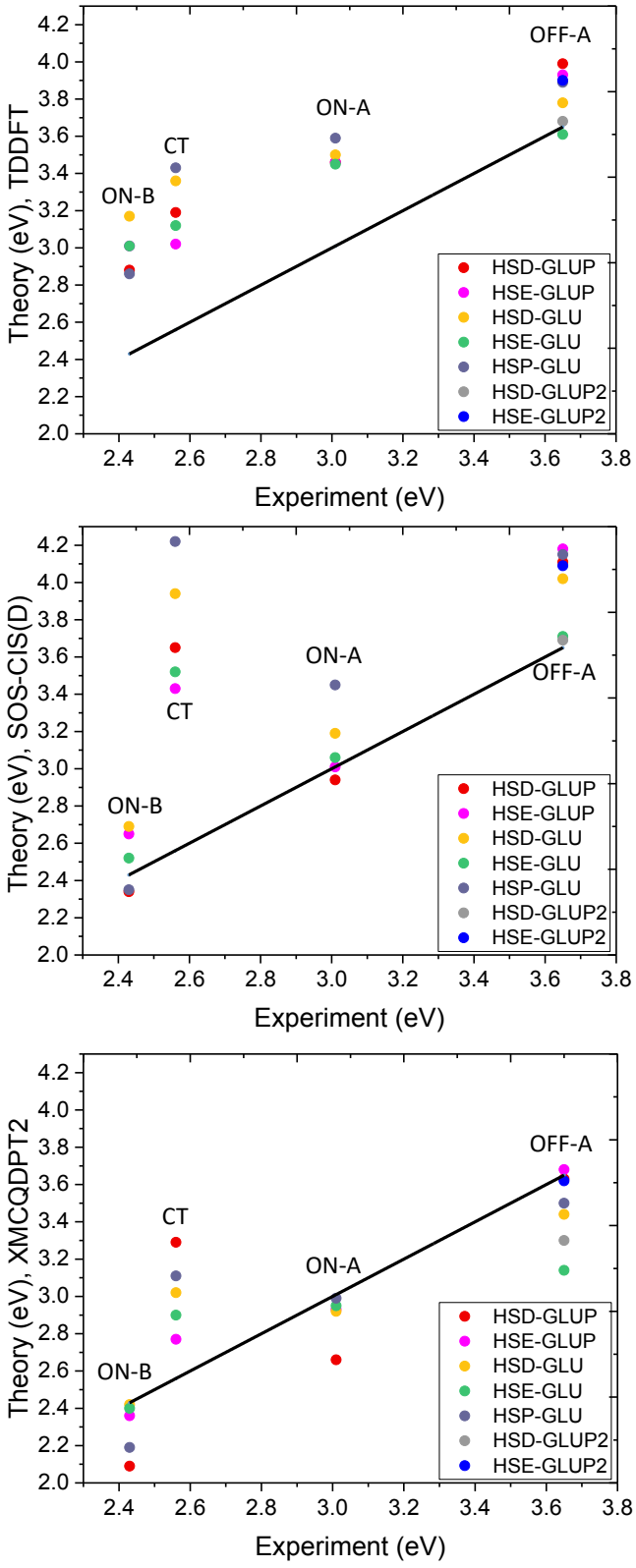


Figure 5.12: Excitation energies for different model systems shown against the experimental values. Top: TD-DFT/aug-cc-pVDZ; middle: SOS-CIS(D)/aug-cc-pVDZ; bottom: XMCQDPT2/aug-cc-pVDZ/cc-pVDZ. Extended QM + correction.

5.3.3 Implications of the CT state and possible mechanism for photoreaction

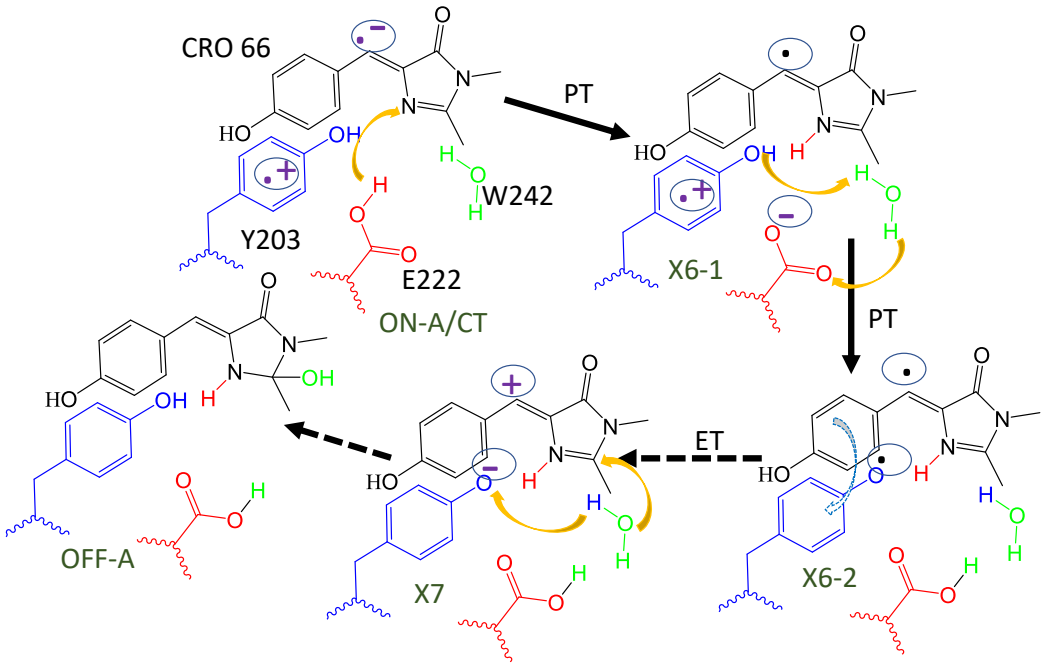


Figure 5.13: Proposed reaction initiated by the population of the CT state. Solid orange arrows show proton transfer and dashed blue arrows show electron transfer. AIMD and excited-state optimization reveal that the steps leading to the formation of X6-2/X7 are nearly barrierless and proceed on the scale of ∼100-200 fs. The last two steps (shown by dashed arrows), back electron transfer from Chro to Tyr203, nucleophilic addition of OH^- to Chro, and reprotonation of Tyr203, are hypothesized. The structures of the possible intermediates are defined in Fig. 5.24 in the Appendix G.

Fig. 5.4 shows the essential steps of Dreiklang's photocycle and outlines the revised photoconversion mechanism via the CT state. The CT state can be populated either via direct excitation or by non-adiabatic transition from the LE state. This is followed by a rapid proton transfer from a nearby residue. The protonated neutral radical chromophore loses the extra electron and undergoes nucleophilic addition of OH^- from the nearby water; this is the slowest, rate-determining step. Below we describe the computational support for the proposed mechanism.

To investigate possible excited-state pathways, we carried out geometry optimization and AIMD simulations for the CT and LE states (for the on-A-HSE-GLUP structure). Fig. 5.13 shows the structural transformation along the AIMD/optimization trajectories. Fig. 5.14 shows additional details of the AIMD simulations: energy profiles of the two lowest electronic states (Kohn-Sham reference state and the lowest TD-DFT state) and the charges of the key residues (chromophore and Tyr203) in these two states. The abrupt changes in the charges clearly indicate the instances of proton transfer.

In the CT state, both the optimization and AIMD simulations show rapid (on the scale of $\sim$100-250 fs) and barrierless proton-transfer steps leading to the formation of the protonated chromophore; this can be rationalized by an increased basisity of the imidozalinone nitrogen caused by the electron attachment. First, the proton is transferred from Glu222 to the imidozalinone nitrogen (this happens within 50 fs). Then Glu222 is reprotonated via proton transfer from Tyr203 (acidified as a result of the electron transfer to the chromophore) via a water-mediated pathway. This process is completed in 200-250 fs. At this point, the CT state is energetically nearly degenerate with the reference Kohn-Sham state (S_0); or, in other words, the Chro-Tyr203 radical pair (neutral protonated radical chromophore and neutral deprotonated Tyr203 radical, X6-2 structure) is nearly isoenergetic with the closed-shell ion-pair state (in which the chromophore is protonated and positively charged and Tyr203 is deprotonated and negatively charged, X7 structure). Hence, one can assume effective back-electron transfer resulting in the formation of the ground-state X7 intermediate.

Based on these observations, X6-2 (Chro$^{\cdot}$-Tyr203$^{\cdot}$ radical pair) or X7 (Chro^{+}-Tyr203^{-} ion pair) are our candidates for the intermediate X observed spectroscopically in the time-resolved study[10]. Experimentally[10], strong transient absorption was observed at 2.67-2.88 eV (with 100 fs kinetics). At the nanosecond scale, the formation of the intermediate X adsorbing at 2.76

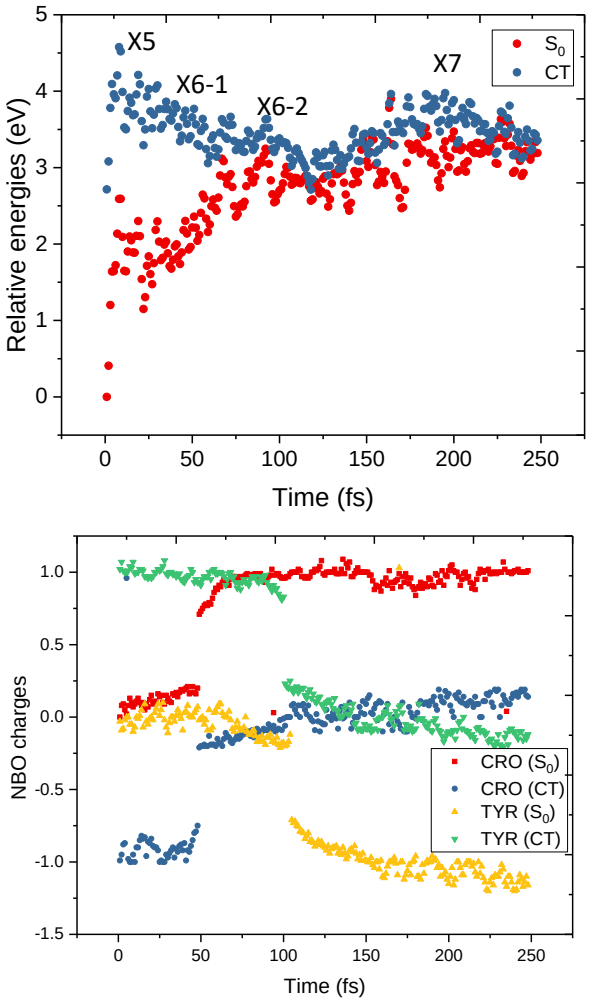


Figure 5.14: Left: Energies of the Kohn-Sham reference state (S_0) and the CT state along the AIMD trajectory on the CT potential energy surface. Right: Charges on the chromophore and Tyr203 in the Kohn-Sham reference state and the CT state (lowest TDDFT state). Labels X5, X6, and X7 denote points along the trajectories when structures resembling these intermediates are formed (see Fig. 5.24 in the Appendix F; X6-1 refers to HSE-GLU; X6-2 refers to HSE-GLUP2).

eV was observed. Hence, both short-time transient absorption and longer time-scale absorption occurs at about 2.8 eV, which is 0.2 eV red-shifted relative to the absorption of the A form.

At the geometry taken from AIMD trajectory at time ∼248 fs, the excitation energy of the ion pair X7 (computed as the lowest bright transition from the Kohn-Sham reference state) is 3.56 eV (oscillator strength 0.40), which is too high compared to the experimental absorption

of X. On the other hand, the excitation energy of the radical pair X6-2 (computed as the lowest bright transition from the lowest TDDFT state) is 2.93 eV (oscillator strength 0.22), which is close to the experimental value. The large difference in the excitation energies of the two structures can be easily rationalized: protonation of the closed-shell neutral chromophore should lead to a blue-shift relative to the parent on-A form, whereas the absorption of the radical anion (chromophore with the additional electron) or protonated neutral radical is expected to be red-shifted relative to the respective closed-shell parent species. Hence, X6-2 appears to be a good candidate for the hot I* intermediate formed on the femtosecond timescale[10]. Further changes in its excitation energy (leading to a small blue shift in the absorption of X relative to I*) are anticipated as the result of the structural relaxation of the protein.

A back electron-transfer step (Chro→Tyr203) would result in the formation of X7 in which the chromophore is positively charged and, therefore, appears to be a good candidate for nucleophilic attack by the nearby water, leading to the formation of the hydrated chromophore and reprotonation of Tyr203. Our preliminary calculations indicate that this step would need to overcome a barrier—the scan along the water-imidozalinone distance (Fig. 5.29 in the Appendix G) yields a barrier of ∼25 kcal/mol, which is very similar to the barrier of the thermal recovery reaction. This is a relatively crude estimate, which should be regarded as an upper bound on the barrier; more accurate estimates will be a subject of future studies. The delayed appearance of the hydrated chromophore is consistent with such a barrier. We note that, in contrast to the thermal recovery reaction, the reaction may still be rather fast, because of the high excess energy available to the system (see left panel in Fig. 5.14). We validated (by AIMD simulations) that once the system reaches this transition state, the dynamics swiftly proceeds downhill, leading to the formation of the hydrated chromophore and reprotonated Tyr203. The AIMD simulations also show that the reverse reaction, from X7 back to the neutral chromophore and reprotonated Tyr203, is very efficient and can compete with the final step of the nucleophilic addition. This

competition between the (slow) nucleophilic addition step and the (fast) reverse reaction, along with other possible channels, is likely to be responsible for a small quantum yield of the photo-transformation, despite the fast and barrierless initial steps.

In contrast to the CT state, geometry optimization and AIMD simulation (1 ps long trajectory) on the LE potential energy surface do not show any significant structural changes, i.e., no evidence of the ultrafast ESPT from the chromophore (Fig. 5.28) posited in Ref. 10.

In summary, the following picture of Dreiklang's photocycle emerges from the results of our theoretical modeling:

1. Excitation of the anionic form (peak B in the ON-state) leads to fluorescence.

2. Excitation of the neutral form (peak A in the ON-state) leads to non-adiabatic transition to the CT state, from which photochemical transformation ensues. It can also lead to ESPT and fluorescence from the anionic state (as in the main photocycle of wt-GFP), but this channel is strongly suppressed.

This picture differs from the mechanism outlined in Ref. 10, where it was proposed that photochemistry unravels in the anionic state, formed by ESPT of photoexcited form A. We note that the ESPT mechanism does not explain why there is no photoconversion upon the direct excitation of the anionic form. In contrast, our proposed mechanism via the CT state, which can only be populated by the excitation of the neutral form, explains the essential trait of Dreiklang: the decoupling of the fluorescence excitation (produced via the anionic form) from the photoconversion (produced by excitation of the neutral form).

Ref. 10 invoked ESPT because of the observed isotope effect. But this effect can be explained by concerted proton transfer to the chromophore in CT state. Ref. 10 invoked ESPT to explain the observed short-time dynamics (510 fs) and commented that this process is an order of magnitude faster than ESPT in GFP (2 ps). This is inconsistent with the lack of strong

fluorescence following the excitation of peak A and increased pKa of the chromophore, which greatly reduces the thermodynamic drive for proton transfer in the excited state. Our AIMD simulations on the LE PES show no evidence of the ultrafast ESPT. The authors of Ref. 10 also commented that the putative ESPT in Dreiklang is significantly less sensitive to H/D exchange than ESPT in GFP, deuterium slowing the observed kinetics by a factor of 1.5 instead of 5. Our simulations strongly suggest that what is seen on the femtosecond time scale is formation of the radical pair Chro$^{\cdot}$-Tyr203$^{\cdot}$ in which the chromophore is protonated on imidozalinone's nitrogen and Tyr203 is deprotonated. Our dynamics show 250 fs time for proton transfers, but one needs also to include time for non-adiabatic transition from the LE state populating the CT state. Sub-picosecond timescales are very likely and there should be some isotope effect.

In addition, our revised photocycle is consistent with the following observations. As pointed out in Ref. 7, the essential difference between the parent system (Citrine) and Dreiklang is the upshift of the pKa of the ON-state of the chromophore (7.2 versus 5.7), which increased the effectiveness of photoconversion. Larger pKa suggests that the ESPT from the neutral form is suppressed, making the population of the CT state more competitive. Note that the fluorescence excitation spectrum (Fig. 1B from Ref. 7) shows that very little fluorescence is produced by excitation of the peak A. We note that the photoconversion is achieved by continuous irradiation in the course of $\sim$5 s, which suggests that the quantum yield for this process is relatively small.

Ref. 7 emphasized that Tyr203 and Glu222 (and Gly65) are crucial for Dreiklang function. The authors also comment that in the fluorescent-state, Tyr203 and Glu222 form hydrogen bonds to a water molecule and thereby stabilize it in close vicinity to the C65 of the chromophore, a situation that is different in the nonswitchable GFP (avGFP-S65T). This strengthens the argument that the reaction may proceed by concerted proton transfer from water-to-Glu222-to-Chro.

We conclude by noting that neither the ET step (population of the CT state) nor the subsequent barrierless proton-transfer steps should be affected by temperature, meaning that these steps would not be suppressed at cryogenic temperatures. We also note that previous studies indicate that the photoinduced recovery of the ON-state is likely to be barierrless[20]. These observations suggest that Dreiklang could be a good starting point for developing photoswitchable fluorescent proteins that can operate at low temperatures, as desired for cryogenic super-resolution imaging applications[54,55].

5.4 Conclusion

In this contribution, we investigated properties of the fluorescent protein Dreiklang using high-level electronic structure methods combined with QM/MM and dynamics simulations. The results allowed us to quantify the spectral consequences of possible protonation states of the key residues around the chromophore and to refine the properties of the low-lying excited states. The key finding is that the neutral (protonated) ON-state of Dreiklang features a low-lying state of CT character (Tyr203→Chro), which is energetically close to the LE and is strongly affected by hydrogen bonding and thermal motions. Once this state is populated (either by direct photoexcitation or via non-adiabatic transition), the system undergoes a cascade of proton transfer steps leading to the protonation of the chromophore (on imidozalinone's nitrogen) and formation of the neutral Chro-Tyr203 radical pair, nearly iso-energetic with the ion-pair state (in which Tyr203 is in deprotonated anionic state and the chromophore is positively charged). This structure appears to be a good candidate for nucleophilic addiction of hydroxide to the chromophore, coupled with reprotonation of Tyr203.

This mechanism is consistent with the available experimental data. The disrupted hydrogen-bonding network around the chromophore and its reduced acidity explain why the canonical ESPT route is strongly suppressed, making the CT channel competitive. The key role of the CT

state, which is only accessible by photoexcitation of the on-A form, explains the unique feature of Dreiklang, the decoupling of fluorescence from photoswitching.

5.5 Appendix A: Definitions of protonation states

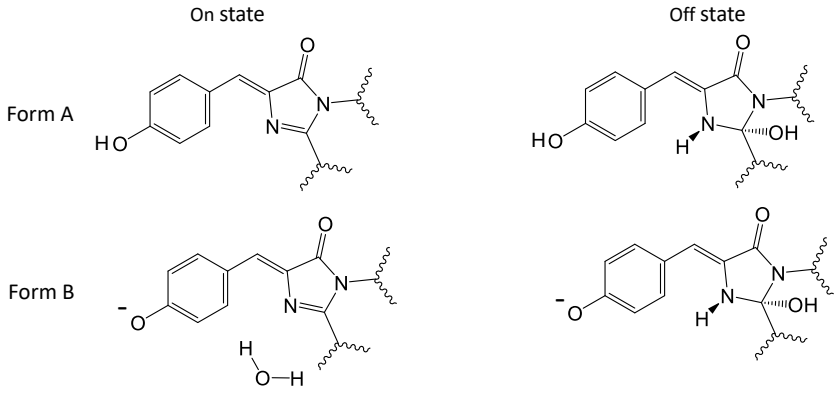


Figure 5.15: Definition of chromophore states.

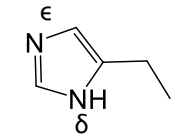
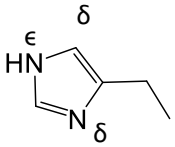
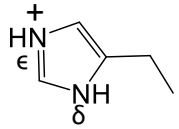
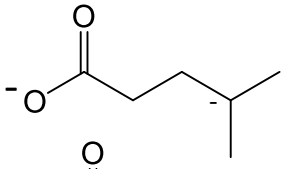
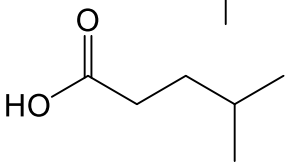
	Charge	
HSD/HID	0	
HSE/HIE	0	
HSP/HIP	+1	
GLU	-1	
GLUP	0	

Figure 5.16: Definition of protonation states of Glu222 and His145 in Dreiklang. GLUP can exist in two conformations: As shown or protonated on the other oxygen (GLUP2).

5.6 Appendix B: Computational details

In addition to the structures obtained by the computational protocol described in the main manuscript, we also consider the structures from our previous study[20] in which we started with 3ST2 structure and used QM/MM optimization with electrostatic embedding, as implemented in *NWChem*. The QM part was described by M06-L/cc-pVDZ and the MM part was described by the *AMBER* forcefield. In these calculations[20], QM included the chromophore, side chains of Gln94, Arg96, His145, Tyr203, Ser205, and Glu222, and seven water molecules. This definition is similar to our extended QM. We note that these model structures also included additional water molecule, which is present in 3ST3 structure (OFF-state) but not seen in 3ST2 and 3ST4. The comparisons between the two protocols quantify the effect of the level of theoretical treatment.

The key structural parameters two sets of structures are compared in Tables 5.8-5.13 below and graphically in Figs. 5.18 and 5.19.

5.7 Appendix C: Forcefield parameters for the neutral hydrated chromophore

To derive missing forcefield parameters (for the OFF-form of the chromophore) we followed a protocol described in our previous work[37,38]. The key equations and the values of the forcefield parameters are given below.

$$\Delta q_{(on,charmm-qm)} = q_{on,charmm} - q_{NBO(on,qm)} \tag{5.1}$$

$$q_{(off,charmm)} = q_{NBO(off,qm)} + \Delta q_{(on,charmm-qm)} \tag{5.2}$$

$$E = k(b - b_0)^2 \tag{5.3}$$

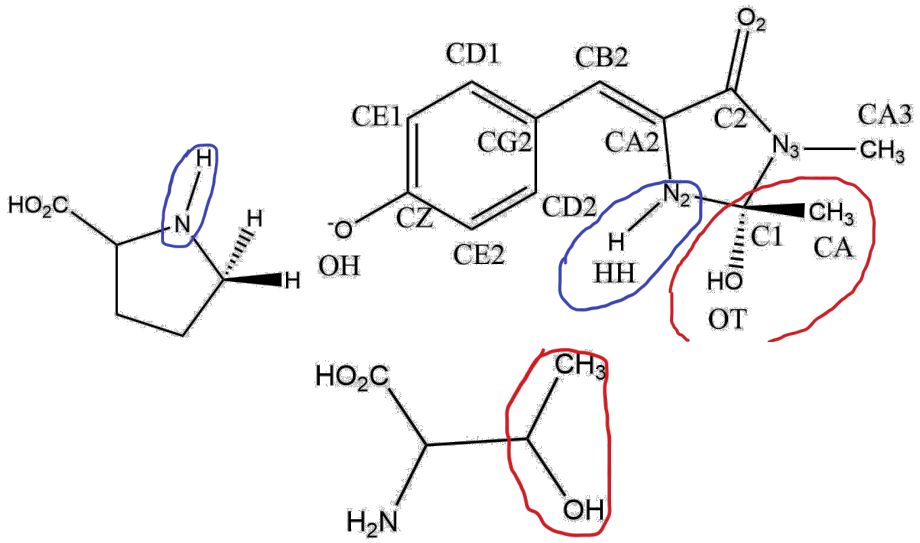


Figure 5.17: From left to right: proline, chromophore in off-state, threonine.

$$k_{off,param} = \frac{k_{off,theory}}{k_{on,theory}} \times k_{on-charmm} \tag{5.4}$$

b_0 is the equilibrium bond length.

$$E = k(A - A_0)^2 \tag{5.5}$$

$$k_{off,param} = \frac{k_{off,theory}}{k_{on,theory}} \times k_{on-charmm} \tag{5.6}$$

A_0 is the equilibrium bond angle.

$$E = k[1 + cos(n\phi - \delta)] \tag{5.7}$$

where n is the phase, δ is the optimized dihedral angle.

$$k_{off,param} = \frac{k_{off,theory}}{k_{on,theory}} \times k_{on-charmm} \tag{5.8}$$

Table 5.2: Partial charges in the OFF-state.

Atom, off	on, charmm	on, qm	$\Delta q_{(on,charmm-qm)}$	$\Delta q_{adjusted}$	off, qm	off, charmm
C1(threonine)	0.10	0.15	-0.05	-0.02	0.67	0.69
N2 (proline)	-0.74	-0.28	-0.46	-0.43	-0.68	-0.25
N3	-0.64	-0.52	-0.12	-0.09	-0.56	-0.47
C2	0.8	0.74	0.06	0.09	0.74	0.65
O2	-0.61	-0.60	-0.01	0.02	-0.63	-0.65
CA2	0.24	0.05	0.19	0.22	0.12	-0.10
CB2	-0.10	-0.09	-0.01	0.02	-0.24	-0.26
HB2	0.1	0.28	-0.18	-0.15	0.27	0.42
CG2	0.00	-0.11	0.11	0.14	-0.09	-0.23
CD1	-0.115	-0.14	0.025	0.06	-0.185	-0.245
HD11	0.115	0.27	-0.155	-0.12	0.25	0.37
CD2	-0.115	-0.14	0.025	0.06	-0.185	-0.245
HD21	0.115	0.27	-0.155	-0.12	0.25	0.37
CE1	-0.115	-0.27	0.155	0.19	-0.275	-0.465
HE11	0.115	0.27	-0.155	-0.12	0.25	0.37
CE2	-0.115	-0.27	0.155	0.19	-0.275	-0.465
HE21	0.115	0.27	-0.155	-0.12	0.25	0.37
CZ	0.11	0.38	-0.27	-0.24	0.34	0.58
OH	-0.54	-0.68	0.14	0.17	-0.70	-0.87
OHH	0.43	0.52	-0.09	-0.06	0.50	0.56
OT (threonine)	-0.78	-0.65	-0.13	-0.10	-0.75	-0.65
HT(threonine)	0.50	0.44	0.06	0.09	0.50	0.41
HH (proline)	0.41	0.11	0.30	0.33	0.44	0.11
CA3		-0.18				-0.18
HA31		0.09				0.09
HA32		0.09				0.09
C		0.51	–	–	–	0.51
O		-0.51 –	–	–	–	-0.51
N		-0.47 –	–	–	–	-0.47
HN		0.31 –	–	–	–	0.31
CA		-0.02				-0.02
HA1		0.09				0.09
HA2		0.09				0.09

Table 5.3: Optimized bond lengths (in Å) involving key atoms.

Bonds	$B_{on,charmm}$	$B_{on,opt}$	$B_{off,opt}$
C1-N2 (proline)	1.434	1.46	1.45
N2-CA2	1.40	1.41	1.39
N2-HH (proline)	0.997	1.02	1.01
C1-OT (threonine)	1.42	1.40	1.41
C1-CA	1.49	1.49	1.53
C1-N3	1.39	1.38	1.45

Table 5.4: Parameterization of the force constant k for bond lengths in kcal/mol/Å^2.

Bonds	$k_{on,charmm}$	$k_{on,theory}$	$k_{off,theory}$	$k_{off,theory}/k_{on,theory}$	$k_{off,param}$
C1-N2 (proline)	320	1156.50	1226.15	1.06	339.27
N2-CA2	400	940.00	1257.53	1.34	535.12
N2-HH (proline)	440	976.40	1044.80	1.07	470.82
C1-OT (threonine)	428	847.14	744.22	0.88	376.00
C1-CA	354	562.25	542.79	0.965	341.75
C1-N3	320	1156.50	1226.15	1.06	339.27

Table 5.5: Optimized bond angles (in degrees) involving key atoms.

Angles	$A_{on,charmm}$	$A_{on,opt}$	$A_{off,opt}$
N2-C1-N3	114.0	113.99	102.31
C1-N2-CA2	106.0	106.18	111.21
HH-N2-C1 (proline)	117.0	111.44	115.90
HH-N2-CA2	117.0	111.44	117.8
OT-C1-CA (threonine)	110.1	112.6	110.31
OT-C1-N2 (threonine)	110.1	112.6	107.90
OT-C1-N3 (threonine)	110.1	112.6	111.39
N2-CA2-CB2	129.5	129.58	130.17
N2-CA2-C2	108.3	108.73	106.54
C1-N3-C2	107.9	113.47	108.26
CA-C1-N3 (threonine)	113.5	111.6	112.01

Table 5.6: Parameterization of the force constant k for bond angles in kcal/mol/rad$^{\circ 2}$.

Angles	$k_{on,charmm}$	$k_{on,theory}$	$k_{off,theory}$	$k_{off,theory}/k_{on,theory}$	$k_{off,param}$
N2-C1-N3	130.0	444.28	347.33	0.78	101.6
C1-N2-CA2	130.0	438.63	259.98	0.59	77.05
HH-N2-C1 (proline)	35.0	89.73	79.69	0.89	31.08
HH-N2-CA2	35.0	89.73	79.69	0.89	31.08
OT-C1-CA (threonine)	75.7	232.18	259.37	1.12	84.56
OT-C1-N2(N3) (threonine)	75.7	232.18	259.37	112	84.56
N2-CA2-CB2	45.8	151.23	169.42	1.12	51.3
N2-CA2-C2	130.0	472.5	376.50	0.797	103.6
C1-N3-C2	130.0	498.24	305.6	0.61	79.7
CA-C1-N3 (threonine)	70.0	180.72	179.47	0.99	69.5

Table 5.7: Parameterization of the force constant k for dihedral angles; δ in degrees, k in kcal/mol.

Angles	$k_{on,charmm}$	n	δ	$k_{off,theory}/k_{on,theory}$	$k_{off,param}$
OT-C1-N2-HH	0.16	3	180	0.263	0.053
OT-C1-N2-CA2	0.20	3	0	13.83	2.213
CA-C1-N2-HH	0.16	3	180	0.263	0.053
HH-N2-CA2-CB2	0.16	3	0	0.263	0.053
HH-N2-CA2-C2	0.20	3	180	13.83	2.213
HH-N2-C1-N3	0.20	3	180	13.83	2.213
CA-C1-N3-CA3	0.16	3	0	0.263	0.053
OT-C1-N3-CA3	0.16	3	0	0.263	0.053
CA-C1-N3-C2	0.20	3	180	13.83	2.213

5.8 Appendix D: Structures of model systems

Table 5.8: Comparison of the distances (in Å) from MD and QM/MM simulations with crystal structure 3ST4 (ON-state). The chromophore is neutral (A-form). 'md' denotes structures averaged over equilibrium MD trajectories. 'opt' and 'opt2' denote the QM/MM optimized structures obtained with present protocol and with the protocol from Ref. 20, respectively.

D		3ST4	HSE-GLU	HSE-GLUP	HSD-GLU	HSD-GLUP	HSP-GLU
d1	md	3.52	3.42(0.3)	3.96(0.8)	3.74(0.4)	3.63(0.4)	4.68(1.0)
	opt	–	3.31	3.92	3.53	4.2	3.78
	opt2	–	3.45	3.50	3.49		3.46
d2	md	2.97	3.94(0.3)	3.44(0.3)	3.92(0.3)	3.54(0.4)	4.00(0.3)
	opt	–	3.99	2.78	4.06	2.86	4.00
	opt2	–	3.46	3.09	3.44		3.49
d3	md	2.73	3.42(0.3)	2.77(0.1)	2.74(0.1)	2.77(0.1)	2.74(0.1)
	opt	–	2.58	2.62	2.6	2.67	2.61
	opt2	–	2.77	2.85	2.80		2.81
d4	md	3.81	3.60(0.2)	4.82(0.8)	3.62(0.2)	3.91(0.3)	4.02(0.7)
	opt	–	3.53	3.67	3.52	3.77	3.38
	opt2	–	3.96	3.92	3.96		4.16
d5	md	4.04	4.38(0.4)	4.08(0.3)	4.05(0.4)	3.86(0.3)	4.67(0.4)
	opt	–	—	—	—	—	—-
	opt2	–	—	–	—	—	—
d6	md	3.64	3.85(0.2)	3.82(0.2)	3.88(0.2)	3.87(0.2)	3.82(0.2)
	opt	–	3.69	3.52	3.63	3.6	3.59
	opt2	–	3.85	3.63	3.68		3.67
d7	md	2.63	V.L	2.69(0.1)	2.71(0.1)	2.74(0.1)	3.19(0.8)
	opt	–	2.97	2.59	2.63	2.63	3.59
	opt2	–	2.59	2.66	2.66		2.67
d8	md	3.41	3.18(0.2)	3.43(0.2)	3.49(0.3)	3.43(0.2)	3.6(0.3)
	opt	–	3.25	3.22	3.71	3.27	3.59
	opt2	–	3.05	3.26	3.02		3.01
d9	md	2.89	2.92(0.2)	2.99(0.3)	2.88(0.2)	2.99(0.3)	2.96(0.2)
	opt	–	2.64	2.68	2.78	2.67	2.79
	opt2	–	2.82	2.81	2.79		2.81
d10	md	2.69	2.63(0.1)	2.81(0.2)	2.63(0.1)	2.78(0.2)	2.66(0.1)
	opt	–	2.58	2.7	2.8	2.72	2.73
	opt2	–	2.67	2.75	2.67		2.67

Table 5.9: Comparison of the distances (in Å) from MD and QM/MM simulations with crystal structure 3ST4 (ON-state). The chromophore is neutral (A-form). 'md' denotes structures averaged over equilibrium MD trajectories. 'opt' and 'opt2' denote the QM/MM optimized structures obtained with present protocol and with the protocol from Ref. 20, respectively.

D		3ST4	HSE-GLU	HSE-GLUP	HSD-GLU	HSD-GLUP	HSP-GLU
d11	md	2.73	V.L	3.76(0.8)	2.9(0.2)	3.02(0.3)	2.93(0.3)
	opt	–	3.01	2.7	2.52	2.59	2.89
	opt2	–	2.72	2.69	2.67		2.63
d12	md	3.05	V.L	3.06(0.3)	3.53(0.4)	3.42(0.4)	3.89(0.4)
	opt	–	3.33	2.58	2.83	2.95	3.23
	opt2	–	3.53	2.70	3.10		2.74
d13	md	2.89	V.L	2.84(0.3)	3.4(0.4)	3.41(0.6)	3.87(0.4)
	opt	–	–	–	–	–	–
	opt2	–	—	–	—	—	—
d14	md	4.18	2.67(0.1)	4.28(0.5)	2.66(0.1)	3.48(0.4)	2.68(0.1)
	opt	–	2.7	4.72	2.56	4.32	2.5
	opt2	–	5.06	4.38	5.02		5.03

Table 5.10: Comparison of the distances (in Å) from MD and QM/MM simulations with crystal structure 3ST4 (ON-state). Chromophore is anionic (B-form). 'md' denotes structures averaged over equilibrium MD trajectories. 'opt' and 'opt2' denote the QM/MM optimized structures obtained with present protocol and with the protocol from Ref. 20, respectively.

D		3ST4	HSE-GLU	HSE-GLUP	HSD-GLU	HSD-GLUP	HSP-GLUP
d1	md	3.52	3.90(0.3)	3.19(0.2)	3.57(0.4)	3.32(0.3)	3.05(0.3)
	opt	–	3.66	3.39	3.39	3.47	2.88
	opt2	–		3.30		3.24	3.10
d2	md	2.97	3.98(0.3)	3.17(0.3)	4.62(0.4)	3.13(0.2)	3.07(0.2)
	opt	–	4.02	2.74	4.11	2.9	2.79
	opt2	–		3.47		3.49	3.08
d3	md	2.73	2.72(0.1)	2.75(0.1)	2.74(0.1)	2.72(0.1)	2.73(0.1)
	opt	–	2.52	2.57	2.51	2.57	2.64
	opt2	–		2.74		2.76	2.81
d4	md	3.81	3.78(0.3)	4.83(0.8)	3.73(0.3)	3.97(0.3)	4.42(0.4)
	opt	–	3.93	3.84	3.33	4.06	4.51
	opt2	–		4.01		4.01	3.99
d5	md	4.04	4.22(0.2)	4.74(0.3)	4.15(0.3)	3.82(0.2)	4.14(0.3)
	opt	–	—	—	—	—	—
	opt2	–	–	–	–	–	—
d6	md	3.64	3.96(0.2)	3.78(0.2)	4.00(0.3)	3.84(0.2)	3.94(0.2)
	opt	–	3.84	3.64	3.97	3.67	3.64
	opt2	–		3.68		3.83	3.68
d7	md	2.63	2.79(0.2)	5.24(1.0)	2.8(0.2)	2.75(0.2)	2.79(0.2)
	opt	–	2.86	5.14	2.87	2.59	2.54
	opt2	–		2.74		2.68	2.63
d8	md	3.41	3.00(0.1)	3.86(0.5)	3.01(0.1)	3.72(0.4)	3.88(0.4)
	opt	–	3.19	3.29	3.47	3.23	3.51
	opt2	–		3.34		3.02	3.27
d9	md	2.89	2.83(0.1)	2.86(0.2)	2.84(0.1)	2.84(0.1)	2.85(0.2)
	opt	–	2.71 2.88	2.58	2.77	2.77	
	opt2	–		2.87		2.86	2.80
d10	md	2.69	2.66(0.1)	2.8(0.2)	2.64(0.1)	2.78(0.2)	2.78(0.2)
	opt	–	2.46	2.64	2.89	2.77	2.81
	opt2	–		2.91		2.89	2.75

Table 5.11: Comparison of the distances (in Å) from MD and QM/MM simulations with crystal structure 3ST4 (ON-state). Chromophore is anionic (B-form). 'md' denotes structures averaged over equilibrium MD trajectories. 'opt' and 'opt2' denote the QM/MM optimized structures obtained with present protocol and with the protocol from Ref. 20, respectively.

D		3ST4	HSE-GLU	HSE-GLUP	HSD-GLU	HSD-GLUP	HSP-GLUP
d11	md	2.73	3.67(0.7)	4.54(0.8)	2.9(0.2)	2.9(0.2)	2.97(0.4)
	opt	–	3.95	4.11	2.66	2.74	2.76
	opt2	–		2.62		2.63	2.70
d12	md	3.05	5.41(0.3)	3.07(0.4)	3.43(0.4)	3.25(0.3)	3.33(0.3)
	opt	–	3.31	2.89	3.76	2.85	2.65
	opt2	–		3.37		3.31	2.76
d13	md	2.89	3.38(0.6)	4.73(0.5)	3.25(0.3)	2.81(0.2)	3.24(0.5)
	opt	–	–	–	–	–	—
	opt2	–	—	—	–	–	–
d14	md	4.18	2.68(0.1)	4.33(0.5)	2.65(0.1)	4.1(0.4)	5.24(0.4)
	opt	–	2.65	4.27	2.55	4.59	5.79
	opt2	–		4.90		4.99	4.53

Table 5.12: Comparison of the distances (in Å) from MD and QM/MM simulations with crystal structure 3ST3 (OFF-state). Chromophore is neutral. 'md' denotes structures averaged over equilibrium MD trajectories. 'opt' and 'opt2' denote the QM/MM optimized structures obtained with present protocol and with the protocol from Ref. 20, respectively.

D		3ST3	HSE GLU P	HSD GLU P	HSE GLU P2	HSD GLU P2	HSE GLU	HSD GLU	HSP GLU
d1	md	3.35	3.37 (0.3)	3.75 (0.5)	3.71 (0.6)	3.68 (0.2)	3.63 (0.5)	3.53 (0.4)	2.97 (0.2)
	opt	–	4.1	2.96	3.17	4.2	4.29	3.48	3.38
	opt2	–	3.58	3.27			3.44	3.47	3.46
d2	md	2.46	3.25 (0.2)	3.28 (0.2)	4.21 (0.4)	4.53 (0.4)	3.20 (0.2)	3.23 (0.2)	3.13 (0.1)
	opt	–	2.8	2.85	2.76	2.67	2.66	2.67	2.78
	opt2	–	2.82	2.86			2.62	2.63	2.63
d3	md	2.85	4.23 (0.6)	4.32 (0.4)	3.56 (0.3)	3.49 (0.3)	3.91 (0.4)	3.67 (0.5)	3.51 (0.3)
	opt	–	3.5	3.45	3.13	3.09	2.96	3.03	3.0
	opt2	–	3.17	3.17			2.96	2.96	2.98
d4	md	3.01	2.69 (0.1)	2.7 (0.1)	2.71 (0.1)	2.67 (0.1)	2.68 (0.1)	2.68 (0.1)	2.67 (0.1)
	opt	–	2.66	2.68	2.66	2.67	2.6	2.62	2.61
	opt2	–	2.82	2.79			2.80	2.82	2.82
d5	md	3.82	4.13 (0.3)	4.93 (0.4)	4.84 (0.6)	5.51 (0.7)	5.66 (0.5)	3.81 (0.3)	4.69 (0.4)
	opt	–	3.6	4.12	3.96	4.05	3.85	3.79	3.78
	opt2	–	4.16	3.95			4.02	4.04	4.15
d6	md	3.98	3.91 (0.3)	4.96 (0.4)	4.62 (0.4)	4.14 (0.3)	4.50 (0.4)	3.98 (0.3)	4.44 (0.4)
	opt	–	–	–	–	–	–	–	–
	opt2	–	–	–	–	–	—	–	–
d7	md	3.91	3.98 (0.3)	3.78 (0.2)	3.80 (0.2)	3.94 (0.3)	3.99 (0.3)	4.05 (0.3)	4.21 (0.3)
	opt	–	3.68	3.67	3.66	3.74	3.95	3.7	3.67
	opt2	–	3.70	3.73			3.72	3.69	3.67

Table 5.13: Comparison of the distances (in Å) from MD and QM/MM simulations with crystal structure 3ST3 (OFF-state). Chromophore is neutral. 'md' denotes structures averaged over equilibrium MD trajectories. 'opt' and 'opt2' denote the QM/MM optimized structures obtained with present protocol and with the protocol from Ref. 20, respectively.

D		3ST3	HSE GLU P	HSD GLU P	HSE GLU P2	HSD GLU P2	HSE GLU	HSD GLU	HSP GLU
d8	md	2.59	L.V	3.86 (1.1)	10.4 (7.9)	2.87 (0.4)	L.V	2.64	L.V
	opt	–	2.7	2.71	2.96	2.61	2.71	2.83	2.67
	opt2	–	2.7	2.67			2.63	2.66	2.68
d9	md	2.64	L.V	7.28 (0.7)	7.38 (8.2)	3.97 (0.6)	L.V	2.80	L.V
	opt	–	2.71	4.36	2.62	2.53	2.6	2.62	2.51
	opt2	–	2.65	2.62			2.75	2.67	2.63
d10	md	3.17	L.V	4.79 (0.7)	9.60 (8.3)	5.40 (0.3)	L.V	3.37 (0.4)	L.V
	opt	–	4.5	3.8	2.62	2.83	4.03	2.87	2.54
	opt2	–	2.68	3.55			3.44	3.13	2.75
d11	md	2.80	L.V	3.95 (1.4)	7.37 (8.6)	2.82 (0.2)	L.V	3.21 (0.4)	L.V
	opt	–	–	–	–	–	–	–	–
	opt2	–	–	–	–	–	–	–	–
d12	md	2.67	4.29 (0.9)	4.38 (0.8)	4.72 (0.1)	4.64 (0.5)	2.79 (0.3)	3.04 (0.6)	2.62 (0.1)
	opt	–	3.53	3.6	2.73	2.71	2.68	2.58	2.53
	opt2	–	2.55	2.54			2.68	2.65	2.65
d13	md	2.47	L.V	3.03 (0.4)	2.78 (0.1)	2.89 (0.2)	4.70 (1.8)	4.22 (1.7)	2.81 (0.1)
	opt	–	2.64	2.69	2.6	2.56	3.16	2.56	2.5
	opt2	–	2.65	2.66			2.60	2.58	2.56
d14	md	4.88	3.74 (1.2)	5.18 (0.9)	6.56 (0.4)	6.36 (0.5)	4.52 (0.6)	3.69 (1.0)	4.59 (0.4)
	opt	–	6.02	5.95	4.91	5.04	5.27	4.76	4.82
	opt2	–	4.79	4.91			5.01	4.95	4.93

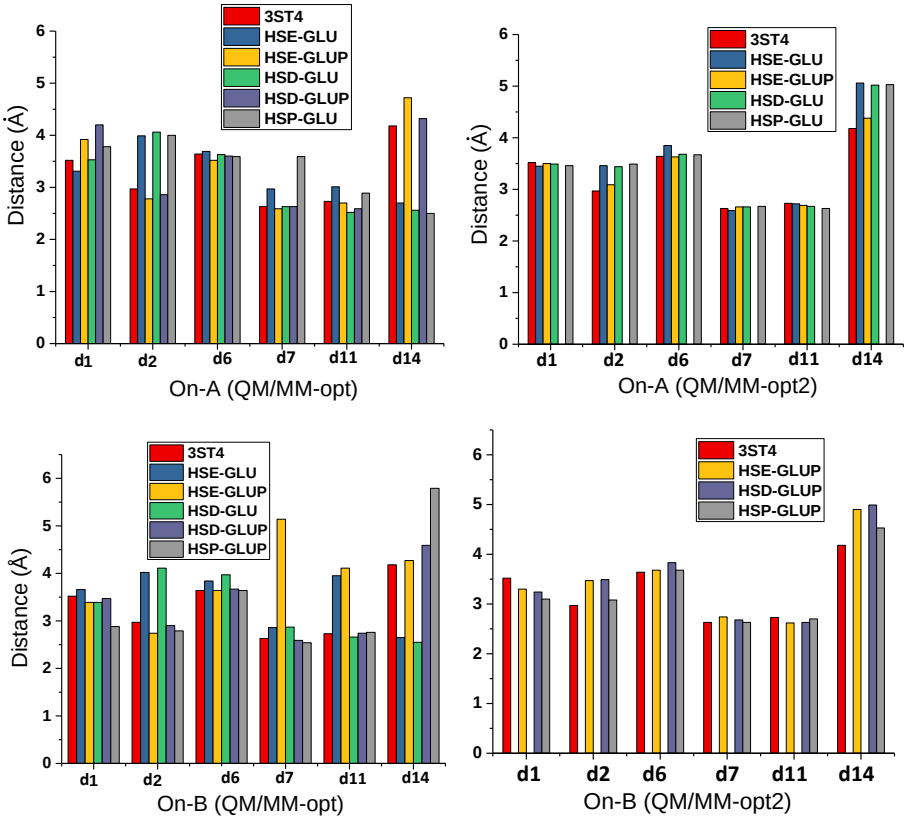


Figure 5.18: Key distances for ON-states: Comparison between crystal structure and QM/MM optimization. OPT1 and OPT2 denote two different protocols (see text). See Fig. 5 in the main text for definitions.

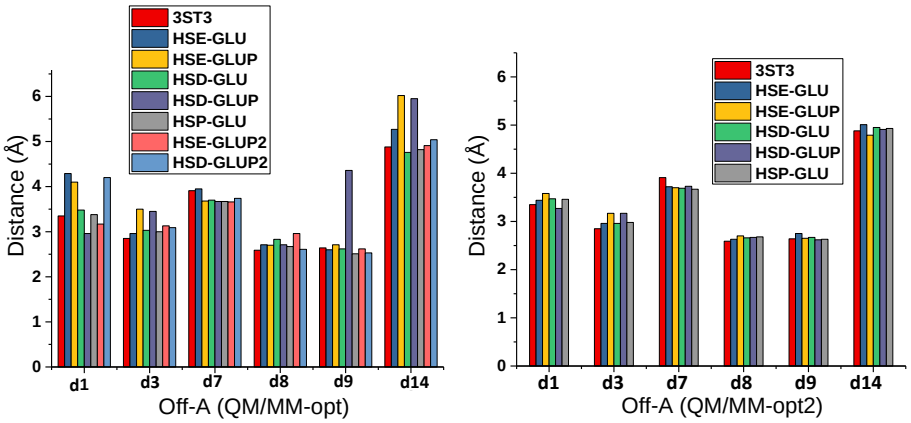


Figure 5.19: Key distances for OFF-states: Comparison between crystal structure, average MD values, and QM/MM optimizations. See Fig. 6 of the main text for definitions.

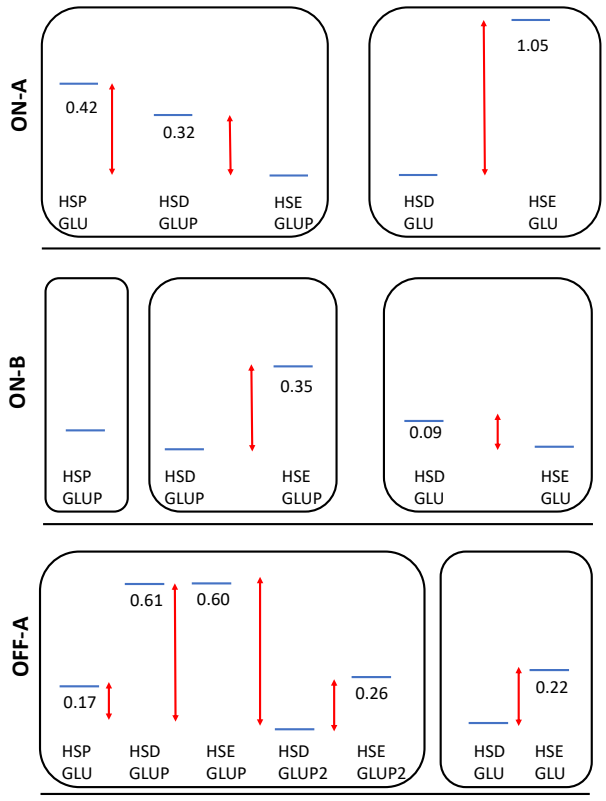


Figure 5.20: Energy ordering (eV) of QM/MM (ONIOM) optimized structures (boxes mark the structures with the same number of atoms in QM).

5.9 Appendix E: Analysis of excited states

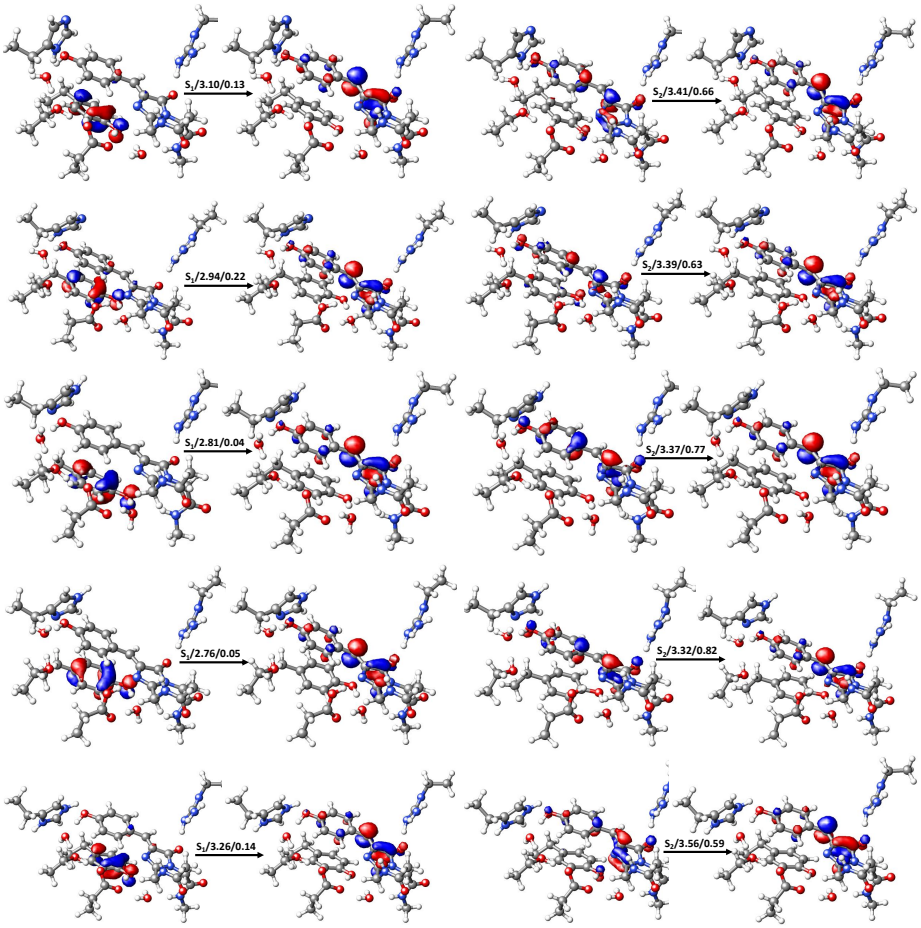


Figure 5.21: NTOs of the lowest excited states of the neutral form and different protonation states of His145 and Glu222; TD-DFT, extended QM. Left: CT state; right: LE state; top-to-bottom: HSD-GLU, HSD-GLUP, HSE-GLU, HSE-GLUP, HSP-GLU.

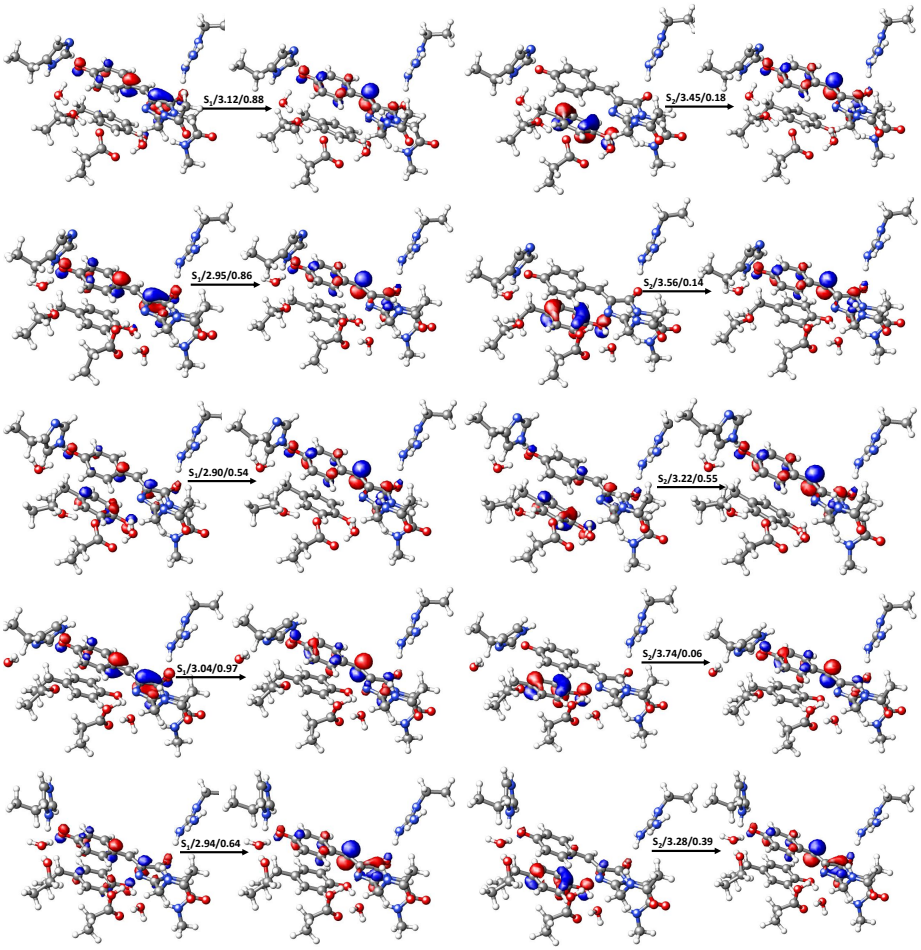


Figure 5.22: NTOs of the lowest excited states of the anionic form and different protonation states of His145 and Glu222; TD-DFT, extended QM. Left: LE state; right: CT state; top-to-bottom: HSD-GLU, HSD-GLUP, HSE-GLU, HSE-GLUP, HSP-GLUP. CT state is pushed to much higher energies and disappears in QM/MM calculations.

Table 5.14: Effect of the protein environment beyond extended QM estimated from the 21 MD snapshots for the neutral chromophore in the ON-state. All energies are in eV; large QM.

System	State	TD-DFT aug-cc-pVDZ QM/MM (MD)	TD-DFT aug-cc-pVDZ QM only (MD)	TD-DFT aug-cc-pVDZ Δ^a	TD-DFT aug-cc-pVDZ QM only (opt)[b]	TD-DFT aug-cc-pVDZ QM/MM-corr[c]
HSD-GLUP	CT	4.01 (0.05)	3.76 (0.17)	+0.25	3.07 (0.26)	3.32 (0.26)
	LE	3.49 (0.56)	3.42 (0.52)	+0.07	3.49 (0.54)	3.56 (0.54)
HSE-GLUP	CT	3.96 (0.10)	3.70 (0.38)	+0.26	2.88(0.06)	3.14 (0.06)
	LE	3.58 (0.52)	3.44 (0.30)	+0.14	3.49 (0.73)	3.63 (0.73)
HSD-GLU	CT	3.84 (0.10)	3.58 (0.19)	+0.26	3.10 (0.13)	3.36 (0.13)
	LE	3.39 (0.49)	3.30 (0.51)	+0.09	3.41 (0.66)	3.50 (0.66)
HSE-GLU	CT	3.93 (0.08)	3.62 (0.11)	+0.31	2.81 (0.04)	3.12 (0.04)
	LE	3.45 (0.56)	3.37 (0.60)	+0.08	3.37 (0.77)	3.45 (0.77)
HSP-GLU	CT	3.88 (0.16)	3.71 (0.12)	+0.17	3.26 (0.14)	3.43 (0.14)
	LE	3.56 (0.63)	3.53 (0.69)	+0.03	3.56 (0.59)	3.59 (0.59)

[a] Δ is the difference in excitation energies in QM/MM and QM only calculation evaluated using structures from 21 MD snapshots.

[b] QM only excitation energies computed using ONIOM optimized structures.

[c] Extrapolated values: QM only excitation energies computed using ONIOM optimized structures plus Δ correction.

Table 5.15: Effect of the protein environment beyond extended QM estimated from the 21 MD snapshots for the anionic chromophore in the ON-state. All energies are in eV; large QM.

System	State	TD-DFT aug-cc-pVDZ QM/MM (MD)	TD-DFT aug-cc-pVDZ QM only (MD)	TD-DFT aug-cc-pVDZ Δ^a	TD-DFT aug-cc-pVDZ QM only (opt)^b	TD-DFT aug-cc-pVDZ QM/MM-corr^c
HSD-GLUP	LE	2.89 (0.91)	2.96 (0.86)	-0.07	3.01 (0.88)	2.94 (0.88)
HSE-GLUP	LE	2.99 (1.02)	3.02 (0.99)	-0.03	3.09 (0.95)	3.06 (0.95)
HSD-GLU	LE	3.07 (0.96)	3.02 (0.90)	+0.05	3.14 (0.98)	3.19 (0.98)
HSE-GLU	LE	3.03 (0.88)	2.92 (0.81)	+0.11	3.04 (0.94)	3.15 (0.94)
HSP-GLUP	LE	3.14 (0.94)	3.22 (0.88)	-0.08	3.05 (0.84)	2.97 (0.84)

[a] Δ is the difference in excitation energies in QM/MM and QM only calculation evaluated using structures from 21 MD snapshots.

[b] QM only excitation energies computed using ONIOM optimized structures.

[c] Extrapolated values: QM only excitation energies computed using ONIOM optimized structures plus Δ correction.

Table 5.16: Effect of the protein environment beyond extended QM estimated from the 21 MD snapshots for the OFF-form (neutral chromophore). All energies are in eV; large QM.

System	State	TD-DFT aug-cc-pVDZ QM/MM (MD)	TD-DFT aug-cc-pVDZ QM only (MD)	TD-DFT aug-cc-pVDZ Δ^a	TD-DFT aug-cc-pVDZ QM only (opt)[b]	TD-DFT aug-cc-pVDZ QM/MM-corr[c]
HSD-GLUP	LE	4.10 (0.44)	3.97 (0.39)	+0.13	3.86 (0.41)	3.99 (0.41)
HSE-GLUP	LE	3.89 (0.51)	3.88 (0.48)	+0.01	3.92 (0.33)	3.93 (0.33)
HSD-GLU	LE	4.12 (0.56)	4.03 (0.36)	+0.09	3.69 (0.53)	3.78 (0.53)
HSE-GLU	LE	3.88 (0.40)	3.79 (0.41)	+0.09	3.52 (0.57)	3.61 (0.57)
HSP-GLU	LE	3.89 (0.63)	3.85 (0.57)	+0.04	3.85 (0.60)	3.89 (0.60)
HSD-GLUP2	LE	3.79 (0.66)	3.63 (0.59)	+0.16	3.52 (0.64)	3.68 (0.64)
HSE-GLUP2	LE	3.77 (0.40)	3.74 (0.32)	+0.03	3.87 (0.58)	3.90 (0.58)

[a] Δ is the difference in excitation energies in QM/MM and QM only calculation evaluated using structures from 21 MD snapshots.

[b] QM only excitation energies computed using ONIOM optimized structures.

[c] Extrapolated values: QM only excitation energies computed using ONIOM optimized structures plus Δ correction.

Table 5.17: TD-DFT excitation energies (eV) of the two lowest states of protein-bound neutral chromophore in the ON-state with different basis sets and different size of QM region; oscillator strength is shown in parentheses.

System	State	Extended QM cc-pVDZ QM only	Extended QM mixed basis[a] QM only	Extended QM aug-cc-pVDZ QM only	Large QM aug-cc-pVDZ QM only
HSD-GLUP	LE	3.43 (0.72)	3.40 (0.61)	3.39 (0.63)	3.49 (0.54)
	CT	2.91 (0.16)	2.96 (0.24)	2.94 (0.22)	3.07 (0.26)
HSE-GLUP	LE	3.38 (0.87)	3.34 (0.82)	3.32 (0.82)	3.39 (0.73)
	CT	2.71 (0.03)	2.80 (0.05)	2.76 (0.05)	2.88 (0.06)
HSD-GLU	LE	3.46 (0.73)	3.41 (0.65)	3.41 (0.66)	3.44 (0.60)
	CT	3.09 (0.08)	3.12 (0.14)	3.10 (0.13)	3.16 (0.11)
HSE-GLU	LE	3.42 (0.80)	3.38 (0.75)	3.37 (0.77)	3.48 (0.67)
	CT	2.74 (0.03)	2.85 (0.05)	2.81 (0.04)	2.88 (0.02)
HSP-GLU	LE	3.62 (0.65)	3.57 (0.58)	3.56 (0.59)	3.58 (0.52)
	CT	3.26 (0.09)	3.28 (0.14)	3.26 (0.14)	3.32 (0.13)

[a] mixed basis: aug-cc-pVDZ for the chromophore and tyrosine and cc-pVDZ for rest of QM.

Table 5.18: TD-DFT excitation energies (eV) of the two lowest states of protein-bound anionic chromophore in the ON-state; oscillator strength is shown in parentheses.

System	State	Extended QM cc-pVDZ QM only	Extended QM aug-cc-pVDZ QM only	Large QM aug-cc-pVDZ QM only
HSD-GLUP	LE	3.03 (0.86)	2.95 (0.86)	3.01 (0.88)
HSE-GLUP	LE	3.12 (0.98)	3.04 (0.97)	3.09 (0.95)
HSD-GLU	LE	3.10 (0.79)	3.12 (0.88)	3.14 (0.98)
HSE-GLU	LE	2.89 (0.36)	2.90 (0.54)	3.04 (0.94)
HSP-GLUP	LE	2.96 (0.49)	2.94 (0.64)	3.05 (0.84)

Table 5.19: TD-DFT excitation energies (eV) of the two lowest states of protein-bound neutral chromophore in the OFF-state; oscillator strength is shown in parentheses.

System	State	Extended QM cc-pVDZ QM only	Extended QM aug-cc-pVDZ QM only	Large QM aug-cc-pVDZ QM only
HSD-GLUP	LE	4.03 (0.19)	3.86 (0.41)	3.80 (0.58)
HSE-GLUP	LE	3.97 (0.63)	3.92 (0.33)	3.93 (0.56)
HSD-GLU	LE	4.07 (0.11)	3.69 (0.53)	3.70 (0.49)
HSE-GLU	LE	3.57 (0.56)	3.52 (0.57)	3.58 (0.54)
HSP-GLU	LE	3.89 (0.58)	3.85 (0.60)	3.83 (0.52)
HSD-GLUP2	LE	3.58 (0.64)	3.52 (0.64)	3.57 (0.60)
HSE-GLUP2	LE	3.95 (0.56)	3.87 (0.58)	3.83 (0.52)

Table 5.20: Excitation energies (eV) of the two lowest states of protein-bound neutral chromophore in the ON-state; oscillator strength is shown in parentheses. Extended QM.

System	State	TD-DFT aug-cc-pVDZ QM only	SOS-CIS(D) aug-cc-pVDZ QM only	XMCQDPT2 cc-pVDZ QM only	XMCQDPT2[a] cc-pVDZ QM only	XMCQDPT2 aug-cc-pVDZ /cc-pVDZ QM only
HSD-GLUP	LE	3.39 (0.63)	2.87 (1.04)	2.62 (0.50)		2.59 (0.59)
	CT	2.94 (0.22)	3.40 (0.08)	2.98 (0.12)		3.04 (0.17)
HSE-GLUP	LE	3.32 (0.82)	2.87 (0.99)	2.64 (0.57)	2.88 (0.31)	2.79 (0.41)
	CT	2.76 (0.05)	3.17 (0.13)	2.89 (0.23)	3.09 (0.05)	2.51 (0.28)
HSD-GLU	LE	3.41 (0.66)	3.10 (0.97)	2.89 (0.72)	2.85 (0.40)	2.83 (0.24)
	CT	3.10 (0.13)	3.68 (0.04)	3.01 (0.03)	3.12 (0.03)	2.76 (0.28)
HSE-GLU	LE	3.37 (0.77)	2.98 (0.94)	2.70 (0.30)	2.74 (0.56)	2.87 (0.45)
	CT	2.81 (0.04)	3.21 (0.10)	2.98 (0.29)	3.08 (0.01)	2.59 (0.18)
HSP-GLU	LE	3.56 (0.59)	3.42 (0.93)	3.02 (0.14)	2.90 (0.16)	2.96 (0.26)
	CT	3.26 (0.14)	4.05 (0.04)	3.10 (0.15)	2.93 (0.02)	2.94 (0.11)

[a] Using structures and QM definition from the old protocol[20].

Table 5.21: Excitation energies of the protein-bound anionic chromophore in the ON-state; oscillator strength is shown in parentheses. Extended QM.

System	state	TD-DFT aug-cc-pVDZ QM only	SOS-CIS(D) aug-cc-pVDZ QM only	XMCQDPT2 cc-pVDZ QM only	XMCQDPT2[a] cc-pVDZ QM-only	XMCQDPT2 aug-cc-pVDZ /cc-pVDZ QM-only
HSD-GLUP	LE	2.95 (0.86)	2.41 (1.29)	2.30 (0.94)	2.39 (0.89)	2.16 (0.87)
HSE-GLUP	LE	3.04 (0.97)	2.68 (1.41)	2.58 (0.96)	2.37 (0.93)	2.39 (0.98)
HSD-GLU	LE	3.12 (0.88)	2.64 (1.34)	2.50 (0.93)		2.37 (0.92)
HSE-GLU	LE	2.90 (0.54)	2.41 (1.29)	2.41 (0.90)		2.29 (0.92)
HSP-GLUP	LE	2.94 (0.64)	2.43 (1.23)	2.39 (0.84)	2.41 (0.85)	2.27 (0.81)

[a] Using structures and QM definition from the old protocol[20].

Table 5.22: Excitation energies of the protein-bound neutral chromophore in the OFF-state; oscillator strength is shown in parentheses. Extended QM.

System	TDDFT aug-cc-pVDZ QM only	SOS-CIS(D) aug-cc-pVDZ QM only	XMCQDPT2 cc-pVDZ QM only	XMCQDPT2[a] cc-pVDZ QM-only	XMCQDPT2 aug-cc-pVDZ /cc-pVDZ QM-only
HSD-GLUP	3.86 (0.41)	3.98 (0.83)	4.06 (0.52)	3.79 (0.21)	3.50 (0.63)
HSE-GLUP	3.92 (0.33)	4.17 (0.79)	4.11 (0.53)	4.00 (0.61)	3.67 (0.58)
HSD-GLU	3.69 (0.53)	3.93 (0.76)	3.51 (0.58)	3.99 (0.66)	3.35 (0.55)
HSE-GLU	3.52 (0.57)	3.62 (0.80)	3.30 (0.63)	3.97 (0.43)	3.05 (0.55)
HSP-GLU	3.85 (0.60)	4.11 (0.72)	3.60 (0.55)	3.94 (0.60)	3.46 (0.60)
HSD-GLUP2	3.52 (0.64)	3.53 (0.87)	3.34 (0.52)		3.14 (0.52)
HSE-GLUP2	3.87 (0.58)	4.06 (0.78)	3.92 (0.21)		3.59 (0.54)

[a] Using structures and QM definition from the old protocol[20].

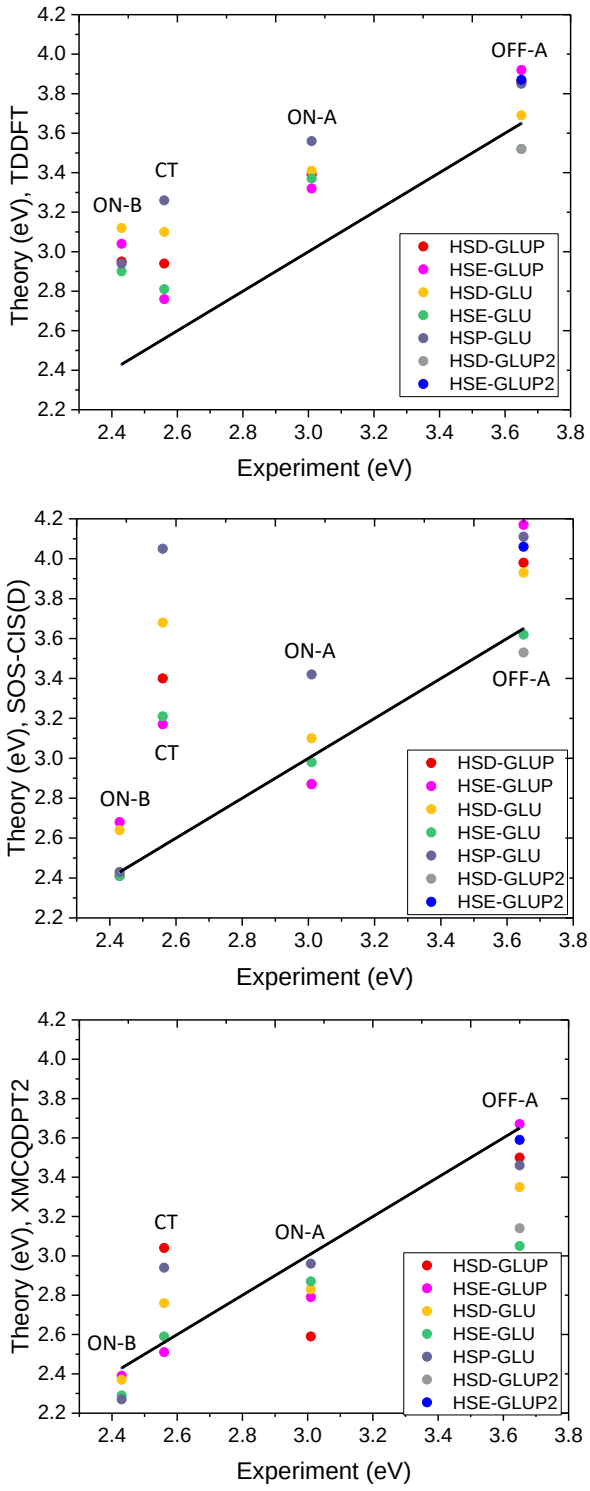


Figure 5.23: Excitation energies for different model systems shown against the experimental values. Top: TD-DFT/aug-cc-pVDZ; middle: SOS-CIS(D)/aug-cc-pVDZ; bottom: XMCQDPT2/aug-cc-pVDZ/cc-pVDZ. Extended QM.

5.10 Appenidx F: Structures of possible intermediates

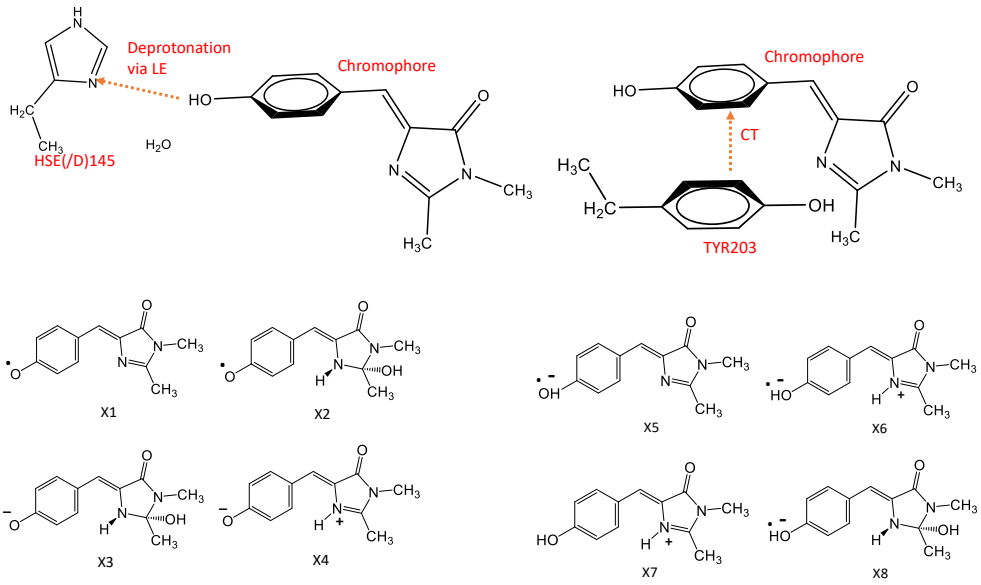


Figure 5.24: Two possible initial steps for Dreiklang photoconversion. Ref. 10 proposed that the photoconversion begins by ESPT (left), forming anionic chromophore, which undergoes further transformation. Following this route, one can consider structures X1-X4 as possible candidates for reaction intermediate X. We propose an alternative mechanism via CT state (right). Following this route, one can consider structures X5-X8 as possible candidates for reaction intermediate X.

We considered several structures of the intermediates. Fig. 5.24 shows 2 possible scenarios for initiating photoconversion. Ref.[10] proposed that the photoconversion begins by ESPT, forming anionic chromophore, which undergoes further transformation. Following this route, one can consider structures X1-X4 as possible candidates for reaction intermediate X. As explained in the main text, there are several major objections to this mechanism. We propose an alternative mechanism via CT state. Following this route, one can consider structures X5-X8 as possible candidates for reaction intermediate X.

Intermediate X5 corresponds to the chromophore in the CT state ($Chro^{\cdot -}$). Intermediate X5 is the result of proton transfer to $Chro^{\cdot -}$, forming neutral radical. X7 is the result of the

protonated chromophore after back-transfer of the electron. X8 is the result of the hydrated chromophore which still has the extra electron.

5.11 Appendix G: Optimization and AIMD simulations: Additional results

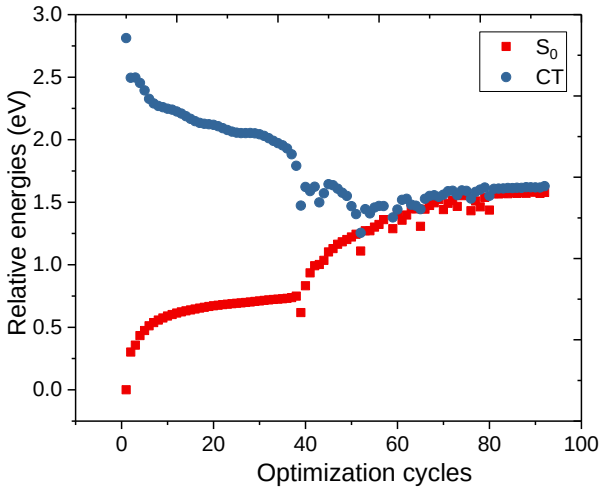


Figure 5.25: Energies of the Kohn-Sham reference state (S_0) and CT state along optimization path (on-A-HSE-GLUP structure).

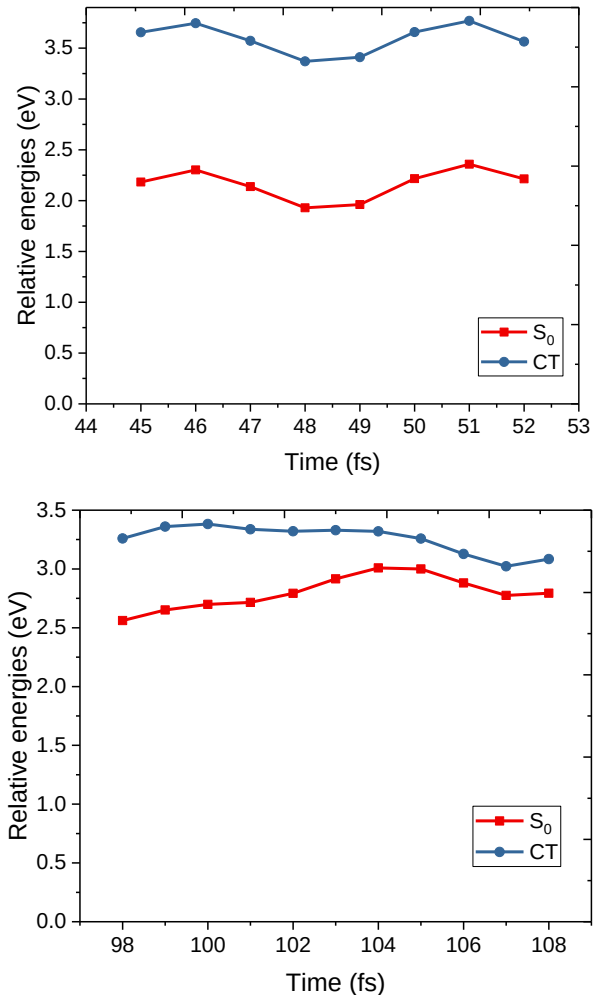


Figure 5.26: Ground and excited state during the first two steps of the reaction in CT state (on-A-HSE-GLUP structure). Left: 1st step — proton abstraction by chromophore's N from protonated Glu222. Right: 2nd step — proton transfer from Tyr203 to deprotonated Glu222.

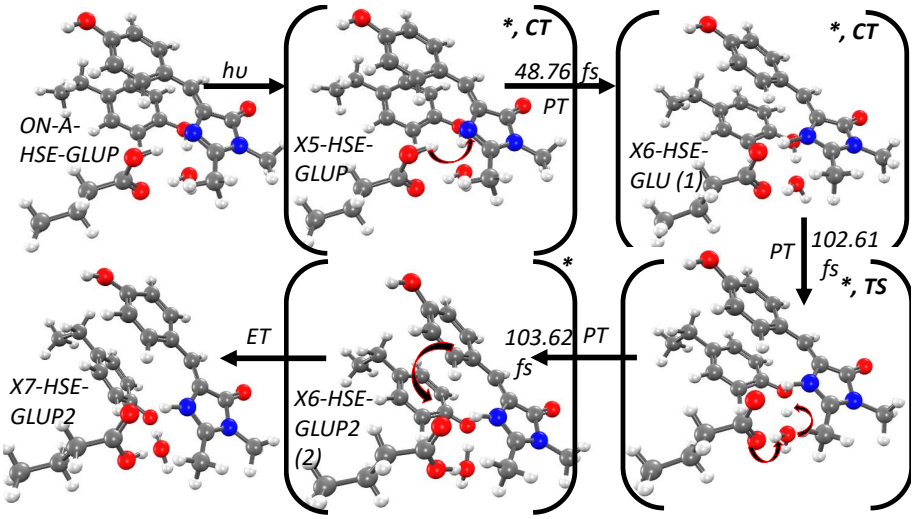


Figure 5.27: Analysis of the AIMD trajectory on the CT state (on-A-HSE-GLUP structure).

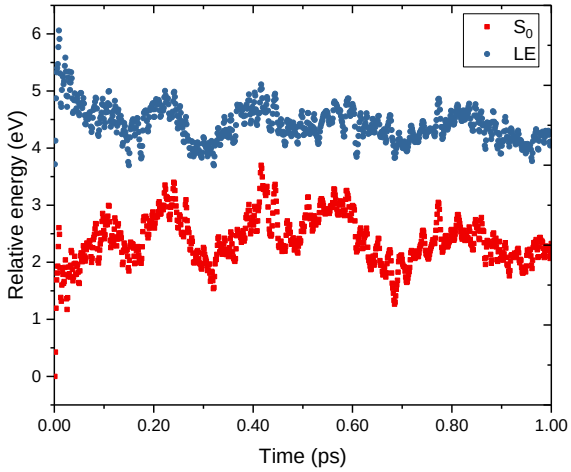


Figure 5.28: Energies of the Kohn-Sham reference state (S_0) and the LE state (2nd TD-DFT state) along the AIMD trajectory on the LE potential energy surface.

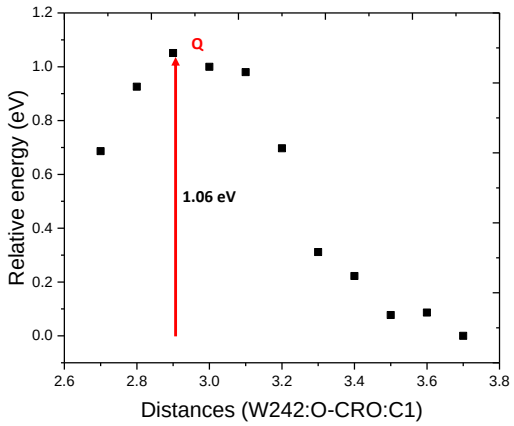


Figure 5.29: Relaxed energy profile on the ground state surface (starting from X7 intermediate) along hydration reaction coordinate defined as W242:O-CRO:C1 distance. Zero energy corresponds to the energy of the reference state of the structure at t=248 fs, roughly corresponding to X7. ONIOM, ωB97X-D/aug-cc-pVDZ/CHARMM27.

Table 5.23: Average excitation energies (eV) of the two lowest states of protein-bound neutral chromophore in the ON-state computed using structures from 21 MD snapshots; oscillator strength is shown in parentheses. Large QM.

System	State	TDDFT aug-cc-pVDZ QM/MM (MD)
HSD-GLUP	LE	3.49 (0.56)
	CT	4.01 (0.05)
HSE-GLUP	LE	3.58 (0.52)
	CT	3.96 (0.10)
HSD-GLU	LE	
	CT	
HSE-GLU	LE	
	CT	
HSP-GLU	LE	3.52 (0.49)
	CT	3.72 (0.02)

Table 5.24: Average excitation energies (eV) of the two lowest states of protein-bound anionic chromophore in the ON-state computed using structures from 21 MD snapshots; oscillator strength is shown in parentheses. Large QM.

System	State	TDDFT aug-cc-pVDZ QM/MM (MD)
HSD-GLUP	LE	2.99 (0.85)
HSE-GLUP	LE	
HSD-GLU	LE	
HSE-GLU	LE	
HSP-GLUP	LE	2.98 (0.86)

Table 5.25: Average excitation energies (eV) of the two lowest states of protein-bound hydrated chromophore in the OFF-state computed using structures from 21 MD snapshots; oscillator strength is shown in parentheses. Large QM.

System	State	TDDFT aug-cc-pVDZ QM/MM (MD)
HSD-GLUP	LE	
HSE-GLUP	LE	3.90 (0.85)
HSD-GLU	LE	3.75 (0.52)
HSE-GLU	LE	
HSP-GLUP	LE	3.77 (0.51)
HSD-GLUP2	LE	
HSE-GLUP2	LE	

5.12 Chapter 5 References

[1] A. Acharya, A. M. Bogdanov, K. B. Bravaya, B. L. Grigorenko, A. V. Nemukhin, K. A. Lukyanov, and A. I. Krylov, Photoinduced chemistry in fluorescent proteins: Curse or blessing?, Chem. Rev. **117**, 758 (2017).

[2] K. Nienhaus and G.U. Nienhaus, Fluorescent proteins for live-cell imaging with super-resolution, Chem. Soc. Rev. **43**, 1088 (2014).

[3] H.C. Ishikawa-Ankerhold, R. Ankerhold, and G.P.C. Drummen, Advanced fluorescent microscopy techniques — FRAP, FLIP, FLAP, FRET and FLIM, Molecules **17**, 4047 (2012).

[4] R. Ando, H. Mizuno, and A. Miyawaki, Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting, Science **306**, 1370 (2004).

[5] M. Andresen, A.C. Stiel, J. Fölling, D. Wenzel, A. Schönle, A. Egner, *et al.*, Photoswitchable fluorescent proteins enable monochromatic multilabel imaging and dual color fluorescence nanoscopy, Nat. Biotechnol. **26**, 1035 (2008).

[6] D.M. Chudakov, A.V. Feofanov, N.N. Mudrik, S. Lukyanov, and K.A. Lukyanov, Chromophore environment provides clue to "kindling fluorescent protein" riddle, J. Biol. Chem. **278**, 7215 (2003).

[7] T. Brakemann, A.C. Stiel, G. Weber, M. Andresen, I. Testa, T. Grotjohann, M. Leutenegger, U. Plessmann, H. Urlaub, C. Eggeling, M.C. Wahl, S.W. Hell, and S. Jakobs, A reversibly photoswitchable GFP-like protein with fluorescence excitation decoupled from switching, Nat. Biotechnol. **29**, 942 (2011).

[8] R. M. Wachter, M. A. Elsliger, K. Kallio, G. T. Hanson, and S. J. Remington, Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein, Structure **6**, 1267 (1999).

[9] M. Ormö, A.B. Cubitt, K. Kallio, L.A. Gross, R.Y. Tsien, and S.J. Remington, Crystal structure of the aequorea victoria green fluorescent protein, Science **273**, 1392 (1996).

[10] F. Lacombat, P. Plaza, M.-A. Plamont, and A. Espagne, Photoinduced chromophore hydration in the fluorescent protein Dreiklang is triggered by ultrafast excited-state proton transfer coupled to a low-frequency vibration, J. Phys. Chem. Lett. **8**, 1489 (2017).

[11] R.Y. Tsien, The green fluorescent protein, Annu. Rev. Biochem. **67**, 509 (1998).

[12] B. L. Grigorenko, A. V. Nemukhin, I. V. Polyakov, D. I. Morozov, and A. I. Krylov, First-principle characterization of the energy landscape and optical spectra of the green fluorescent protein along A-I-B proton transfer route, J. Am. Chem. Soc. **135**, 11541 (2013).

[13] S.R. Meech, Excited state reactions in fluorescent proteins, Chem. Soc. Rev. **38**, 2922 (2009).

[14] J.J. van Thor, Photoreactions and dynamics of the green fluorecent protein, Chem. Soc. Rev. **38**, 2935 (2009).

[15] M. Chattoraj, B.A. King, G.U. Bublitz, and S.G. Boxer, Ultra-fast excited state dynamics in green fluorescent protein: Multiple states and proton transfer, Proc. Nat. Acad. Sci. **93**, 8362 (1996).

[16] K. Brejc, T.K. Sixma, P.A. Kitts, S.R. Kain, R.Y. Tsien, M. Ormö, and S.J. Remington, Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein, Proc. Nat. Acad. Sci. **94**, 2306 (1997).

[17] S.J. Remington, Green fluorescent protein: A perspective, Protein Sci. **20**, 1509 (2011).

[18] R. Simkovitch, A. Huppert, D. Huppert, S.J. Remington, and Y. Miller, Proton transfer in wild-type GFP and S205V mutant is reduced by conformational changes of residues in the proton wire, J. Phys. Chem. B **117**, 11921 (2013).

[19] B. L. Grigorenko, A. I. Krylov, and A. V. Nemukhin, Molecular modeling clarifies the mechanism of chromophore maturation in the green fluorescent protein, J. Am. Chem. Soc. **139**, 10239 (2017).

[20] B. L. Grigorenko, I. Polyakov, A. I. Krylov, and A. V. Nemukhin, Computational modeling reveals the mechanism of fluorescent state recovery in the reversibly photoswitchable protein Dreiklang, J. Phys. Chem. B **123**, 8901 (2019).

[21] M. H. M. Olsson, C. R. Sondergaard, M. Rostkowski, and J. H. Jensen, PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions, J. Chem. Theory Comput. **7**, 525 (2011).

[22] N. Foloppe and A. D. MacKerell, All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data, J. Comput. Chem. **21**, 86 (2000).

[23] N. Reuter, H. Lin, and W. Thiel, Green fluorescent proteins: Empirical force field for the neutral and deprotonated forms of the chromophore. Molecular dynamics simulations of the wild type and S65T mutant, J. Phys. Chem. B **106**, 6310 (2002).

[24] F. Weinhold and C. R. Landis, Natural bond orbitals and extensions of localized bonding concepts, Chem. Ed.: Res. & Pract. Eur. **2**, 91 (2001).

[25] J.-D. Chai and M. Head-Gordon, Systematic optimization of long-range corrected hybrid density functionals, J. Chem. Phys. **128**, 084106 (2008).

[26] J.-D. Chai and M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom-atom dispersion interactions, Phys. Chem. Chem. Phys. **10**, 6615 (2008).

[27] Y. Zhang and W. Yang, A challenge for density functionals: Self-interaction error increases for systems with a noninteger number of electrons, J. Chem. Phys. **109**, 2604 (1998).

[28] V. Polo, E. Kraka, and D. Cremer, Electron correlation and the self-interaction error of density functional theory, Mol. Phys. **100**, 1771 (2002).

[29] M. Lundber and P.E.M. Siegbahn, Quantifying the effects of the self-interaction error in DFT: When do the delocalized states appear?, J. Chem. Phys. **122**, 224103 (2005).

[30] A. Dreuw, J.L. Weisman, and M. Head-Gordon, Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange, J. Chem. Phys. **119**, 2943 (2003).

[31] S. Grimme, Accurate description of van der Waals complexes by density functional theory including empirical corrections, J. Comput. Chem. **25**, 1463 (2004).

[32] Y. M. Rhee and M. Head-Gordon, Scaled second order perturbation corrections to configuration interaction singles: efficient and reliable excitation energy methods, J. Phys. Chem. A **111**, 5314 (2007).

[33] A. I. Krylov, Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: The hitchhiker's guide to Fock space, Annu. Rev. Phys. Chem. **59**, 433 (2008).

[34] A.A. Granovsky, Extended multi-configuration quasi-degenerate perturbation theory: The new approach to multi-state multi-reference perturbation theory, J. Chem. Phys. **134**, 214113 (2011).

[35] K. B. Bravaya, M. G. Khrenova, B. L. Grigorenko, A. V. Nemukhin, and A. I. Krylov, Effect of protein environment on electronically excited and ionized states of the green fluorescent protein chromophore, J. Phys. Chem. B **115**, 8296 (2011).

[36] K. B. Bravaya, O. M. Subach, N. Korovina, V. V. Verkhusha, and A. I. Krylov, Insight into the common mechanism of the chromophore formation in the red fluorescent proteins: The elusive blue intermediate revealed, J. Am. Chem. Soc. **134**, 2807 (2012).

[37] A. M. Bogdanov, A. Acharya, A. V. Titelmayer, A. V. Mamontova, K. B. Bravaya, A. B. Kolomeisky, K. A. Lukyanov, and A. I. Krylov, Turning on and off photoinduced electron transfer in fluorescent proteins by π-stacking, halide binding, and Tyr145 mutations, J. Am. Chem. Soc. **138**, 4807 (2016).

[38] T. Sen, A. V. Mamontova, A. V. Titelmayer, A. M. Shakhov, A. A. Astafiev, A. Acharya, K. A. Lukyanov, A. I. Krylov, and A. M. Bogdanov, Influence of the first chromophore-forming residue on photobleaching and oxidative photoconversion of EGFP and EYFP, Int. J. Mol. Sci. **20**, 5229 (2019).

[39] Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A.T.B.; Wormit, M.; Kussmann, J.; Lange, A.W.; Behn, A.; Deng, J.; Feng, X., et al., Advances in molecular quantum chemistry contained in the Q-Chem 4 program package, Mol. Phys. **113**, 184 (2015).

[40] A. I. Krylov and P. M. W. Gill, Q-Chem: An engine for innovation, WIREs: Comput. Mol. Sci. **3**, 317 (2013).

[41] A. A. Granovsky, PC GAMESS/Firefly, http://classic.chem.msu.su/gran/gamess.

[42] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. **26**, 1781 (2005).

[43] F. Plasser, M. Wormit, and A. Dreuw, New tools for the systematic analysis and visualization of electronic excitations. I. Formalism, J. Chem. Phys. **141**, 024106 (2014).

[44] K. Takaba, Y. Tai, H. Eki, H.-A. Dao, Y. Hanazono, K. Hasegawa, K. Mikia, and K. Takeda, Subatomic resolution x-ray structures of green fluorescent protein, IUCrJ **6**, 387 (2019).

[45] A. V. Luzanov, A. A. Sukhorukov, and V. E. Umanskii, Application of transition density matrix for analysis of excited states, Theor. Exp. Chem. **10**, 354 (1976), Russian original: Teor. Eksp. Khim., 10, 456 (1974).

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

[47] E. Epifanovsky, I. Polyakov, B. L. Grigorenko, A. V. Nemukhin, and A. I. Krylov, Quantum chemical benchmark studies of the electronic properties of the green fluorescent protein chromophore: I. Electronically excited and ionized states of the anionic chromophore in the gas phase, J. Chem. Theory Comput. **5**, 1895 (2009).

[48] A. V. Nemukhin, B. L. Grigorenko, M. Khrenova, and A. I. Krylov, Computational challenges in modeling of representative bioimaging proteins: GFP-like proteins, flavoproteins, and phytochromes, J. Phys. Chem. B **123**, 6133 (2019).

[49] A. I. Krylov, From orbitals to observables and back, J. Chem. Phys. **153**, 080901 (2020).

[50] P. Kimber and F. Plasser, Toward an understanding of electronic excitation energies beyond the molecular orbital picture, Phys. Chem. Chem. Phys. **22**, 6058 (2020).

[51] R. B. Vegh, K. B. Bravaya, D. A. Bloch, A. S. Bommarius, L. M. Tolbert, M. Verkhovsky, A. I. Krylov, and K. M. Solntsev, Chromophore photoreduction in red fluorescent proteins is responsible for bleaching and phototoxicity, J. Phys. Chem. B **118**, 4527 (2014).

[52] B. L. Grigorenko, A. V. Nemukhin, D. I. Morozov, I. V. Polyakov, K. B. Bravaya, and A. I. Krylov, Toward molecular-level characterization of photo-induced decarboxylation of the green fluorescent protein: Accessibility of the charge-transfer states, J. Chem. Theory Comput. **8**, 1912 (2012).

[53] B. L. Grigorenko, A. V. Nemukhin, I. V. Polyakov, M. G. Khrenova, and A. I. Krylov, A light-induced reaction with oxygen leads to chromophore decomposition and irreversible photobleaching in GFP-type proteins, J. Phys. Chem. B **119**, 5444 (2015).

[54] P. D. Dahlberg, A. M. Sartor, J. Wang, S. Saurabh, L. Shapiro, and W. E. Moerner, Identification of PAmKate as a red photoactivatable fluorescent protein for cryogenic super-resolution imaging, J. Am. Chem. Soc. **140**, 12310 (2018).

[55] P. D. Dahlberg, A. Saurabh, A. M. Sartor, J. Wang, P. G. Mitchell, W. Chiu, L. Shapiro, and W. E. Moerner, Cryogenic single-molecule fluorescence annotations for electron tomography reveal in situ organization of key proteins in Caulobacter, Proc. Nat. Acad. Sci. **117**, 13937 (2020).

@articlecase2004amber, title=AMBER, ver. 8, author=Case, DA and Darden, TA and Cheatham III, TE and Simmerling, CL and Wang, J and Duke, RE and Luo, R and Merz, KM and Wang, B and Pearlman, DA and others, journal=University of California at San Francisco, San Francisco, CA, year=2004

Chapter 6: BrUSLEE and his shadow: Two persistent excited-state populations within a GFP mutant

6.1 Introduction

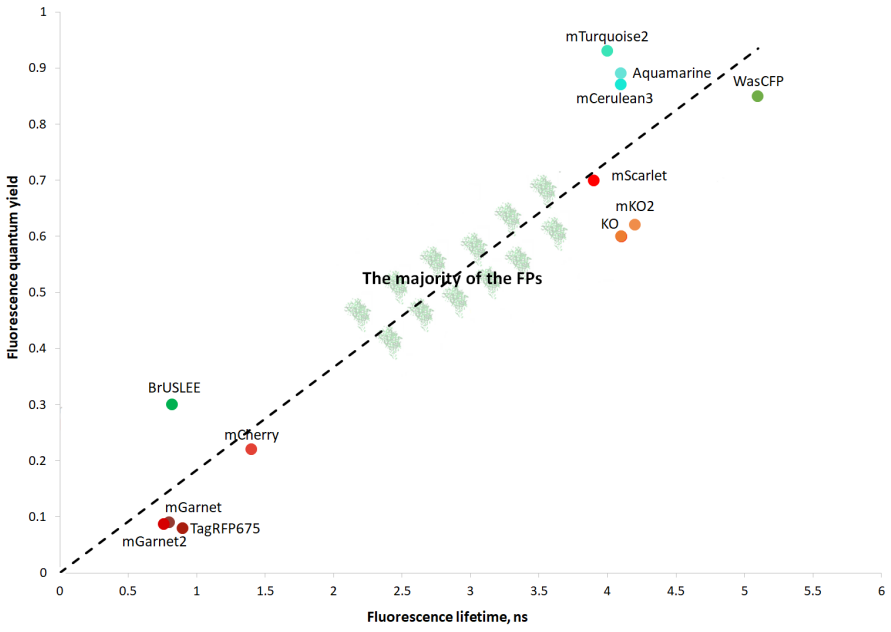


Figure 6.1: Fluorescence quantum yield versus fluorescence lifetime for selected FPs.

In 1994, the green fluorescent protein (avGFP) from *Aequorea Victoria* jellyfish was used to implement a genetically encoded fluorescent label for *in vivo* imaging[1]. The unique structure of

the chromophore formed by the protein's own amino acid residues, the possibility of targeted labeling inside a living cell, low toxicity, relative ease of use, and the ability of tuning its properties by genetic engineering have made fluorescent proteins (FPs) an essential molecular tool for biological imaging[2–4]. Biomedical research often requires to monitor multiple macromolecules or subcellular structures, and to record signal of the fluorescent indicators. The efficiency is limited mainly by the spectral properties of fluorophores. Thus, hundreds of the probes of different origins described to date (including dozens of FP variants) can provide a reliable simultaneous detection in 3-5 spectral channels only[5]. Namely, bright and photostable fluorophores are widespread in green-yellow range but are relatively rare in the blue and far-red parts of the spectrum[2,6–8].

The design of FPs with properties matching particular applications requires understanding of how the structure of the protein relates to its photophysical and photochemical properties. Despite intense research efforts aiming to unravel fundamental aspects of the FP photocycle[9], many questions remain unanswered, including structural determinants of fluorescence lifetime and quantum yield and the limits of their tunability.

In the most basic case of a single emissive state, the population of excited fluorophores (Chro*) decays via two competing first-order processes[10,11]:

$$Chro^* \xrightarrow{k_r} Chro + h\nu, \tag{6.1}$$

$$Chro^* \xrightarrow{k_{nr}} Chro, \tag{6.2}$$

where k_r is the radiative (intrinsic fluorescence) rate constant and k_{nr} describes all quenching channels. The overall decay of the excited-state chromophore is also described by the first-order kinetics with $k = k_r + k_{nr}$ and the corresponding apparent (measured) fluorescence lifetime $\tau = \frac{\ln(2)}{k}$. If non-radiative channels are much slower than the radiative rate ($k_{nr} \ll k_r$), then the apparent excited-state lifetime corresponds to the intrinsic fluorescence lifetime

($\tau_r = \frac{\ln(2)}{k_r}$). It is expected that at cryogenic conditions, when various quenching channels are suppressed, the apparent lifetime represents the intrinsic fluorescence lifetime. In contrast, if the radiationless decay is fast ($k_{nr} > k_r$), then the apparent lifetime τ reflects the kinetics of the radiationless decay and is shorter than τ_r. If several distinct populations of fluorophores are present, the fluorescence kinetics becomes multi-exponential and the above simple relationships between rate constants and lifetimes are no longer true. Multi-exponential fluorescence decay (spectral heterogeneity) arises due to structural heterogeneity, such as different conformations or protonation states of fluorophores, or different local environments.

The fluorescence quantum yield (FQY) is determined by the competition between radiative and radionionless decay:

$$FQY = \frac{k_r}{k_r + k_{nr}} = \frac{\tau}{\tau_r} = \frac{1}{1 + \frac{\tau_r}{\tau_{nr}}}, \tag{6.3}$$

where τ_r denotes the intrinsic fluorescence lifetime and $\tau_{nr} = \frac{\ln(2)}{k_{nr}}$ represents a timescale associated with non-radiative decay. Thus, large FQY can be attained by either suppressing non-radiative decay rates (i.e., increasing τ_{nr}) or increasing the radiative decay rate (i.e., decreasing τ_r). The intrinsic radiative lifetime is inversely proportional[39] to the oscillator strength of the transition (f_l) and to the square of corresponding transition energy (E). In atomic units

$$\frac{1}{\tau_r} = \frac{E^2 f_l}{2\pi (c')^3 \epsilon} \tag{6.4}$$

where c' is the speed of light in the medium and ϵ is the dielectric constant. The extinction coefficient (EC) is proportional to the oscillator strength of the transition.

The radiationless decay constant represents the sum of all non-radiative excited-state decay channels. It can vary widely among different fluorophores and is strongly affected by fluorophore's environment. In contrast, the radiative decay constant is an intrinsic property of the

fluorophore and, therefore, is expected to be the same for chemically identical chomophores. Conversely, for FPs with the same type of chromophores, the variations in FQY can be attributed to the variations in non-radiative decay rates, so that the correlation between the apparent fluorescence lifetime τ and FQY is expected.

This conjecture can only be tested by direct measurements of k_r for hundreds of FPs. However, experiments are usually carried out at room temperature and, therefore, reflect the apparent fluorescence lifetimes. FPs with self-maturing chromophores exhibit a broad range of the apparent fluorescence lifetimes—some members having much longer (3.9-5.1 ns) and some having much shorter (0.76-1.6 ns) lifetimes than the average[6]. Fig. 6.1 shows FQYs against fluorescence lifetimes for a variety of FPs. Some FPs exhibit a good correlation between τ and FQY, combining either a long lifetime with a high quantum yield or a short lifetime with a low-to-moderate quantum yield. The first group comprises mCerulean3 (τ=4.1 ns/FQY=0.87)[13], Aquamarine (τ=4.1 ns/FQY=0.89)[14], mTurquoise2 (τ=4.0 ns/FQY=0.93)[15], mScarlet (τ=3.9 ns/FQY=0.7)[16], and WasCFP (τ=5.1 ns/FQY 0.85)[17]. The second group contains mCherry (τ=1.4 ns/FQY=0.22), TagRFP675 (τ=0.9 ns/FQY=0.08)[18], mGarnet (τ=0.8 ns/FQY=0.09)[19], mGarnet2 (τ=0.76 ns/FQY=0.087)[20]. There is a also third group—deviants for which the correspondence between τ and FQY is less pronounced. For example, in the orange KO and mKO proteins, the impressive 4.1-4.2 ns lifetimes go together with moderate FQYs[21,22]. Green BrUSLEE, a new FP introduced here, also belongs to this group, featuring short lifetime (τ=0.8 ns) and a moderate FQY of 0.12-0.3.

Because the radiative lifetime is inversely proportional to the oscillator strength (as per Eq. (6.4)), it is expected that large EC (i.e, large oscillator strength) would result in the decrease of the $\frac{\tau_r}{\tau_{nr}}$ term in Eq. (6.3), giving rise to an increased FQY. This is generally the case – brighter FPs often have larger FQY. However, there are interesting exceptions. For example, some of the proteins listed above (for example, cyan mCerulean, Aquamarine, and mTurquoise2), feature

relatively low EC and high FQY/long τ, and in the pair of orange KO/mKO, an increase in FQY is accompanied by a decrease in EC. These exceptions can be rationalized by assuming that in these FPs changes in τ_r are compensated by changes in τ_{nr}. Hence, the process of tuning up optical properties of FPs requires simultaneous optimization of oscillator strength (which defines brightness and fluorescence lifetime) and non-radiative decay rates. Because the former is largely the property of the chromophore and the latter largely depends on the interactions of the chromophore with its immediate environment, it should be possible to tune them independently.

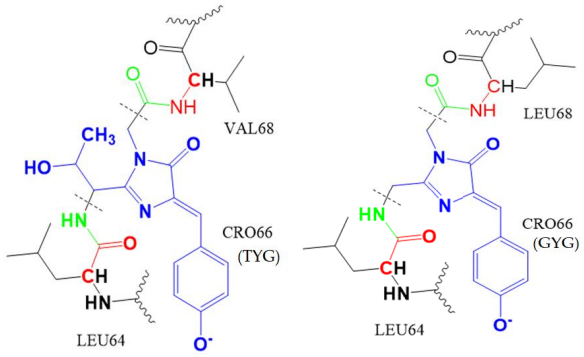


Figure 6.2: Structure of the chromophore in EGFP (left) and the 3 mutants studied in this Chapter (right). In EGFP, the chromophore is formed by the threonine-tyrosine-glycine (TYG) triad whereas in T65G mutants the chromophore is formed by the glycine-tyrosine-glycine (GYG) triad. The conjugated core of both chromophores is the same, but the TYG chromophore has additional electron-donating group. The twisting motion is described by dihedral angles ϕ (phenolate flip around the single bond) and τ (imidozalinone flip around the double bond); see Fig. 6.12.

Photophysical properties of the fluorescent proteins are determined by an interplay between chromophore's intrinsic electronic structure, its interactions with the surrounding residues, and several competing excited-state processes[9,23]. Oscillator strength, which is the key determinant of EC, depends on the transition dipole moment, and is affected by the size of the conjugated π-system. Electron donating groups attached to the chromophore can lead to an increased f_l. Chromophore twisting disrupts conjugation and reduces f_l; hence, deviations from planarity

are expected to lead to dimmer FPs. One interesting feature of the GFP-type chromophore (shown in Fig. 6.2) is that it is rigid in the ground state (torsional barriers along ϕ and τ of around 30 kcal/mol), but becomes rather floppy in the excited state (torsional barrier drops to $\sim$3 kcal/mol). Because of this flexibility, the bare chromophore is non-fluorescent—twisting motion leads to an effective radiationless decay. Only when constrained by the protein environment (or another matrix), which prevent it from twisting, the chromophore becomes fluorescent. Hence, hydrogen-bond network around the protein-bound chromophore has a major effect on its excited-state dynamics and fluorescent properties. Because the rigidity of the chromophore in the ground state, the changes in hydrogen-bond network due to mutations do not necessarily lead to prominent structural changes (i.e., the chromophore remains planar in the course of thermal motions), but can have a profound effect on the excited-state dynamics and, consequently, the non-radiative decay rate.

In this Chapter we introduce the BrUSLEE protein and investigate mechanistic details of its photophysical properties. BrUSLEE is a descendant of the popular EGFP[24] and differs from it by 3 mutations: T65G/Y145M/F165Y. These mutations were inspired by the previous study[25], which identified the involvement of the respective residues in photoinduced electron transfer ultimately leading to photobleaching. BrUSLEE—BRight Ultimately Shorttime Enhanced Emitter—demonstrates an unusual combination of high fluorescence brightness and short lifetime, which prompted us to investigate structural determinants of its photophysical properties by time-resolved fluorescence measurements and atomistic simulations. We also considered the T65G and T65G/Y145M mutants. Below we often refer to the double mutant (T65G/Y145M) as Duo and to the triple mutant (T65G/Y145M/F165Y; BrUSLEE) as Trio. In addition to an unusual FQY/τ combination, BrUSLEE also shows multi-exponential fluorescence decay, revealing 2 distinct subpopulations, co-existing in a wide temperature range

(4-300 K). The fluorescence lifetimes of these emissive states change considerably with temperature, converging to low temperature limits that are vastly different from each other and from that of the parental EGFP. As discussed below, crystal structure and 15N-NMR spectroscopy of BrUSLEE show no obvious structural heterogeneity. Atomistic simulations suggested that the heterogeneity arises due to co-existing populations of different protonation states of the chromophore-adjacent titratable residues. In particular, different protonation states of His148 alter the hydrogen-bonding network around the chromophore, affecting significantly effect on its twisting flexibility in the excited state. Simulations also explain trends in τ and FQY by the changes in the electronic properties of the chromophore and hydrogen-bond network around it due to mutations. In particular, the T65G mutation[26] increases conformational flexibility of the chromophore in the excited state, leading to faster τ_{nr}; at the same time it increases the oscillator strength of the transition, leading to shorter τ_r. Consequently, despite the reduction in excited-state lifetime, relatively large FQY is observed.

6.2 Results and discussion

6.2.1 Structure analysis

Fig. 6.3 shows superimposed x-ray structures of EGFP and BrUSLEE[45,46] and Fig. 6.4 compares hydrogen-bond network around the chromophore. MD simulations (discussed in the Appendix A1 and below) yield average structural parameters that agree well with the crystal structures and provide additional insight into the thermal range of motion of the chromophore and the key residues; the simulations also provide structural data for the T65G and Duo mutants. In addition to comparison of the structures, we also studied FRET between the Tryptophan and the chromophore in all 4 systems. By comparing the results with the experimental FRET

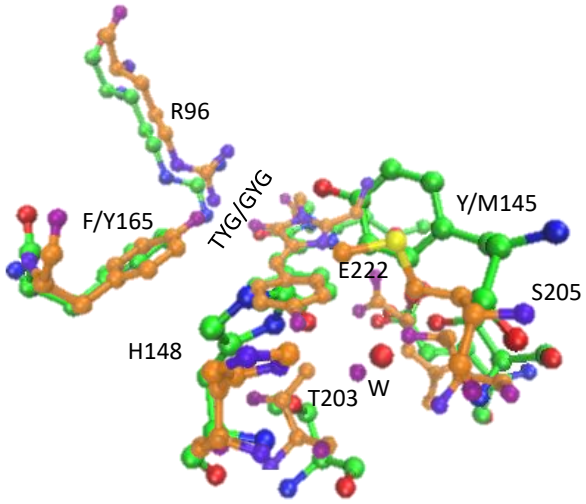


Figure 6.3: Superimposed crystal structures of EGFP (green) and BrUSLEE (orange), with the chromophore's center of mass set at the origin.

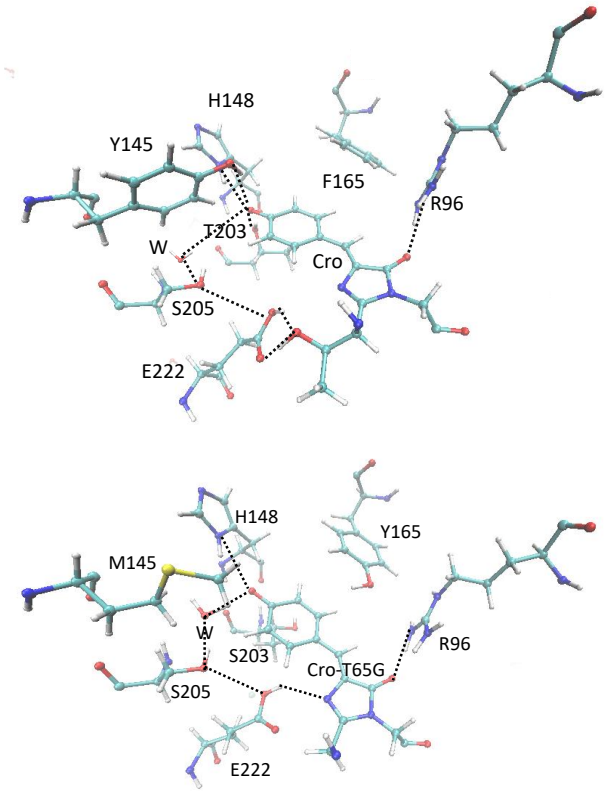


Figure 6.4: Top: Hydrogen-bond network around the chromophore in EGFP and Bottom: BrUSLEE.

measurements we can further validate the simulations. FRET experiments and simulations are discussed in the Appendix C.

The mutations affect both the chromophore is structure and its interactions with the nearby residues. As discussed in our previous work in Chapter 4, T65G mutation has significant effect on the structure and results in weakening of the hydrogen-bond network around the chromophore. In the mutants, there is no hydrogen bond between Glu222 and Thr65 (since it is substituted by Gly); instead a new bond between Glu222 and N-imidazoline is formed (see Fig. 6.4). Phe165Tyr mutation leads to the formation of the hydrogen-bond chain Tyr165. . .Arg96. . .O=C-imidazoline. Another important feature of BrUSLEE is that the spatial fixation of the chromophore's tyrosine (Tyr66) is weakened. First, there is now no hydrogen bond between Thr203 and Tyr66 due to changes in Thr203 side-chain conformation (it is twisted away from the chromophore). Second, hydrogen bonding between His148 and OH-Tyr66 is weaker than in the parental protein (bond length is 3.52 Å vs 2.89 in EGFP). Third, Tyr145Met mutation leads to the increased range of motion of the chromophore (Tyr66 movements). Overall, in EGFP one can count up to 9 hydrogen bonds around the chromophore (Chro-HSD148, Chro-W, Chro-Thr203, CHro-GLUP(2), Chro-Arg96, W-Ser205, Ser205-Glup222, Chro-Tyr145), whereas in BrUSLEE only 6 hydrogen bonds can be formed.

6.2.2 Time-resolved fluorescence

All four mutants exhibit maximum emission at approximately 510 nm (2.43 eV), which is characteristic for EGFP. In this region, fluorescence decay of EGFP is dominated by a characteristic lifetime of 2.8 ns (88.7 %) and a minor component $\sim$2.0 ns (11.3 %). Mutations lead to the appearance of the fast (sub-nanosecond) component and a significant reduction of average fluorescence lifetime (Table 6.1).

Table 6.1: Lifetime distributions of EGFP and the mutants at 510 nm (2.43 eV).

System	τ_1, ns	A_1, %	τ_2, ns	A_2, %	τ_3, ns	A_3, %	$\langle\tau\rangle$, ns
EGFP			2.0	11.3	2.8	88.7	2.71
T65G	0.82	88.5	2.0	11.5			0.96
T65G_Y145M	0.52	91.0	1.5	9.0			0.61
T65G_Y145M_F165Y	0.51	83.3	1.4	16.4	2.3	0.3	0.66

Lifetimes represent fluorescence decay measured at 29°C. Average lifetime is computed as $\langle\tau\rangle = \sum_i A_i\tau_i$.

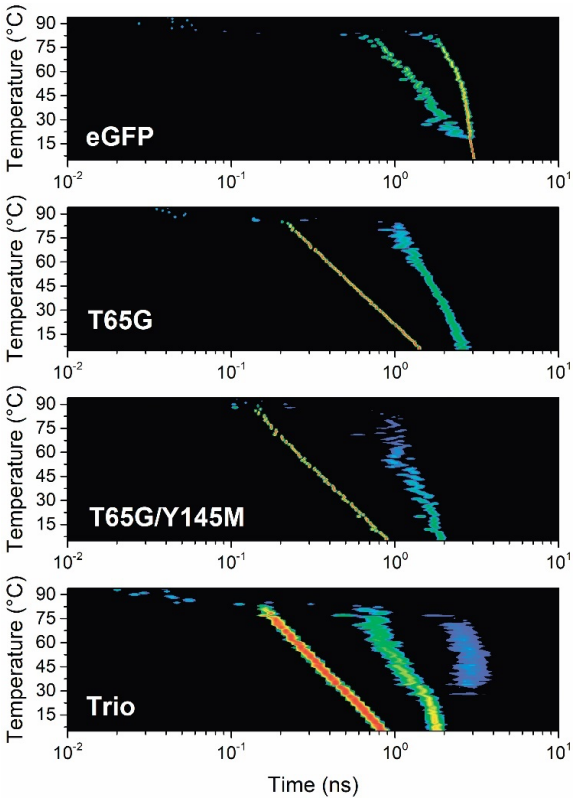


Figure 6.5: Temperature dependence of fluorescence lifetimes in EGFP and the mutants. Fluorescence decay was measured at 510 nm under 470 nm excitation by 50 ps FWHM laser pulses. Color represents the logarithm of the amplitude of the corresponding component (Data courtsey to Bogdanov *et al.*

Temperature dependence of fluorescence lifetimes of EGFP and its mutants (presented as Arrhenius plots in Fig. 6.5) reveals principally linear parts for fast components, while transitions of slow components are characterized by more complex s-shaped dependency, except

T65G/Y145M double mutant in which slow component almost disappears at high temperatures. Such nonlinear behavior is usually attributed to gradual changes of protein's conformation.

Several linear parts of temperature dependencies can be used to estimate effective activation energies (E_a) for internal conversion; the results are given in Table 6.2. We observe that the fastest component has the largest E_a in all systems and that the respective value is similar among mutants and in the entire temperature range. Its value (4.4-8 kcal/mol) is close to the computed torsional barrier around the double bond of the isolated chromophore in the excited state (3.59 kcal/mol)[26], suggesting that the fast component corresponds to the chromophore twisting. The excited-state dynamics simulations (discussed in Section 6.2.3) confirm that the timescale of the torsional motion is indeed similar to the timescales of the fast components.

The E_a for the slow components are much smaller, suggesting that lower-frequency vibrational motions are responsible for radiationless relaxation of the chromophore locked in the planar configuration. For EGFP, E_a extracted from low temperatures equals 0.59 kcal/mol ($\sim$205 cm^{-1}). At high temperatures energy barriers for the slow components increased up to 2.2-3.5 kcal/mol. A striking feature of the triple mutant (BrUSLEE) is the presence of a slow component of fluorescence decay, which has lifetime and activation energy close that of EGFP.

Table 6.2: Activation energies (kcal/mol) for internal conversion of EGFP and its mutants.

Range	EGFP		T65G		T65G/Y145M		BrUSLEE		
	τ_2	τ_3	τ_1	τ_2	τ_1	τ_2	τ_1	τ_2	τ_3
Below 45°C		0.59	4.78	1.81	4.47	2.28	4.41	1.24	
Above 45°C	3.76	2.93	4.78	3.21	4.47	2.28	4.41	3.22	1.79

Freezing initially causes reduction of lifetimes for all samples. Since at low temperatures lifetime of EGFP is lower than at room temperature (1) we can not use it as tau0 to calculate the quantum yield and (2) we have to postulate that this protein has (at least) two states with high fluorescent lifetimes. Almost no dependence of average lifetime of EGFP on temperature in 10 200 K region. Together with the absence of broadening of spectrum it suggests that

the chromophore is locked in a specific configuration which is not sensitive to temperature. Mutants also reach this state, but at significantly lower temperatures. Trio shows the highest average lifetimes at deep temperatures.

6.2.3 Computational results

Analysis of ground state structure from MD simulations

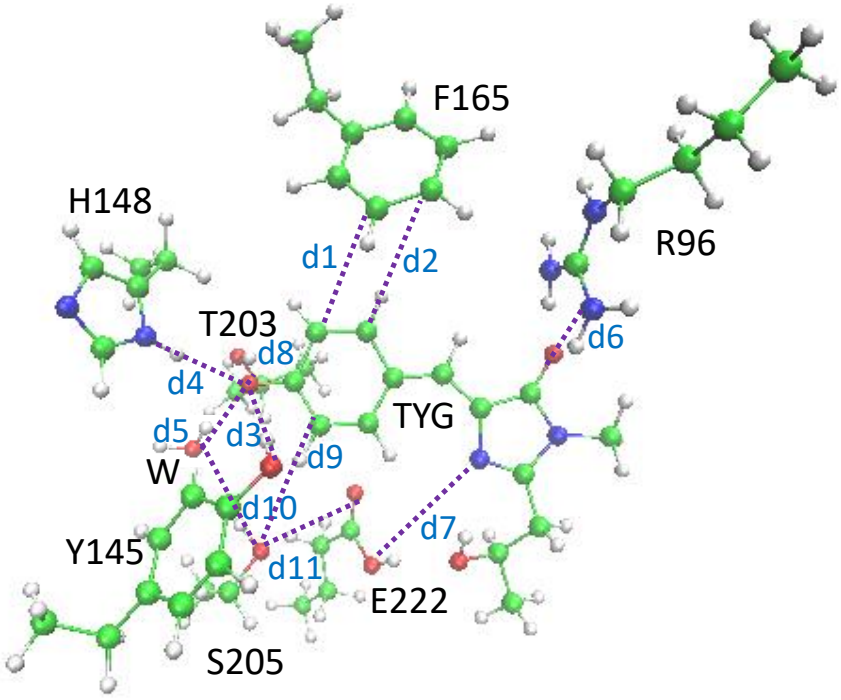


Figure 6.6: Definition of the key distances in EGFP. d1 = CRO66:CE1-PHE165:CE2; d2= CRO66:CD1-PHE165:CZ; d3 = CRO66:OH-TYR145:OH; d4 = CRO66:OH-HSD148:ND1; d5 = CRO66:OH-W84:OH2; d6 = CRO66:O2-ARG96:NH2; d7 = CRO66:N2-GLUP222:OE2; d8 = CRO66:OH-THR203:OG; d9 = CRO66-CE2-SER205:OG; d10 = SER205:OG-W84:OH2; d11 = SER205:OG-GLUP222:OE2.

Figures 3 and 4 compare EGFP and BrUSLEE (Trio) crystal structures. In this section, we analyze the results of equilibrium MD simulations for EGFP, BrUSLEE, and the two mutants;

for each we consider 3 different protonation states of His148. Fig. 6.6 shows the key distances used for structural analysis. Tables 6.6-6.9 and Figure 6.16 show the values of the key structural parameters extracted from the crystal structures and from the MD simulations (averaged along equilibrium trajectories at T=298 K). The main observations are:

- For EGFP, the best overall agreement with the crystal structure is observed for HSD148, whereas for BrUSLEE the best agreement is observed for HSE148.

- Focusing on the distance between His148 and Chro ($d4$) we note that for EGFP HSP148 shows better agreement with the crystal structure. We note that a study of a subatomic resolution X-ray structure of GFP in the neutral (T203I mutant) and anionic (S65T and E222Q mutants) forms[43]. For the neutral form, hydrogen atom densities show that the chromophore is in the neutral form, His148 is in HSD form, and Glu222 is in anionic form, which is consistent with our choices of protonation states in neutral GFP. For the anionic form, the maps confirm that Glu222 is in neutral form (in agreement with the proton wire picture), but His148 is positively charged (HSP)—this suggests that in the ground state there is an additional proton involved in protonation equilibrium.

- Comparing EGFP and BrUSLEE, BrUSLEE possess less planar chromophore compared to that in EGFP. Distance in Cro-HIS148 is larger in BrUSLEE compared to EGFP. This may be indicative of the fact that HSE is suitable protonation state for BrUSLEE where as HSD is suitable for EGFP. THR203 exists in different conformation in BrUSLEE compared to EGFP. Distance in SER205-GLUP222 is larger in BrUSLEE compared to EGFP. This is indicative of a weaker hydrogen bonding around the chromophore.

Chromophore planarity and hydrogen-bond pattern

The key structural parameters related to the photophysical properties are chromophore planarity and the number of hydrogen bonds around the chromophore. We also analyzed partial

stacking between the chromophore and residue 165 (Phe165 in EGFP) but found that it does not change significantly among the mutants and does not correlate with photophysical properties. Fig. 6.12 shows the key parameters characterizing the planarity of the chromophore. Deviation from the planarity can be characterized by the sum of the two torsion angles: $\Delta = \phi + \tau$. Fig. 4 in the main draft shows the hydrogen-bond network around the chromophore. Hydrogen bonds were characterized by the *VMD hbond analyzer* plugin, with the distance cutoff in polar atoms set to 3.5 Å and the angle cutoff set to 30°.

Table 6.3 Shows the chromophore's planarity and the range of twisting motion is sensitive to mutations and depends on the protonation state of His148. The chromophore is most planar in Duo-HSD, EGFP-HSP, and EGFP-HSD (main form). BrUSLEE shows quite noticeable deviations from planarity for all 3 protonation states of His148. The average number of hydrogen bonds around the chromophore is smaller in the mutants than in EGFP. For each structure, the number of hydrogen bonds is smallest for HSE, because this form cannot form hydrogen bonds with the chromophore.

Table 6.3: Chromophore planarity and the number of hydrogen bonds around the chromophore. (Averaged over 400 snapshots from MD at 298 K, standard deviation is in parenthesis).

Mutant	HIS148	Δ	H-bond
EGFP	HSD	6.88 (5.08)	5.50 (0.98)
	HSE	8.96 (6.67)	4.16 (0.80)
	HSP	6.41 (4.67)	4.45 (1.00)
T65G	HSD	8.48 (6.10)	3.69 (0.95)
	HSE	8.87 (6.51)	3.76 (0.89)
	HSP	8.74 (5.97)	4.49 (0.96)
Duo	HSD	6.21 (4.58)	4.31 (0.87)
	HSE	7.20 (5.32)	3.44 (0.90)
	HSP	7.89 (5.56)	4.54 (1.06)
BrUSLEE	HSD	10.06 (6.87)	4.19 (0.83)
	HSE	8.33 (5.41)	3.54 (0.63)
	HSP	7.98 (5.75)	4.35 (0.78)

Ground-state structure analysis and populations of different protonation states

We considered 3 different protonation states of His148: HSD (protonated at N_δ), HSE (protonated at N_ϵ), and HSP (protonated at both nitrogens), see Fig. 6.10 in the Appendix A1. The computed Gibbs free energy differences are summarized in Table 6.12 and the respective populations are shown graphically in Fig. 6.7. At 298 K, the calculated Gibbs free energies suggest that in EGFP and T65G, the main protonation state are HSD (87% and 86%, respectively), with HSP (13% and 14%) being also present. In contrast, in Duo and Trio, the main protonation state is HSE (97 and 70%). In BrUSLEE, the two other states are also present (HSD 29% and HSP 2%). These computed populations correlate well with the populations extracted from fluorescence decay.

The calculations at T=100 K show that the distinct populations can be present at low energies. This is because Gibbs free energies include entropic factor and are also temperature-dependent. In EGFP, the population of HSD drops to 68 %. In BrUSLEE, the population of the main form (HSE) increases to 91 %, with the rest being HSD (the population of HSP drops below 1%). The most important thing to note is that distinct population can coexist in a wide temperature range.

Excited-state dynamics

The results of MD simulations on the excited states are shown in Fig. 6.8.Fig. 6.9 shows the population decay of planar population due to excited-state twisting. For each protonation state, we observe nearly perfect linear fit of the twisting kinetics, which means that in our simulations there are no inter-converting conformers that could give rise to multi-exponential fluorescence decay. The respective lifetimes (shown in each panel of Fig. 6.9) are obtained by linear fit.

As documented in our previous study[26] (see also Chapter 4), the twisting rate is different in mutants because T65G mutation weakens the hydrogen-bond network around the chromophore.

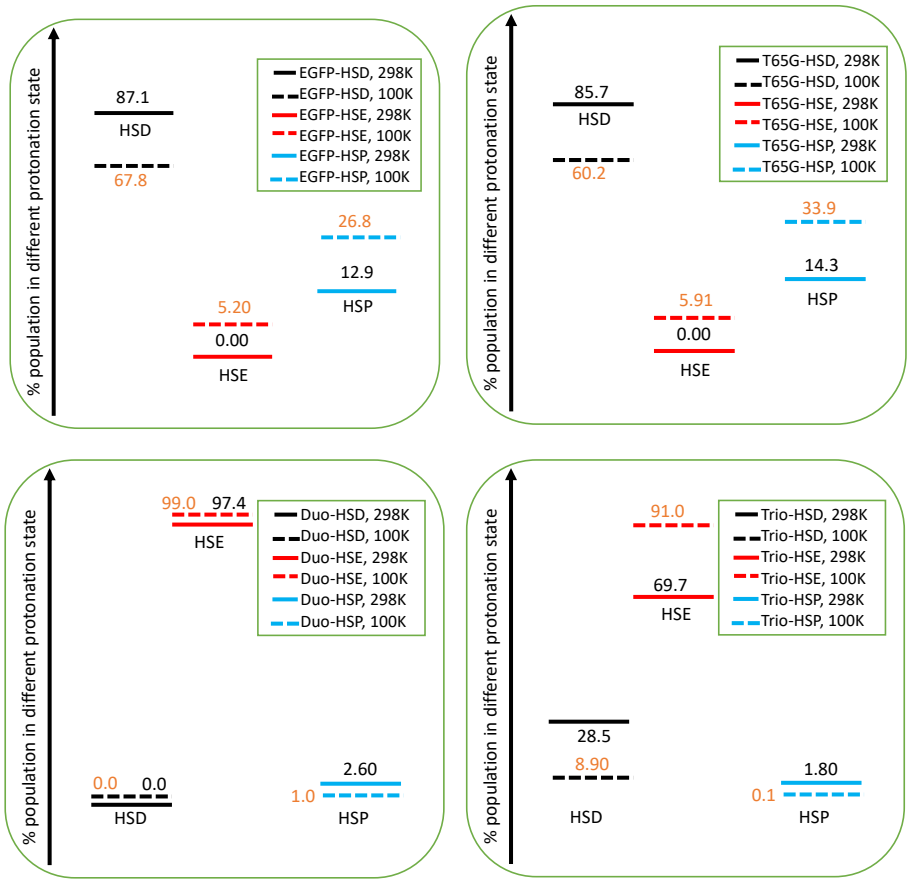


Figure 6.7: Relative populations (at 298 K and 100 K) of different protonation states of His148 in EGFP (top left), T65G (top right), Duo (bottom left), and BrUSLEE (Trio) (bottom right).

Protonation states of His148 also affects the twisting rate—in particular, HSE exhibits the fastest twisting, which can be explained by the inability of this state to form a hydrogen bond with the chromophore. These markedly different lifetimes suggest that the observed multi-exponential fluorescence decay might be due to the co-existence of multiple protonation states of His148.

Table 6.4 shows computed populations and average excited-state lifetimes and estimated FQY. The radiationless times extracted from twisting kinetics agree rather well with experimental results.

The extracted radiationless lifetimes are given in the main text. The percentage of the planar chromophore ($\phi < 50°$) after 3 ns is indicated on the figure.

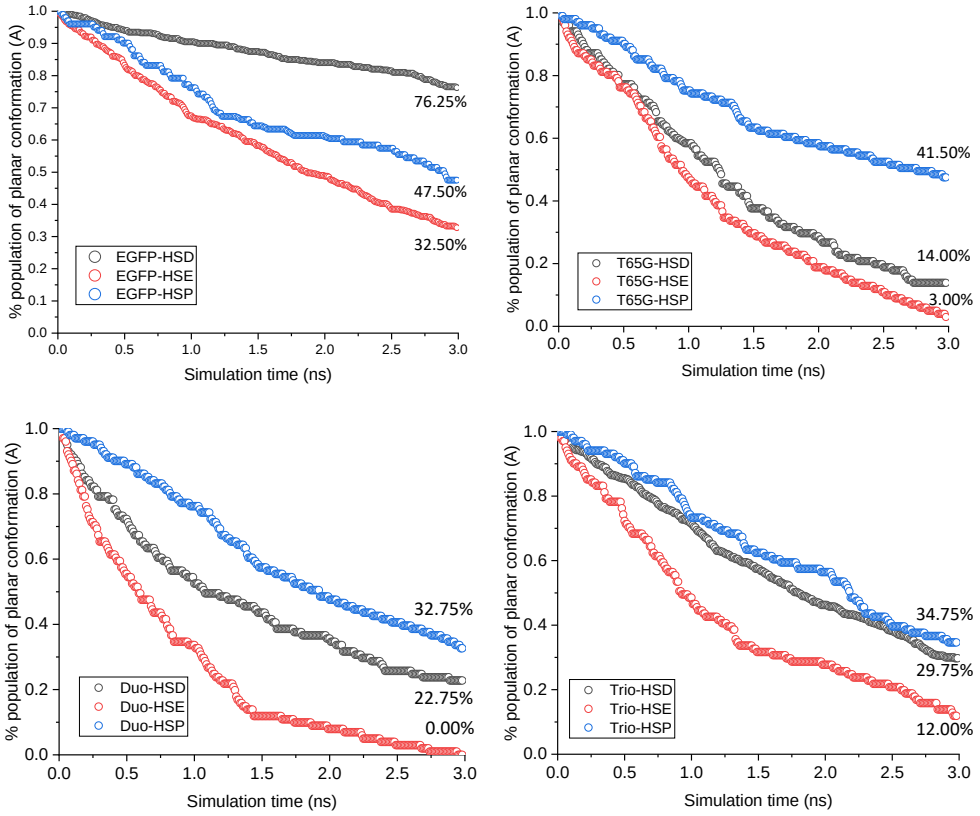


Figure 6.8: Evolution of planar population in excited-state molecular dynamics simulations of EGFP, T65G, Duo, and BrUSLEE (Trio). The numbers indicate the surviving population of the planar chromophore after 3 ns of dynamics.

6.3 Conclusions

Photophysical properties of EGFP, T65G, Duo, BRUSLEE are determined by an interplay between chromophores intrinsic electronic structure, its interactions with the surrounding residues, and several competing excited-state processes. We begin connecting the macroscopic observables (extinction coefficients, brightness, and photostability) with the microscopic

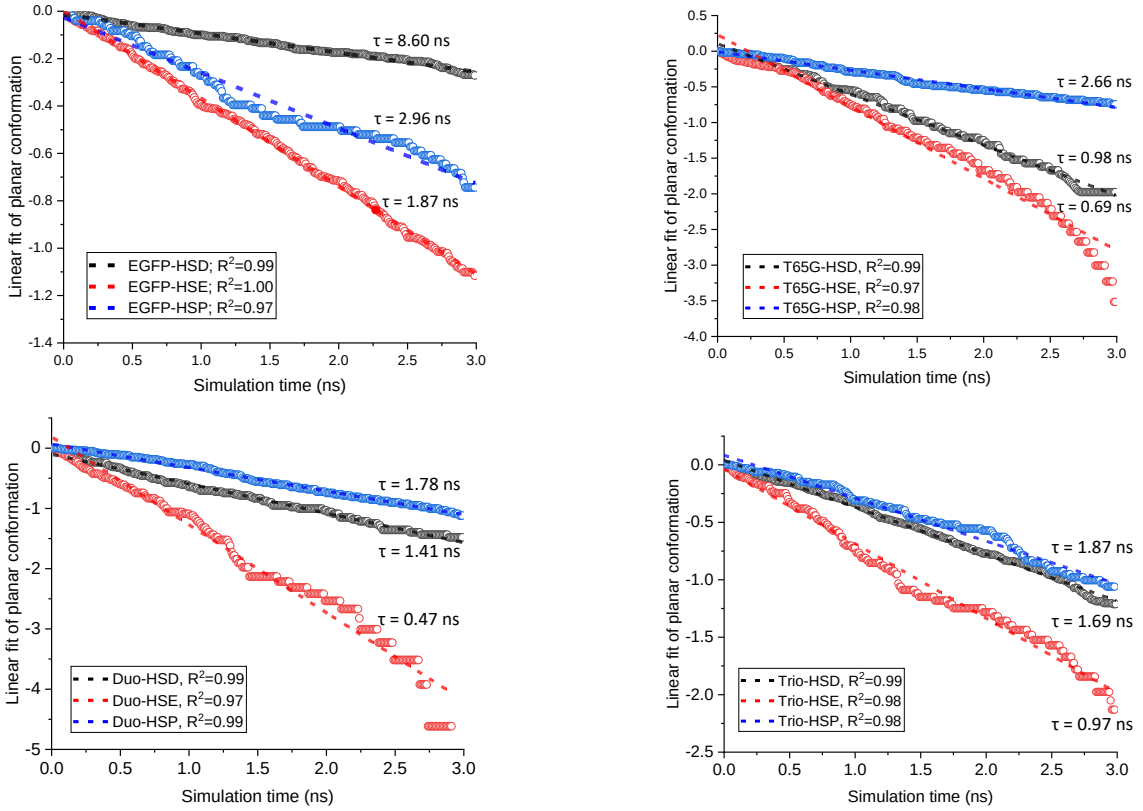


Figure 6.9: Excited-state dynamics: Decay of planar population in EGFP, T65G, Duo, and BrUSLEE. Lifetimes are obtained as linear fit for ln[A].

Table 6.4: Computed values of average lifetime (in ns), percentage population of each protonation states, and fluorescent quantum yield. Experimental values are given in parenthesis.

Mutant	HIS148	τ, theory (exp)	population, theory (exp)	$\langle\tau\rangle$	FQY	$\langle FQY\rangle$, theory (exp)
EGFP	HSD	3.93 (2.8)	0.871 (0.887)	3.69 (2.71)	0.54	0.51 (0.60)
	HSE	1.48	—		0.21	
	HSP	2.10 (2.0)	0.129 (0.113)		0.29	
T65G	HSD	0.85 (0.82)	0.857 (0.885)	1.00 (0.95)	0.13	0.15 (0.10)
	HSE	0.63	—		0.09	
	HSP	1.90 (2.0)	0.143 (0.115)		0.28	
Duo	HSD	1.16	—	0.46 (0.61)	0.17	0.07 (0.08)
	HSE	0.44 (0.52)	0.974 (0.91)		0.07	
	HSP	1.40 (1.5)	0.026 (0.09)		0.21	
Trio	HSD	1.34 (1.4)	0.285 (0.164)	1.00 (0.65)	0.20	0.15 (0.3)
	HSE	0.85 (0.51)	0.697 (0.833)		0.13	
	HSP	1.46 (2.3)	0.018 (0.003)		0.22	

properties of the chromophores. The most interesting features of EGFP and BRUSLEE are the existence of multiple decay channels, which are temperature dependent. As one can see from population analysis, different population of HIS148 twists in excited-state at different time (varying in the range of 1-8 ns). This explains the reason behind multiple subpopulations in mutants. Structural analysis, especially hydrogen-bonding analysis, indicates more hydrogen bonds prevent the twisting in excited-state. For example, HSD form possesses the maximum number of hydrogen bonds in mutants whereas HSE possesses the lowest number of hydrogen bonds, resulting in a faster twist of that form. Competition of enthalpy and entropy at different temperature explains free energy change upon change in protonation state and temperature dependence of that process in equilibrium. On the other hand, intrinsic lifetime is a function of excitation energy and oscillator strength.

To conclude, with the help of series of electronic structure calculations and MD simulations (both in ground and excited-state) we have rationalized the properties of the newly developed variant of EGFP, BRUSLEE, which will lead us designing new FPs with desirable properties.

6.4 Appendix A: Computational details

6.4.1 Appendix A1: Model structures and ground-state dynamics

We begin with the crystal structures of EGFP and BrUSLEE (Trio). EGFP structure was taken from protein data bank (PDB) id: 2Y0G[27]. The mutants are built from the 2Y0G crystal structure by single (T65G) and double (T65G-Y145M, Duo) mutations using the *VMD Mutator* plugin. Hydrogen atoms were added using the *VMD* plugin and a modified (to include the chromophore) *CHARMM27* topology file. Protonation states of titratable residues were initially assigned by *Propka*[28] and then manually set for the chromophore and His148.

In EGFP, the chromophore is deprotonated. The two most important residues near the chromophore are His148 and Glu222. Glu222 can be GLU (anionic) or GLUP (protonated); *Propka*[28] suggested GLUP state (pKa 9.2) for the Glu222, which was validated by geometry optimizations and MD simulations in our previous work[26,29]. Hence, in this study we consider Glu222 to be protonated in all structures. For His148, we considered 3 different protonation states (shown in Fig. 6.10) for each system: HSD (protonated at N_δ), HSE (protonated at N_ϵ), and HSP (protonated on both N, positively charged).

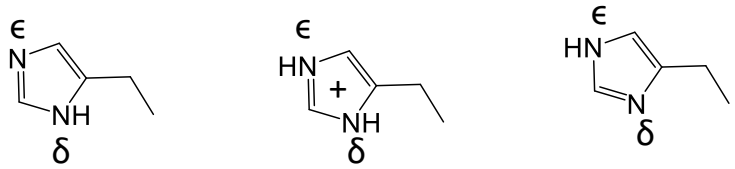


Figure 6.10: Different protonation states of histidine: HSD (left), HSP (middle), and HSE (right).

Charged amino acids on the surface were locally neutralized by adding counterions close ($\sim$4.5 Å) to them. Charged residues that do not form salt bridges inside the protein barrel were also neutralized by adding appropriate counter-ions at the surface. This protocol resulted in the addition of 21 Na^+ and 14 Cl^- for the HSD and HSE structures, and and 20 Na^+ and 14 Cl^- for the HSP structures. The proteins were solvated in water boxes producing a solvation layer of 15 Å. The TIP3P water model was used to describe water.

Ground-state MD simulations were performed using these solvated neutralized model structures as follows:

1. Minimization using steepest descent algorithm for 2,000 steps (protein, crystal water, counterions).

2. Minimization using steepest descent algorithm for 2,000 steps of the fully solvated structure (keeping protein frozen), with the subsequent equilibration of the solvent (keeping

the protein frozen) for 500 ps with 1 fs time step using the NPT (isobaric-isothermal) ensemble.

3. Full equilibration of the system for 2 ns (with 1 fs time step) with periodic boundary condition (PBC) using the NPT ensemble (Noose-Hoover barostat with Langevin dynamics).

4. Production run for 2 ns with 1 fs time step using the NPT ensemble. Pressure and temperature were kept at 1 atm and 298 K.

These simulations provided snapshots (taken from the production run, step 4) representing ground-state equilibrium dynamics and were used to analyze ground-state structures and hydrogen-bond pattern around the chromophore. They also served as a starting point for calculating free energies of different protonation states of His148 (See Appendix D), for computing excitation energies by the QM/MM protocol (Appendix A2), and as starting structures to perform MD simulations on the excited-state surfaces (Appendix A3).

MD simulations were performed with NAMD[30].

6.4.2 Appendix A2: QM/MM setup for excited-state calculations

We computed electronic properties (vertical excitation energies, oscillator strengths) using snapshots from equilibrium trajectories (production runs in the MD simulations) using the following QM/MM scheme. The chromophore and selected residues were included in the QM region and the rest of the system was treated as fixed MM point charges via electrostatic embedding. Figure 6.11 shows the definition of the QM subsystem. Hydrogen atoms were added at the QM/MM boundary to saturate the valencies. Point charges on the atoms adjacent to the QM/MM boundary (such as red and green atoms in Fig. 2 in the main text) were set to zero and the excess charge was redistributed over the rest of the atoms of the respective residues to avoid over-polarization of the QM atoms at QM/MM boundary.

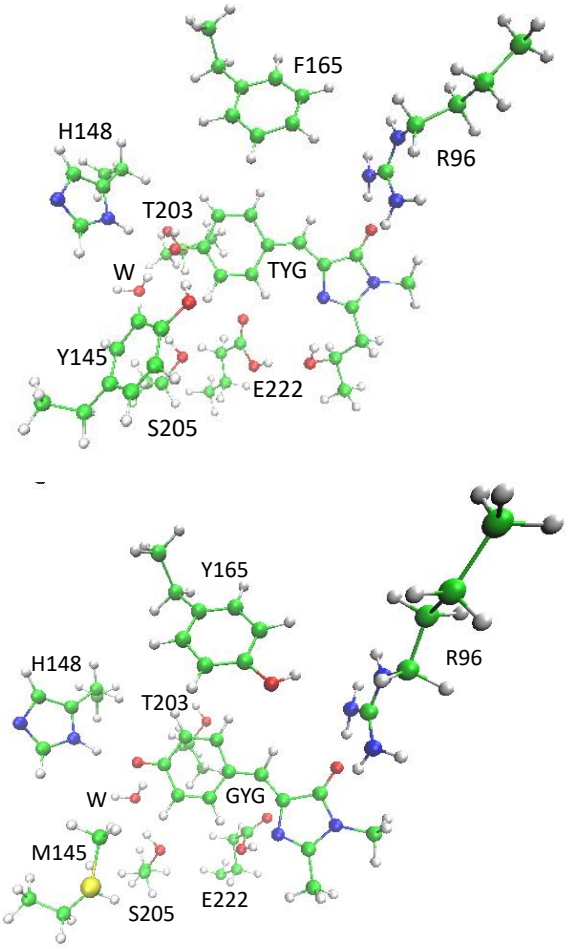


Figure 6.11: Top: Residues involved in QM/MM calculations of EGFP, Bottom: BrUSLEE. Chromophore, water, residues 145, 148, 165, 96, 203, 205, 222 were included in the QM region in calculations of spectra and electronic properties.

Electronic structure calculations were performed at the ωB97X-D/aug-cc-pVDZ[31,32] level of theory. All quantum chemistry and QM/MM calculations were carried out using the *Q-Chem* electronic structure package[33,34].

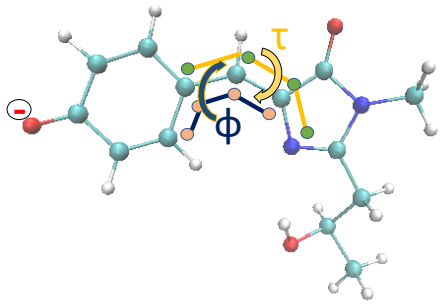


Figure 6.12: Definition of the two torsional angles ϕ and τ describing chromophore twisting. ϕ describes twist around the single bond (phenolate flip) and τ describes twist around the double bond (imidozalinone flip).

6.4.3 Appendix A3: Molecular dynamics simulations on the excited-state surfaces

Following the same procedure as in our earlier work[26], we modified forcefield parameters of the chromophore to account for the changes in the bonding pattern upon photoexcitation. Specifically, we changed the parameters for methyne bond-lengths, angles, and torsional potential, as well as selected partial charges. The values of these parameters were computed in the ground and excited states using DFT (ωB97X-D/aug-cc-pVDZ). We then computed the *difference* (for charges, bond-lengths, and angles) or *ratio* (for force constants) in the ground and excited states and used these values to either shift or scale the respective parameters from the *CHARMM27* forcefield. The resulting forcefield parameters are given in Ref. 26. Below we explain the key differences.

The most important parameters are the two torsional angles ϕ and τ (see Fig. 6.12). The PES scans (Fig. 6.13) show that the minimum in the ground state corresponds to the planar chromophore, whereas in the excited state the planar structure is separated by relatively low barriers from the two minima corresponding to the strongly twisted chromophore. We fitted the excited-state potential to reproduce the location of the new minima (Fig. 6.13). From this fit,

we extracted the force contact. The resulting torsional potential for excited-state calculations has the following form:

$$E = k[1 + cos(n\phi - \delta)], \tag{6.5}$$

where n is periodicity, δ is phase, ϕ being the optimized torsional angle. The parameters are given in Table 6.5 and the resulting torsional potential (computed with the modified forcefield is shown in Fig. 6.14. As one can see, our fit reproduces the barriers for twisting reasonably well, but does not reproduce the depth of the well of the twisted structures (the fitted potential is too shallow). Hence, to prevent the trajectories from re-crossing, in the excited-state MD simulations we simply stop the trajectories once they twist by more than a specified threshold value (50°).

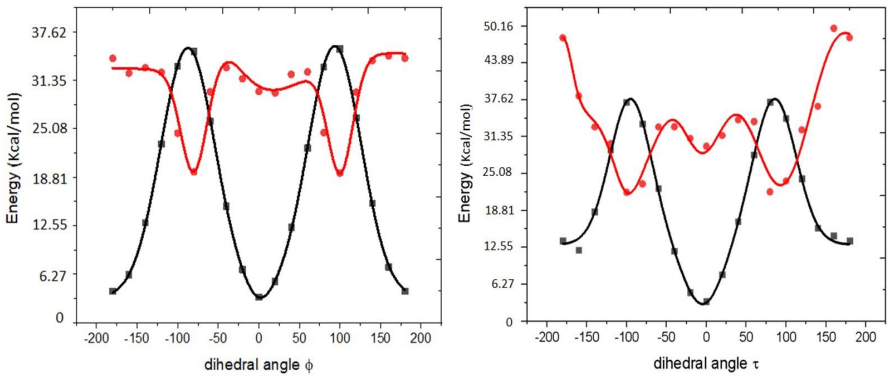


Figure 6.13: Ground- and excited-state torsional potentials for ϕ (twisting of the phenolic ring) and τ (twisting of the imidazolinone ring) of the bare HBDI chromophore. Black dots are *ab initio* calculations whereas red and black lines mark *ab initio* force-field. The barrier heights for twisting along ϕ and τ in the excited state are 3.5 kcal/mol and 3.2 kcal/mol, respectively. The respective ground-state barriers are 32.1 and 34.9 kcal/mol. Reproduced from Ref. 26.

Table 6.5: Parameterized force constant and periodicity (n) for torsional potentials for angles ϕ and τ.

Dihedral	$k_{gs,charmm}$	n (gs)	n (ex)	$k_{gs,qm}$	$k_{ex,qm}$	$k_{ex,charmm}$
ϕ	2.7	2	4	15.05	3.79	0.68
τ	3.9	2	4	14.99	4.90	1.27

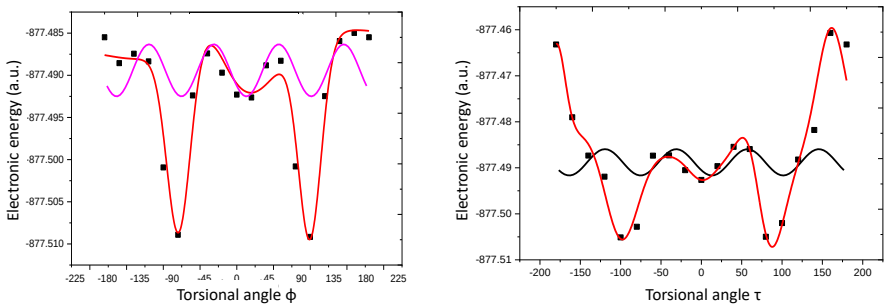


Figure 6.14: Excited-state torsional potentials for ϕ (left) and τ (right) of the bare HBDI chromophore. Red curves: fit to *ab initio* calculations (from which the parameters were extracted). Pink and black curves: torsional potentials computed with the modified force-field.

6.4.4 Appendix A4: Ab initio molecular dynamics (AIMD)

As an additional validation of our force-field parameters, we carried out *ab initio* molecular dynamics (AIMD) simulations on the excited-state surfaces. These calculations were performed using the ONIOM embedding with large QM (shown in Fig. 6.11), ωB97X-D/cc-pVDZ, and *CHARMM27* force-field. 11 trajectories intimated from random snapshots and with initial velocities corresponding to 298 K thermal distribution were propagated for 3 ns with 1 fs time step (3,0000,000 steps) with constant energy (NVE) ensemble. All atoms were allowed to move, except for the link atoms, which were pinned to their positions from the MM-relaxed structures.

6.4.5 Appendix A5: Calculation of free-energy difference between different protonation states of His148

To compute free-energy differences, we employ the thermodynamic cycle shown in Fig. 6.15. This approach[36,37], called quantum mechanical thermodynamic cycle perturbation (QTCP), allows one to compute high-level QM/MM free energy changes between two states A and B based on classical (MM) sampling and a relatively modest amount of QM/MM calculations. In this approach, the free energy change between A and B described by QM/MM

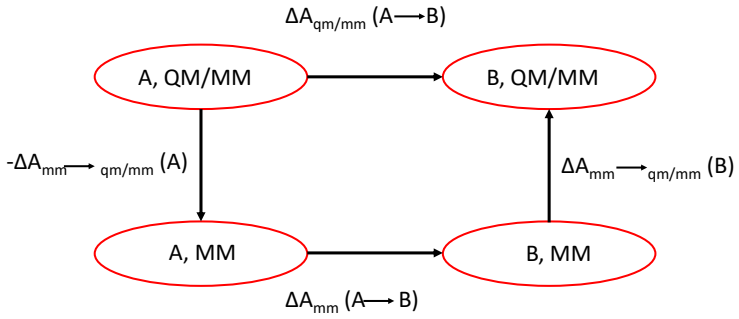


Figure 6.15: The quantum mechanical thermodynamic cycle perturbation (QTCP) method employing a thermodynamic cycle to calculate QM/MM free-energy changes[36].

is calculated as the sum of three terms: (1) free energy change between A described by MM and by QM/MM (-$\Delta A_{mm\rightarrow qm/mm}$ (A)), (2) the free energy change between A and B, with both described by the MM potential (ΔA_{mm}(A→B)), and (3) the free energy change between B described by the MM potential and by QM/MM (Δ $A_{mm\rightarrow qm/mm}$(B)). Hence

$$\Delta A_{qm/mm}(A \rightarrow B) = -\Delta A_{mm\rightarrow qm/mm}(A) + \Delta A_{mm}(A \rightarrow B) + \Delta A_{mm\rightarrow qm/mm}(B), \quad (6.6)$$

$$\Delta A_{mm}(A \rightarrow B) = -k_B T \ln\langle e^{-[E^{tot}_{mm}(B) - E^{tot}_{mm}(A)]/k_B T}\rangle_{mm,A}, \quad (6.7)$$

$$\Delta A_{mm\rightarrow qm/mm} = -k_B T \ln\langle e^{-[E^{tot}_{qm/mm}(X) - E^{tot}_{mm}(X)]/k_B T}\rangle_{mm,X}, \quad (6.8)$$

where k_b = 0.0257 eVK^{-1}. To adapt this scheme to different protonation states, we follow the strategy by Warshel[38]. For HSP→HSD energy difference, this scheme means:

$$\begin{aligned}\Delta A_{qm/mm}(HSP \rightarrow HSD) = {} & k_B T \ln\langle e^{-[E^{tot}_{qm/mm}(HSP) - E^{tot}_{mm}(HSP)]/k_B T}\rangle_{mm,HSP} \\ & -k_B T \ln\langle e^{-[E^{tot}_{mm}(HSD) - E^{tot}_{mm}(HSP)]/k_B T}\rangle_{mm,HSP} \\ & -k_B T \ln\langle e^{-[E^{tot}_{qm/mm}(HSD) - E^{tot}_{mm}(HSD)]/k_B T}\rangle_{mm,HSD}. \end{aligned} \quad (6.9)$$

Once free energies are computed, one can evaluate the populations of different forms by using Maxwell-Boltzmann equation:

$$\frac{P_A}{P_B} = e^{-\frac{\Delta A_{qm/mm}(A\rightarrow B)}{k_b T}}. \tag{6.10}$$

We considered three protonation states of the HIS148 (HSD, HSE, HSP, see Fig. 6.10) to compute HSP→HSD and HSP→HSE free-energy differences. Free energy difference for HSD→HSE is the computed as the difference between the HSP→HSD and HSP→HSE free-energy differences. The three terms involved were computed as follows:

1. We carried out QM/MM electronic energy calculation on 400 ground-state snapshots from MD using the mechanical embedding scheme (ONIOM) and ωB97X-D/aug-cc-pVDZ. We compute QM/MM (with His148 in QM and the rest of the protein in MM) and pure MM energies (all atoms are in MM) for all protonation states. These energies are used to evaluate ΔA$_{mm\rightarrow qm/mm}$ terms: Eqns. (6.6) and (6.8).

2. We then consider snapshots for the HSP state for each mutant. We remove protons from either δ or ϵ nitrogen (giving rise to the HSD and HSE forms, respectively) and place an extra proton at a fixed position in bulk (making sure there are no bad contacts to water). We then compute MM energies for these two structures (original snapshot and the modified one) and use them them to evaluate ΔA$_{mm}$(HSP→HSD/HSE). These calculations are done for 400 snapshots.

The calculations at room temperature (298 K) were carried out using the respective equilibrium MD simulations. Free-energy calculation at low temperature (100 K) were carried out as follows. For each model system (mutant/protonation state), we took 20 snapshots from room-temperature simulations and re-equilibrated them for 1 ns (with reinitialized velocity). The MD

simulations were performed as described above. From each trajectory, we took 20 snapshots for the QM/MM free-energy calculation (total 20x20 = 400).

6.4.6 Appendix A6: Calculation of radiative lifetimes and extinction coefficients

Radiative lifetime is given by[39]:

$$\frac{1}{\tau_r} = \frac{\omega_0^2 f_{abs}}{2\pi (c')^3 \epsilon}, \tag{6.11}$$

where c' is the speed of light in the medium ($c' = c/n$; c is the speed of light in vacuum and n is the index of refraction) and ϵ is the dielectric constant. For vacuum, ϵ=1 and c=137. The index of refraction of water is 1.33; the refractivity of protein solutions is generally larger, around 1.6 Dielectric constant in proteins is small (i.e., 28)[40,41].

We compute macroscopic extinction coefficient using the following expression[42]:

$$\epsilon(\tilde{\omega}) = \sum_i \frac{N_a e^2}{4 m_e c^2 \epsilon_0 \ln 10 \sqrt{\pi}} \frac{f_i}{\Gamma} \exp\left[-\left(\frac{\tilde{\omega} - \tilde{\omega}_i}{\Gamma} \right)^2 \right], \tag{6.12}$$

where $\epsilon(\tilde{\omega})$ is the molar extinction coefficient measured in $\mathrm{L mol^{-1} cm^{-1}}$; $\tilde{\omega}$ is the excitation wavenumber, N_a is the Avogadro number, e is the electron charge, m_e is the electron mass, c is the speed of light in cm $\mathrm{s^{-1}}$, ϵ_0 is the vacuum permittivity in F $\mathrm{cm^{-1}}$, f_i is the oscillator strength of the state i, and Γ is the broadening factor in $\mathrm{cm^{-1}}$. We used wavenumbers, since the units are L$\cdot$ $\mathrm{mol^{-1}}\cdot$ $\mathrm{cm^{-1}}$ and so Γ is in $\mathrm{cm^{-1}}$. The coefficient is:

$$\frac{N_a e^2}{4 m_e c^2 \epsilon_0 \ln 10 \sqrt{\pi}} = 1.277 \times 10^8 \ \mathrm{L \cdot mol^{-1} \cdot cm^{-2}}. \tag{6.13}$$

The choice of Γ is the biggest uncertainty in the calculations, as we cannot compute it from first principles. In calculation we use Γ= 0.3 eV.

6.5 Appendix B: Analysis of structures from equilibrium MD simulations

Table 6.6: EGFP. Comparison of the key distances in the crystal structure and in MD simulations (T=298 K) considering 3 different protonation states for His148.

distance	2Y0G	6JGI	EGFP-HSD	EGFP-HSE	EGFP-HSP
d1	4.01	4.07	4.02 (0.29)	3.93 (0.34)	4.33 (0.32)
d2	3.94	4.11	4.12 (0.32)	3.88 (0.31)	4.25 (0.31)
d3	4.43	4.39	3.89 (0.56)	5.32 (0.22)	3.21 (0.42)
d4	2.85	2.87	3.28 (0.39)	3.80 (0.26)	2.75 (0.12)
d5	2.62	2.74	3.21 (0.48)	2.74 (0.14)	3.78 (0.41)
d6	2.73	2.75	2.68 (0.08)	2.70 (0.09)	2.72 (0.10)
d7	2.59	2.67	2.85 (0.17)	2.85 (0.15)	2.88 (0.21)
d8	2.66	2.67	2.84 (0.19)	2.70 (0.12)	3.08 (0.39)
d9	3.94	4.18	4.54 (0.43)	4.32 (0.40)	4.97 (0.37)
d10	2.69	2.76	2.92 (0.19)	4.67 (0.44)	2.84 (0.16)
d11	3.88	3.81	4.01 (0.42)	3.87 (0.39)	4.34 (0.52)

6JGI is the crystal structure of the S65T variant of EGFP at 0.85 Å(from Ref. 43).

Table 6.7: T65G. Comparison of the key distances in the crystal structure and in MD simulations (T=298 K) considering 3 different protonation states for His148.

distance	2Y0G	T65G-HSD	T65G-HSE	T65G-HSP
d1	4.01	4.23 (0.39)	4.09 (0.40)	4.36 (0.42)
d2	3.94	4.33 (0.36)	4.01 (0.39)	4.43 (0.42)
d3	4.43	3.80 (0.53)	5.34 (0.42)	3.32 (0.47)
d4	2.85	3.53 (0.45)	3.83 (0.28)	2.73 (0.13)
d5	2.62	3.23 (0.56)	2.83 (0.21)	3.39 (0.44)
d6	2.73	2.68 (0.09)	2.73 (0.09)	2.71 (0.09)
d7	2.59	4.61 (0.89)	3.45 (0.49)	3.54 (0.71)
d8	2.66	2.93 (0.33)	2.94 (0.51)	3.13 (0.33)
d9	3.94	4.29 (0.51)	4.13 (0.50)	4.12 (0.48)
d10	2.69	2.89 (0.24)	4.49 (0.60)	2.88 (0.21)
d11	3.88	4.21 (0.60)	4.49 (0.49)	4.35 (0.45)

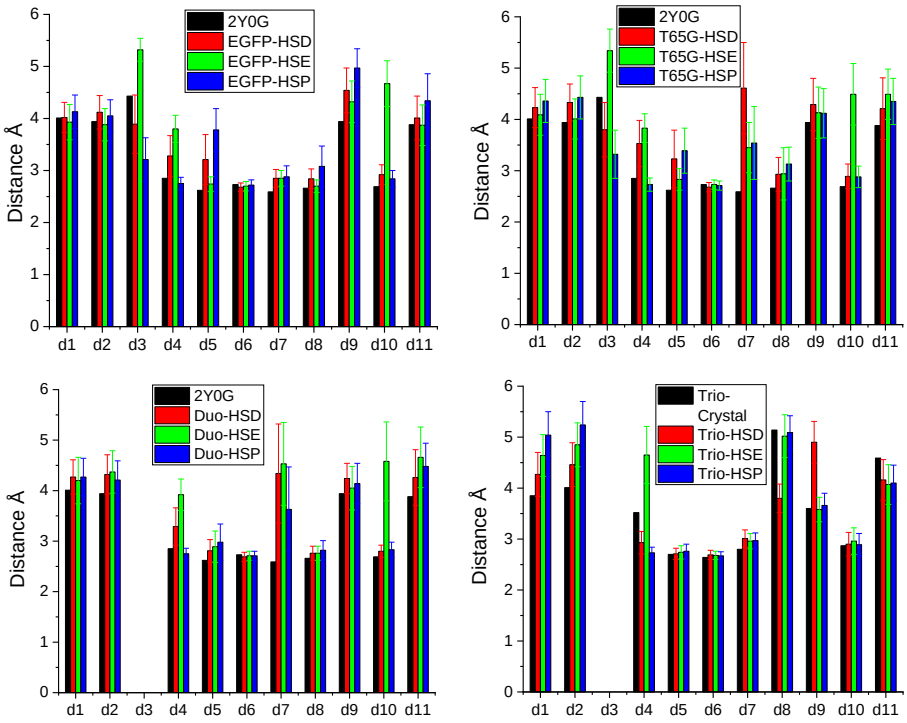


Figure 6.16: Key distances in EGFP (top left), T65G (top right), Duo (bottom left), and BrUSLEE (Trio, bottom right).

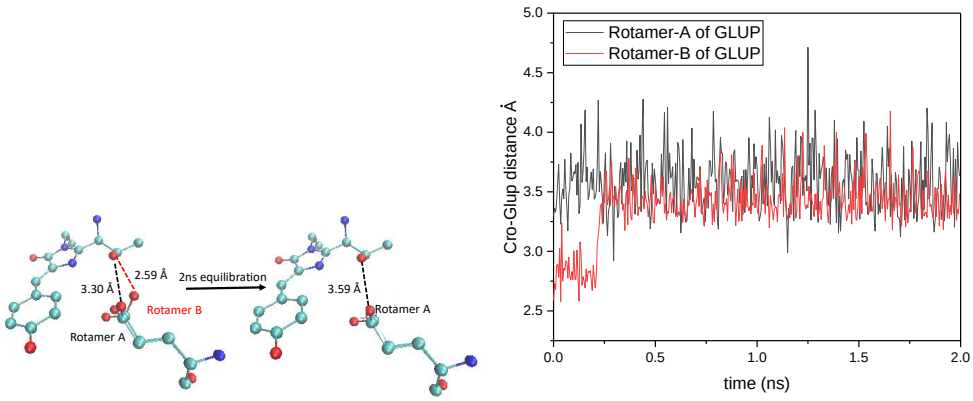


Figure 6.17: Left: Two rotamers of Glu222and the definition of the key distance affected y the rotamers. Right: Equilibrium MD trajectories starting from the two rotameric forms in GLUP222 (T=298 K). The structure of the second rotamer is unstable: it flips after 0.25 ns into the main form and never comes back.

Table 6.8: Duo. Comparison of the key distances in the crystal structure and in MD simulations (T=298 K) considering 3 different protonation states for His148.

distance	2Y0G	Duo-HSD	Duo-HSE	Duo-HSP
d1	4.01	4.27 (0.34)	4.20 (0.46)	4.27 (0.37)
d2	3.94	4.32 (0.39)	4.37 (0.42)	4.21 (0.38)
d3	–	–	–	–
d4	2.85	3.29 (0.37)	3.92 (0.31)	2.75 (0.11)
d5	2.62	2.81 (0.22)	2.89 (0.31)	2.98 (0.36)
d6	2.73	2.69 (0.09)	2.71 (0.09)	2.71 (0.09)
d7	2.59	4.34 (0.98)	4.53 (0.82)	3.63 (0.84)
d8	2.66	2.76 (0.14)	2.76 (0.14)	2.82 (0.19)
d9	3.94	4.24 (0.30)	4.05 (0.43)	4.14 (0.40)
d10	2.69	2.80 (0.12)	4.58 (0.78)	2.83 (0.15)
d11	3.88	4.26 (0.55)	4.66 (0.60)	4.48 (0.46)

Table 6.9: BrUSLEE (Trio). Comparison of the key distances in the crystal structure and in MD simulations (T=298 K) considering 3 different protonation states for His148.

distance	Trio-crystal	Trio-HSD	Trio-HSE	Trio-HSP
d1	3.85	4.27 (0.43)	4.64 (0.41)	5.04 (0.46)
d2	4.01	4.46 (0.43)	4.85 (0.43)	5.24 (0.46)
d3	–	–	–	–
d4	3.52	2.93 (0.22)	4.65 (0.56)	2.73 (0.11)
d5	2.70	2.71 (0.11)	2.74 (0.13)	2.76 (0.14)
d6	2.64	2.69 (0.09)	2.68 (0.08)	2.67 (0.08)
d7	2.80	3.01 (0.17)	2.96 (0.15)	2.97 (0.15)
d8	5.14	3.80 (0.28)	5.02 (0.42)	5.09 (0.33)
d9	3.60	4.90 (0.41)	3.58 (0.24)	3.66 (0.24)
d10	2.87	2.90 (0.23)	2.96 (0.26)	2.89 (0.22)
d11	4.59	4.16 (0.40)	4.07 (0.39)	4.10 (0.35)

6.6 Appendix C: Förster energy transfer between tryptophane and chromophore

Here we describe experiments and simulations of the quenching of the fluorescence of tryptophane (Trp57) at different temperatures, which help to further validate the structures and ground-state equilibrium motions of the protein.

The experiment measures fluorescence lifetime at 360 nm (3.44 eV), which can be interpreted as fluorescence lifetime of Trp. From these raw data efficiency of FRET (E, %) is calculated as:

$$E = (1 - \frac{\tau_{fast}^{da}}{\tau_{slow}^{d}}) \times 100\%, \tag{6.14}$$

where τ_{fast}^{da} and τ_{slow}^{d} are fast and slow components of fluorescence decay at 360 nm. The assumption is that the fast component of fluorescence decay of Trp is due to FRET to the chromophore, whereas slow component is intrinsic Trp excited-state lifetime. The results are shown in Fig. **??**. The actual definition of E is given by the rations of the lifetime (or fluorescence intensity, F) of the donor in the presence and absence of the acceptor:

$$E = 1 - \frac{\tau_D'}{\tau_D} = 1 - \frac{F_D'}{F_D}. \tag{6.15}$$

The efficiency of FRET energy transfer is given by:

$$E = \frac{1}{1 + (r/R_0)^6}, \tag{6.16}$$

where r is the distance between D and A and R_0 is Förster distance:

$$R_0^6 = \frac{2.07 \times 10^4}{128\pi^5 N_A} \frac{\kappa^2 Q_D J}{n^4}, \tag{6.17}$$

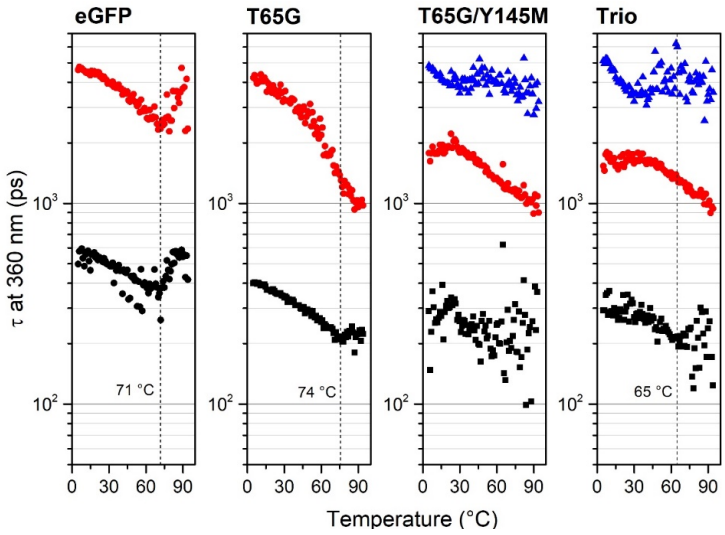


Figure 6.18: Temperature dependence of Trp lifetimes in selected mutants.

where Q_D is the FQY of the donor in the absence of the acceptor, κ^2 is the dipole orientation factor, n is refractive index of the medium (1.33 used here), N_A is the Avogadro number, and J is the spectral overlap of the absorption of the acceptor and emission of the donor:

$$J = \int \bar{f}_D(\lambda)\epsilon_A(\lambda)\lambda^4 d\lambda. \tag{6.18}$$

Here $\bar{f}_D(\lambda)$ is donor's emission spectrum normalized to a unit area and $\epsilon_A(\lambda)$ is molar extinction coefficient of the acceptor. These equations are used in the software *PhotochemCAD*, which was used to analyze the experimental data.

The calculation also requires extinction coefficients; . Q_D is taken as 0.12. κ is:

$$\kappa = \mu_A\mu_D - 3(\mu_D R)(\mu_A R), \tag{6.19}$$

where μ_A and μ_D are normalized transition dipole vectors of the acceptor and the donor, and R is the normalized radius vector between D and A. For randomly oriented D and A (which is

certainly not the case in FPs!), $\kappa = 2/3$. Zero κ corresponds to the perpendicular orientation, $\kappa^2 = 1$ corresponds to parallel orientation, and $\kappa^2 = 4$ to collinear.

While κ can be extracted from the experimental data (as described above), it can also be computed from the equilibrium MD trajectories of mutants using Eq. (6.19); this is described in section 6.6.1 below.

Table 6.10 summarizes the analysis of the experimental FRET measurements. Measured were E, then R_0 were extracted (using r=13.5 Å), then κ was computed from R_0 and J. The experimental data were averaged over lower-temperature measurements, 20-40°. In the analysis of experimental data, Eugene used r=13.5 Å (as in EGFP x-ray structure). We note that the choice of r depends of how one defines the distance (between closest atoms? Or center-of-mass?) and choosing different values affects the values of κ (and vise verse). Specifically, Eqns. (6.16) and (6.17) mean that if one multiplies r by a factor of x, than the values of κ need to be multiplied by x^3 for the $(r/R_0)^6$ remain the same (or, if κ^2 is multiplied by y, then r needs to be multiplied by $y^{1/6}$).

Table 6.10: Summary of the FRET experiments.

System	E% 20°C	R_0, Å	E_λ, $M^{-1}cm^{-1}$	J, cm^6	κ^2
EGFP	87.80	18.7586	55,000	1.8603×10^{-14}	0.07021
T65G	90.59	19.4460	70,000	2.0261×10^{-14}	0.0860
T65G_Y145M	92.80	19.6322	84,500	2.4520×10^{-14}	0.1021
T65G_Y145M_F165Y	92.63	21.3096	86,000	2.8071×10^{-14}	0.0816

$n = 1.33$, r=13.5, Q_D=12%.

6.6.1 Appendix C1: Calculations of the dipole orientation factor from equilibrium MD simulations

The MD simulations were carried out for HSD protonation state. The computed average distance between the donor and acceptor (defined as the CRO66:N2-TRP57:CD2 distance) agrees

well with the crystal structure (CRO66:N2-TRP57:CD2 distance is 13.5 Å in 2Y0G). To compute κ using Eq. (6.19), we carried out TDDFT calculations for 21 QM/MM snapshots. For each snapshot, we performed 2 calculations: one with the chromophore in QM and one with Trp57 in the QM. We used Eq. (6.19) to compute κ. We used two different approaches. In one calculation, we first computed average transition dipole moments for the chromophore and for Trp57 and then computed κ using average values. In the second calculation, we computed κ for each snapshot and averaged them. The second approach is more rigorous as it correctly treats the equilibrium averaging. The first approach can give an idea of how much static orientation of the Chro and Trp is responsible for the observed values. Table 6.11 shows the results of the two calculations. Fig. 6.19 shows experimental versus calculated κ^2 — as you can see, there is an excellent correlation between the theory and experiment (R^2=0.98). The values of κ are relatively small, which is consistent with nearly perpendicular orientation of the two dipoles, $< \theta >$=82-98°. The calculations using average structure overestimate κ, which shows that dynamic fluctuations are important, however, the correlation is very good for both static and dynamic values, meaning that static (average) structures of the mutants are sufficiently different to explain the observed trend.

The differences in the magnitude of κ^2 cannot be reconciled by using different value for the chromophore-Trp57 distance in the experimental analysis or in the calculations. For example, dividing theoretical κ^2 by 2, leads to an adjusted value of $\langle r \rangle$=13.5×1.26=17.01 Å, which is too large. Besides, our current definition of r seems to be consistent with what was used in the experiment, as evidenced by $< r >$. So the discrepancy is likely due to a combined effect of the uncertainties in Q_D, J, and n that enter the equation for R_0. It is quite likely that we accumulate a factor of 2 from the uncertainty in Q_D and J.

Table 6.11: Computed average distance (standard deviation in parenthesis), angle between average transition dipoles, and κ^2. His148 is in the HSD state.

system	$\langle r \rangle$, Å	$\langle \theta \rangle$ (theory^a)	$\langle \theta \rangle$ (theory^b)	κ^2 (theory^a)	κ^2 (theory^b)	κ^2 (exp)
EGFP	13.48 (0.26)	111.59	95.97	0.209	0.143	0.070
T65G	13.14 (0.26)	103.65	91.31	0.309	0.169	0.086
T65G_Y145M	13.31 (0.28)	91.83	96.89	0.472	0.209	0.102
T65G_Y145M_F165Y	13.16 (0.30)	107.6	97.69	0.276	0.154	0.082

[a] Computed using the average values of the dipoles.

[b] Computed by averaging the instantaneous values at each snapshot.

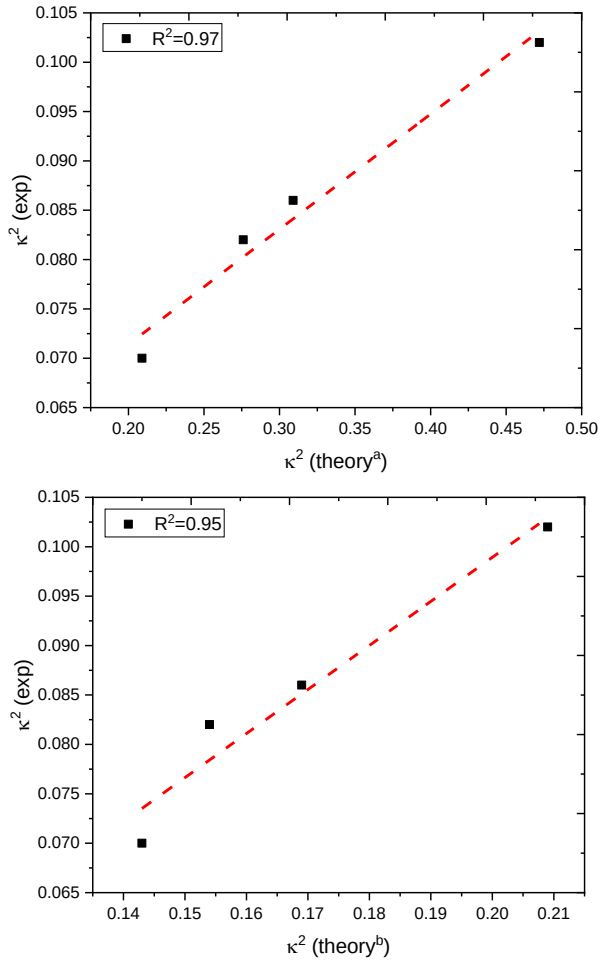


Figure 6.19: Correlation plot of theoretical and experimental dipole orientation factor. Top: Computed using average dipoles. Bottom: Computed by averaging κ^2 at each snapshot.

6.7 Appendix D: Free-energy differences between different protonation states

Table 6.12: Gibbs free-energy differences (in eV) and relative populations of different protonation states at room temperature (298 K) and at 100 K (numbers in parenthesis).

System	HIS148	ΔG_{mm}	HIS148	ΔG (mm $\rightarrow$ qm/mm)	HIS148	ΔG	HIS148	popul.
EGFP	HSP$\rightarrow$HSD	0.085 (-0.033)	HSP	-4.360 (-1.452)	HSP$\rightarrow$HSD	-0.049 (-0.008)	HSP	0.129 (0.268)
	HSP$\rightarrow$HSE	0.064 (-0.023)	HSD	-4.494 (-1.427)	HSP$\rightarrow$HSE	0.125 (0.014)	HSD	0.871 (0.678)
	HSD$\rightarrow$HSE	– –	HSE	-4.299 (-1.415)	HSD$\rightarrow$HSE	0.174 (0.022)	HSE	– (0.052)
T65G	HSP$\rightarrow$HSD	0.067 (-0.026)	HSP	-4.691 (-1.601)	HSP$\rightarrow$HSD	-0.046 (-0.005)	HSP	0.143 (0.339)
	HSP$\rightarrow$HSE	0.029 (-0.017)	HSD	-4.804 (-1.625)	HSP$\rightarrow$HSE	0.180 (0.015)	HSD	0.857 (0.602)
	HSD$\rightarrow$HSE	– –	HSE	-4.540 (-1.569)	HSD$\rightarrow$HSE	0.226 (0.020)	HSE	– (0.059)
Duo	HSP$\rightarrow$HSD	0.045 (0.019)	HSP	-3.832 (-1.297)	HSP$\rightarrow$HSD	0.194 (0.106)	HSP	0.026 (0.010)
	HSP$\rightarrow$HSE	0.020 (-0.011)	HSD	-3.683 (-1.210)	HSP$\rightarrow$HSE	-0.093 (-0.040)	HSD	— (–)
	HSD$\rightarrow$HSE	– –	HSE	-3.945 (-1.326)	HSD$\rightarrow$HSE	-0.101 (-0.146)	HSE	0.974 (0.990)
Trio	HSP$\rightarrow$HSD	0.058 (0.017)	HSP	-3.944 (-1.301)	HSP$\rightarrow$HSD	-0.071 -0.056	HSP	0.018 (0.001)
	HSP$\rightarrow$HSE	0.026 (-0.009)	HSD	-4.073 (-1.374)	HSP$\rightarrow$HSE	- 0.094 -0.076	HSD	0.285 (0.089)
	HSD$\rightarrow$HSE	– –	HSE	-4.064 (-1.368)	HSD$\rightarrow$HSE	-0.023 -0.020	HSE	0.697 (0.910)

Tables 6.12 and 6.13 show computed free energy differences at room temperature (T=298 K) and at T=100 K.

- The entropic factor favors HSD and HSE over HSP: entropy is increasing upon deprotonation, because the number of particles increases. This explains the increase of the HSP fraction at low temperatures for all 4 systems. The HSD-HSP ΔG shows much weaker temperature dependence.

- Seems like distinct populations survive at low T, which is what experiment shows.

Table 6.13: Free-energy differences (in eV) between different protonation states at room temperature (298 K) and at 100 K (numbers in parenthesis), difference in enthalpy (ΔH in eV) and entropy (ΔS in eVK^{-1}).

Mutant	HIS148	ΔG(298K)	ΔG(100K)	ΔH	ΔS	-TΔS(298K)	-TΔS(100K)
EGFP	HSP→HSD	-0.049	-0.008	0.013	2.07×10^{-4}	-0.061	-0.021
	HSP→HSE	0.125	0.014	-0.042	-5.61×10^{-4}	0.167	0.056
	HSD→HSE	0.174	0.022	-0.055	-7.68×10^{-4}	0.229	0.077
T65G	HSP→HSD	-0.046	-0.005	0.016	2.07×10^{-4}	-0.061	-0.021
	HSP→HSE	0.180	0.015	-0.068	-8.33×10^{-4}	0.248	0.083
	HSD→HSE	0.226	0.020	-0.084	-1.04×10^{-3}	0.310	0.104
Duo	HSP→HSD	0.194	0.106	0.061	-4.44×10^{-4}	0.132	0.044
	HSP→HSE	-0.093	-0.040	-0.013	2.68×10^{-4}	-0.080	-0.027
	HSD→HSE	-0.101	-0.146	-1.68	-2.27×10^{-3}	0.676	0.227
Trio	HSP→HSD	-0.071	-0.056	-0.048	7.57×10^{-5}	-0.022	-0.007
	HSP→HSE	- 0.094	-0.076	-0.067	9.09×10^{-5}	-0.027	-0.009
	HSD→HSE	-0.023	-0.020	-0.018	1.51×10^{-5}	-0.004	-0.001

Analysis of the computed free energies and entropy:

- If the reaction is assumed to be it gas phase, one may expect ΔS to be largely positive in reaction HSP$\rightarrow$HSD/HSE making free energy change more negative. This is due to creation of two particles (one H^+ from one leading to increase in degree of disorderness.

- However, the situation is much more complex in solution and in condense phase. The effect of created H^+ is expected to be minimized due to solvation.

- Enthalpy-entropy compensation (EEC) is widely accepted for playing a key role in protein-ligand binding, protein-protein interaction, solvation in proteins in water etc. This was previously understood as explained by assuming that if a molecular change in the ligand leads to more and/or tighter van der Waals contacts and H-bonds with the substrate (giving a more negative H), this inevitably leads to reduced mobility/ flexibility in either or both components of the interaction, i.e., a reduction in the overall conformational entropy, and that change compensates the enthalpy decrease.

- However, a recent study shows, in macromolecules, a little change in Gibbs free energy and and the large changes in enthalpy and entropy are too great to be a consequence of only conformational changes.it does not follow that conformational changes are the sole contributor to the entropy: overall protein flexibility should be considered to be a greater contributor to change in entropy.

- To check, if EEC exists in EGFP and mutants, we will plot ΔH and -TΔS at 298 K and 100 K to check if they correlates.

- There exists linear correlation in ΔH and -TΔS at 298K and 100K for each mutants. However, trends are not similar for all of them. Trio behaves differently compared to rest of the mutants.

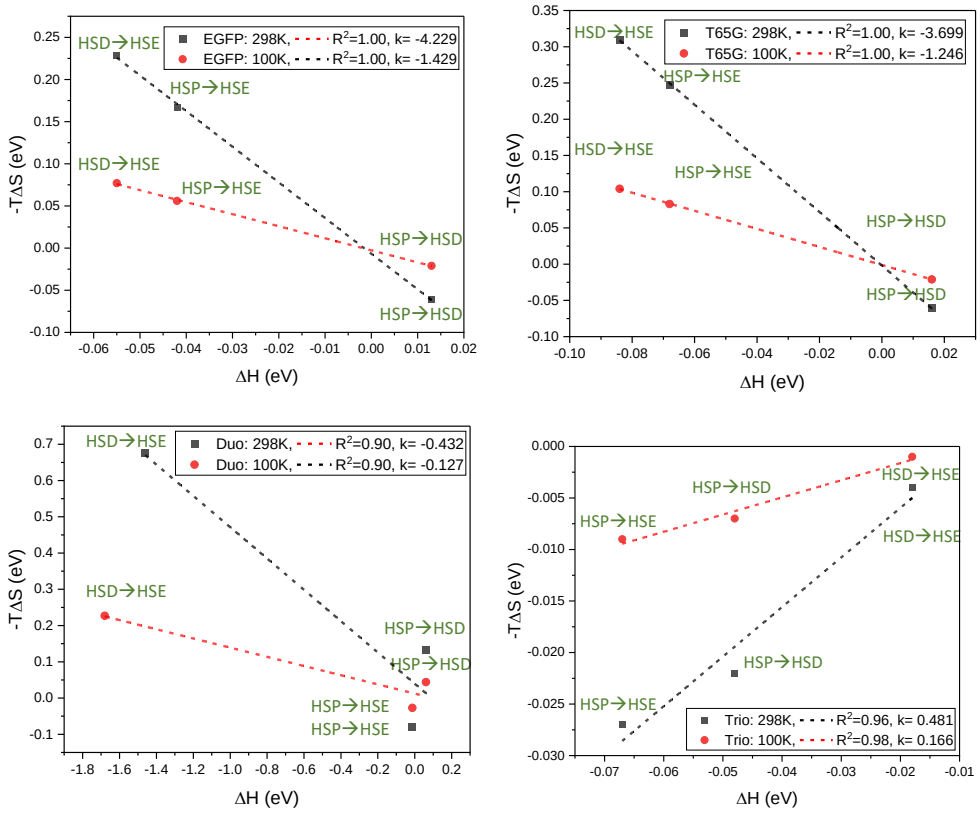


Figure 6.20: Plots of ΔH versus -TΔS in different protonation states of mutants.

- In EGFP, HSP→HSD shows a +ve entropy (as expected), whereas the reaction is also seen to be endothermic (+ve). At room temperature, entropy term dominates over enthalpy. As the temp goes down, enthalpy term starts being dominant changing the favorable protonation state at low temperature. This explains, why population of HSD decreases at 100K compared to 298K.

- For Trio HSP→HSE shows a +ve entropy and the reaction is also seen to be exothermic (-ve). Therefore, here entropy and enthalpy are not competing with each other. Rather it is complementing. At low temperature, HSE remain the dominant population due to large -ve enthalpy. This explains why population of HSE increases upon lowering the temperature.

Using ΔH and ΔS we have extrapolated ΔG at different temperatures which shows change in dominant population at low temperature for EGFP and T65G whereas no change in observed for Duo and Trio; this is shown in Figs. 6.21 and 6.22.

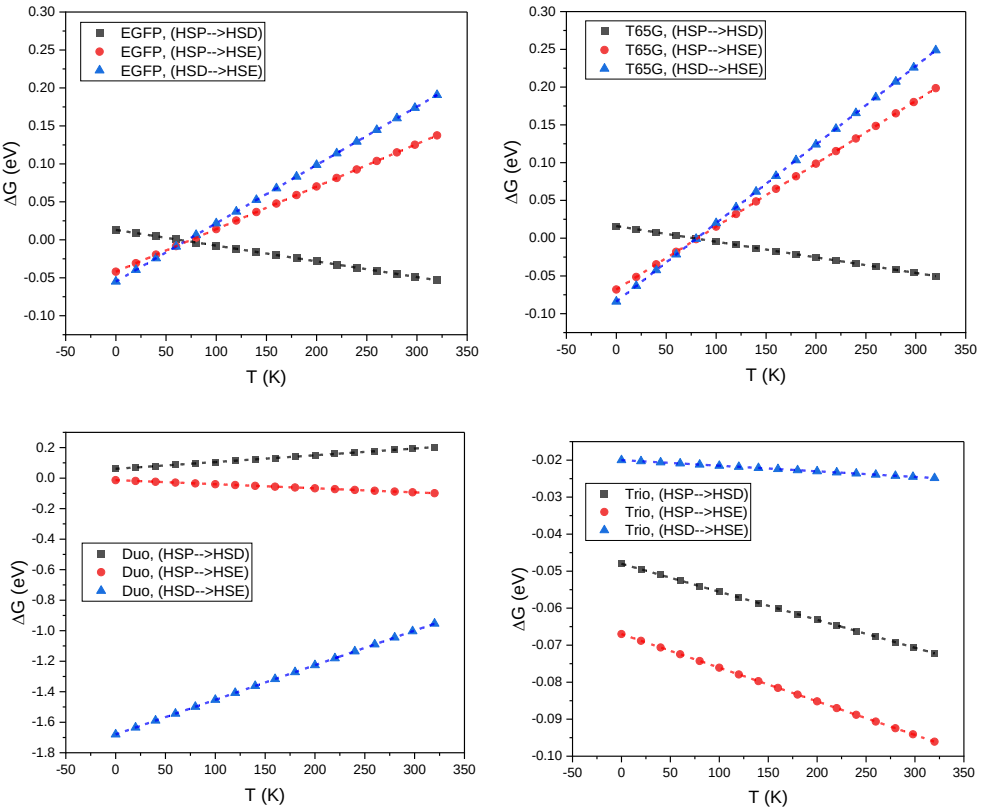


Figure 6.21: Extrapolation of ΔG with respect to temperature in mutants.

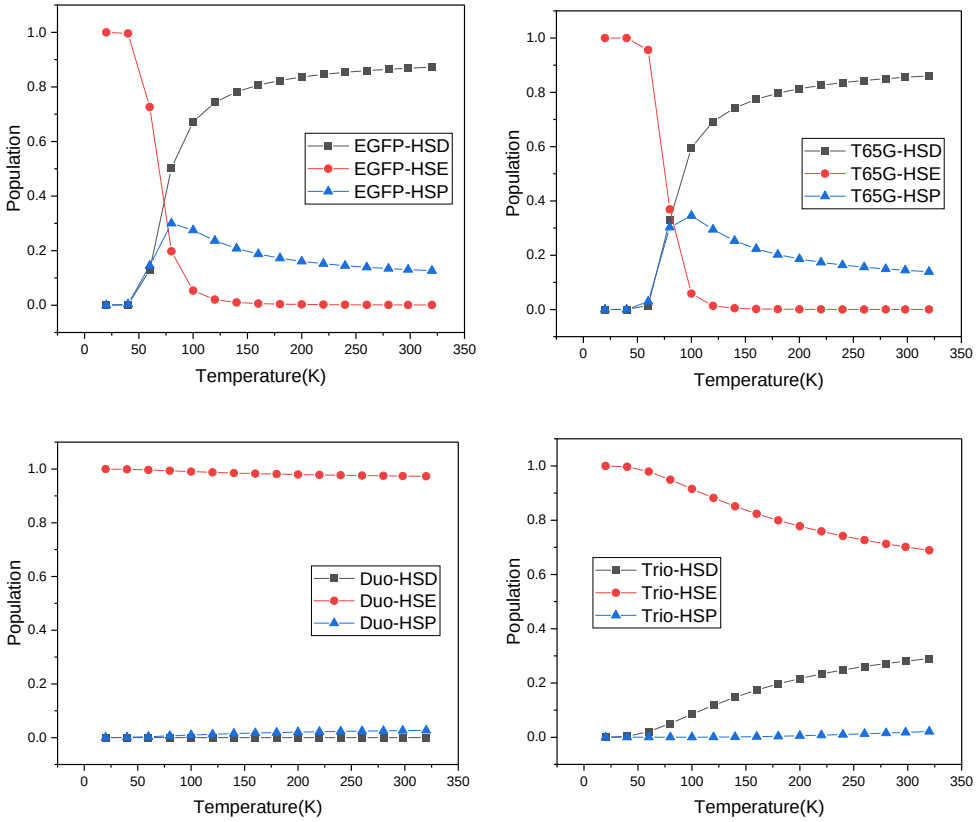


Figure 6.22: Extrapolation of population of different protonation state with respect to temperature in mutants.

6.7.1 Appendix E: AIMD results

Table 6.14 compares twisting times in AIMD trajectories compared with the MD simulations initiated from the same snapshot (same structures but different velocities; see section 6.4.4). Specifically, we record time at which phenolate ring twists (defined as $\phi > 30°$). Fig 6.23 shows this comparison graphically. We observe very close correlation between the MD and AIMD trajectories, which provides validation for our modified fore-field parameters.

Table 6.14: AIMD simulation in 1st excited state for 3 ns showing the twist around ϕ in 11 snapshots for EGFP and BrUSLEE. Time of the twist of the same snapshot in excited-state MD is shown in parenthesis (in ns).

Snapshot	EGFP-HSD	EGFP-HSE	EGFP-HSP	Trio-HSD	Trio-HSE	Trio-HSP
0	2.44 (–)	1.93(1.55)	0.29 (0.56)	0.67 (0.52)	2.27 (–)	0.78 (0.98)
40	2.38 (2.27)	1.00 (0.77)	1.63 (1.47)	2.36 (–)	2.09 (2.26)	1.03 (1.15)
80	– (–)	2.28 (2.06)	2.61 (–)	0.69 (0.93)	2.40 (–)	2.88 (2.32)
120	2.08 (–)	1.14 (1.57)	2.66 (2.57)	2.57 (–)	2.65 (–)	2.79 (–)
160	1.42 (1.65)	0.62 (0.34)	0.74 (0.46)	2.83 (–)	0.19 (0.06)	2.21 (–)
200	1.27 (0.93)	1.55 (1.89)	– (–)	– (2.70)	1.63 (2.14)	1.13 (1.63)
240	2.77 (–)	2.06 (2.28)	2.70 (2.17)	0.39 (0.04)	2.23 (–)	2.06 (1.43)
280	1.69 (1.81)	0.43 (0.60)	1.06 (0.92)	2.31 (–)	2.27 (2.62)	1.85 (1.42)
320	2.36 (–)	1.19 (1.15)	0.28 (0.03)	– (2.98)	0.63 (0.41)	1.33 (0.97)
360	2.41 (–)	1.11 (0.77)	0.84 (1.25)	0.69 (0.71)	0.72 (0.50)	2.16 (–)
400	1.33 (0.83)	2.23 (2.32)	0.79 (0.81)	0.56 (0.37)	1.37 (1.57)	0.97 (0.61)

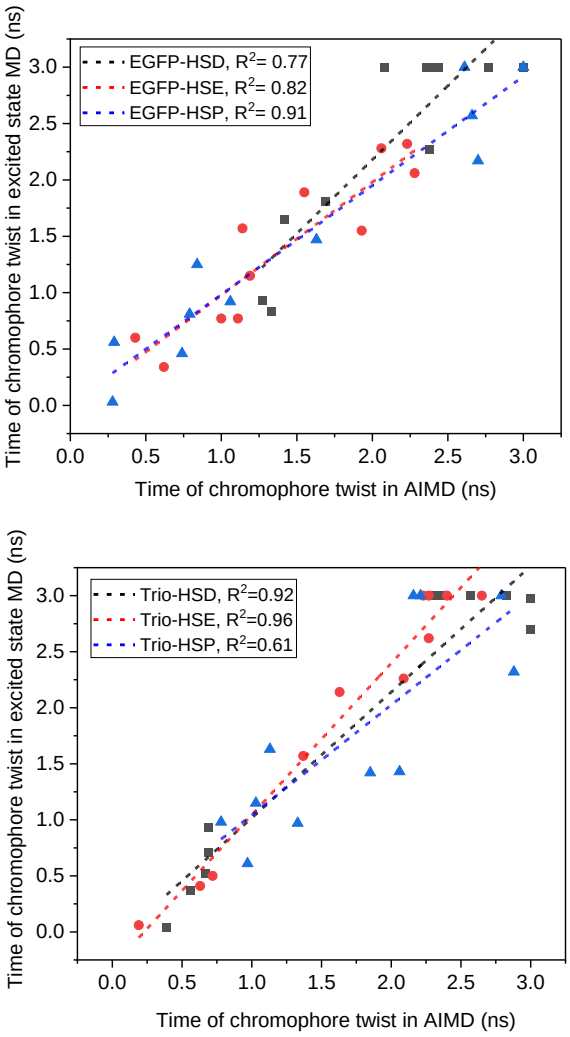


Figure 6.23: Correlation plot of twisting time for MD and AIMD excited-state trajectories initiated from 11 snapshots for each protonation state of His148 of EGFP and BrUSLEE.

6.8 Appendix F: Calculations of radiative and radiationless lifetimes

Radiative lifetimes are computed using Eq. (6.11). For the protein-bound chromophore, we use n=1.6 (as in Ref. 26). Radiationless lifetimes were computed from the excited-state MD by using linear fit of the planar population decay. Then we compute apparent excited-state lifetime as for each form as:

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}} \tag{6.20}$$

$$\tau = \frac{\tau_{nr}\tau_r}{\tau_r + \tau_{nr}} \tag{6.21}$$

$$\tag{6.22}$$

For example, for the HSD form:

$$\tau_{HSD} = \frac{\tau_{nr,HSD}\tau_{r,HSD}}{\tau_{r,HSD} + \tau_{nr,HSD}}. \tag{6.23}$$

Then we use the following procedure to compute apparent excited-state lifetimes and FQY averaged over distinct populations:

$$\langle\tau\rangle = \sum_i A_i\tau_i \tag{6.24}$$

In this case,

$$\langle\tau\rangle = A_{HSD}\tau_{HSD} + A_{HSE}\tau_{HSE} + A_{HSP}\tau_{HSP} \tag{6.25}$$

For FQY:

$$FQY = \frac{\tau_{nr}}{\tau_r + \tau_{nr}} \tag{6.26}$$

For HSD:

$$FQY_{HSD} = \frac{\tau_{nr,HSD}}{\tau_{r,HSD} + \tau_{nr,HSD}} \tag{6.27}$$

$$\langle FQY \rangle = \sum_i A_i FQY_i \tag{6.28}$$

In this case,

$$\langle FQY \rangle = A_{HSD} FQY_{HSD} + A_{HSE} FQY_{HSE} + A_{HSP} FQY_{HSP} \tag{6.29}$$

Table 6.15: Theoretical estimates of radiative lifetime for different mutants. Computed excitation energies and oscillator strengths are also shown. QM/MM absorption energies and oscillator strengths are averaged over 400 snapshots taken from ground-state equilibrium MD simulations. τ_r,rel values are relative lifetimes calculated with respect to τ_r in EGFP-HSD.

Mutant	HIS148	E_{ex}, eV (f_l) (gas)	E_{ex},eV (f_l) (QM/MM)	τ_r, ns (gas, n=1)	τ_r, ns (QM/MM, n=1.6)	τ_r rel, ns
EGFP	HSD	3.101 (1.02)	3.026 (1.06)	29.50	7.254	1.00
	HSE	3.101 (1.02)	3.046 (1.07)	28.25	7.091	0.98
	HSP	3.101 (1.02)	3.019 (1.06)	28.25	7.286	1.00
T65G	HSD	3.123 (1.05)	3.079 (1.14)	28.25	6.496	0.89
	HSE	3.123 (1.05)	3.062 (1.09)	28.25	6.778	0.93
	HSP	3.123 (1.05)	3.067 (1.13)	28.25	6.642	0.91
Duo	HSD	3.123 (1.05)	3.055 (1.14)	28.25	6.622	0.91
	HSE	3.123 (1.05)	3.058 (1.14)	28.25	6.616	0.91
	HSP	3.123 (1.05)	3.096 (1.12)	28.25	6.581	0.91
Trio	HSD	3.123 (1.05)	3.079 (1.13)	28.25	6.577	0.91
	HSE	3.123 (1.05)	3.128 (1.15)	28.25	6.242	0.86
	HSP	3.123 (1.05)	3.061 (1.13)	28.25	6.660	0.92

Table 6.16: Computed non-radiative decay times (ns), populations of different protonation state of His148, and % of non-planar chromophores at the end of the excited-state simulation (3 ns).

Mutant	HIS148	τ_{nr}	population	%planar conformation
EGFP	HSD	8.60	0.871	76.25
	HSE	1.87	—	32.50
	HSP	2.96	0.129	47.50
T65G	HSD	0.98	0.857	14.00
	HSE	0.69	—	3.00
	HSP	2.66	0.143	41.50
Duo	HSD	1.41	—	22.75
	HSE	0.47	0.974	0.00
	HSP	1.78	0.026	32.75
Trio	HSD	1.69	0.285	29.75
	HSE	0.97	0.697	12.00
	HSP	1.87	0.018	34.75

Table 6.17: Computed values of average lifetime (in ns), percentage population of each protonation states, and fluorescent quantum yield. Experimental values are given in parenthesis.

Mutant	HIS148	τ, theory (exp)	population, theory (exp)	$\langle\tau\rangle$	FQY	$\langle FQY\rangle$, theory (exp)
EGFP	HSD	3.93 (2.8)	0.871 (0.887)	3.69 (2.71)	0.54	0.51 (0.60)
	HSE	1.48	—		0.21	
	HSP	2.10 (2.0)	0.129 (0.113)		0.29	
T65G	HSD	0.85 (0.82)	0.857 (0.885)	1.00 (0.95)	0.13	0.15 (0.10)
	HSE	0.63	—		0.09	
	HSP	1.90 (2.0)	0.143 (0.115)		0.28	
Duo	HSD	1.16	—	0.46 (0.61)	0.17	0.07 (0.08)
	HSE	0.44 (0.52)	0.974 (0.91)		0.07	
	HSP	1.40 (1.5)	0.026 (0.09)		0.21	
Trio	HSD	1.34 (1.4)	0.285 (0.164)	1.00 (0.65)	0.20	0.15 (0.3)
	HSE	0.85 (0.51)	0.697 (0.833)		0.13	
	HSP	1.46 (2.3)	0.018 (0.003)		0.22	

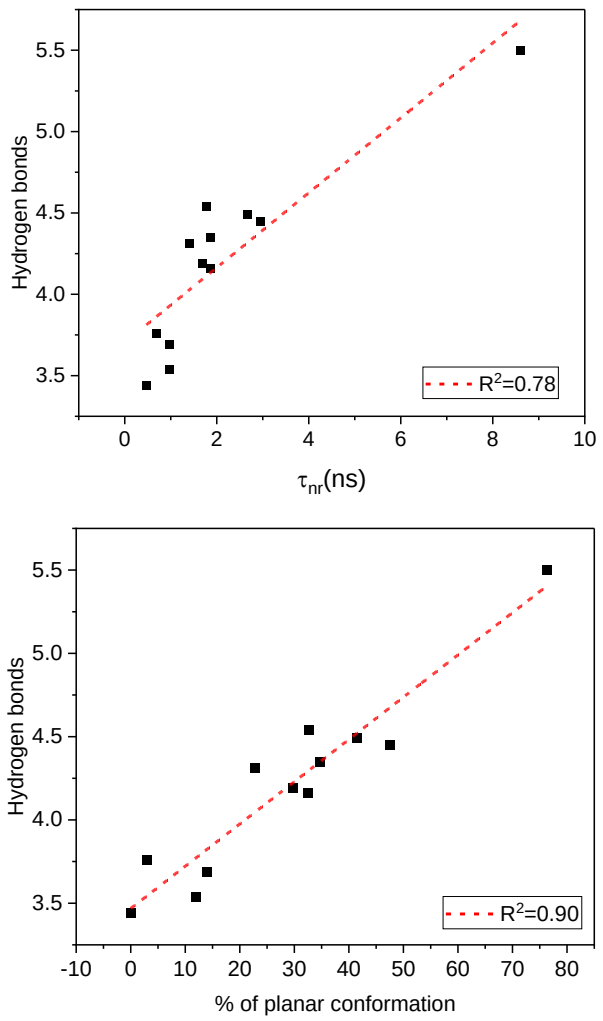


Figure 6.24: Correlation between the average number of hydrogen bonds in the ground state and computed non-radiative lifetime (top) and the % of surviving planar conformation after 3 ns of excited-state dynamics (bottom).

Table 6.18: Computed and experimental values of photophysical parameters EGFP, T65G, Duo, and Trio (in parenthesis, the experimental values are shown).

Protein	Ext. coeff.(exp.)	FQY (exp.)	R.B. (exp.)	FL, ns (exp.)	RPS (exp.)
EGFP	48948.1 (55000)	0.51 (0.6)	1 (1)	3.69 (2.71)	1 (1)
T65G	53668.8 (70000)	0.15 (0.1)	0.32 (0.21)	1.00 (0.95)	3.69 (1.87)
Duo	56774.6 (84500)	0.07 (0.08)	0.16 (0.203)	0.46 (0.61)	8.02 (4.3)
Trio	58157.8 (86000)	0.15 (0.3)	0.35 (0.8)	1.00 (0.65)	3.69 (8.8)

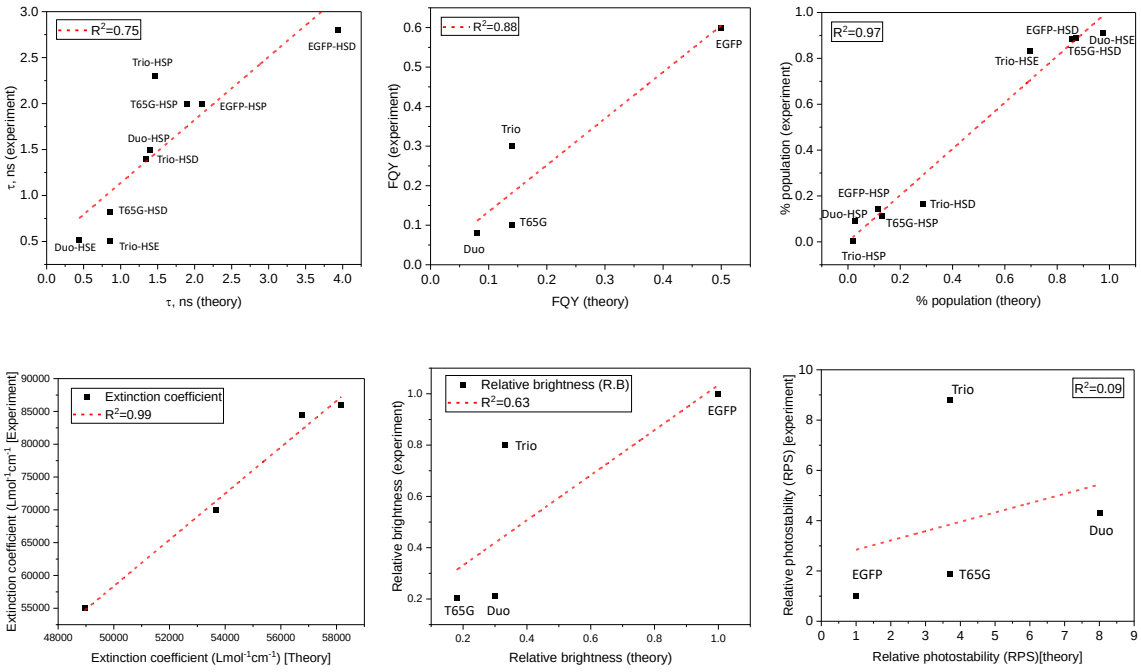


Figure 6.25: Correlation plots in RB, FQY, extinction coefficient, RPS, lifetimes.

6.9 Chapter 6 references

[1] M. Chalfie, Y. Tu, G. Euskirchen, W.W. Ward, D.C. Prasher, Green fluorescent protein as a marker for gene expression, Science **263**, 802 (1994).

[2] D. M. Chudakov, M. V. Matz, S. Lukyanov, K. A. Lukyanov, Fluorescent proteins and their applications in imaging living cells and tissues, Phys. Rev. **90**, 1103 (2010).

[3] K. Lu, C. Q. Vu, T. Matsuda, T. Nagai, Fluorescent protein-based indicators for functional super-resolution imaging of biomolecular activities in living cells, Int. J. Mol. Sci. **20**, 5784 (2019).

[4] E. A. Specht, E. Braselmann, A. E. Palmer, A critical and comparative review of fluorescent tools for live-cell imaging, Annu. Rev. Physiol. **79**, 93 (2017).

[5] N. C. Shaner, G. H. Patterson, M. W. Davidson, Advances in fluorescent protein technology, J. cell Sci. **120**, 4247 (2007).

[6] T. J. Lambert, FPbase: a community-editable fluorescent protein database, Nat. Methods **16**, 277 (2019).

[7] M. A. Mena, T. P. Treynor, S. L. Mayo, P. S. Daugherty, Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library, Nat. Biotech. **24**, 1569 (2006).

[8] P. J. Cranfill, B. R. Sell, M. A. Baird, J. R. Allen, Z. Lavagnino, H. M. De Gruiter, G. Kremers, M. W. Davidson, A. Ustinone, D. W. Piston, Quantitative assessment of fluorescent proteins, Nat. Methods **13**, 557 (2016).

[9] A. Acharya, A. M. Bogdanov, K. B. Bravaya, B. L. Grigorenko, A. V. Nemukhin, K. A. Lukyanov, and A. I. Krylov, Photoinduced chemistry in fluorescent proteins: Curse or blessing?, Chem. Rev. **117**, 758 (2017).

[10] M. Y. Berezin, S. Achilefu, Fluorescence lifetime measurements and biological imaging, Chem. Rev. **110**, 2641 (2010).

[11] T. Ha, P. Tinnefeld, Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging, Annu. Rev. Phys. Chem. **63**, (2012).

[12] T. R. Gosnell, Fundamentals of Spectroscopy and Laser Physics, Camb. Univ. Press 3 (2002).

[13] M. L. Markwardt, G. Kremers, C. A. Kraft, K. Ray, P. Jc. Carnfill, K. A. Wilson, R. N. Day, R. M. Wachter, M. W. Davidson, M. A. Rizzo, An improved cerulean fluorescent protein with enhanced brightness and reduced reversible photoswitching, PloS one **6** e17896 (2011).

[14] M. Erard, A. Fredj, H. Pasquier, D. Beltolngar, Y. Bousmah, V. Derrien, P. Vincent, F. Merola, Minimum set of mutations needed to optimize cyan fluorescent proteins for live cell imaging, Mol. Biosys. **9** 258 (2013).

[15] J. Goedhart, D. Von Stetten, M. Noirclerc-Savoye, M. Lelimousin, L. Joosen, M. A. Hink, L. Van Weeren, T. WJ. Gadella, A. Royant, Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%, Nat. Comm. **3**, 1 (2012).

[16] D. S. Bindels, L. Haarbosch, L. Van Weeren, M. Postma, K. E. Wiese, M. Mastop, S. Aumonier, G. Gotthard, A. Royant, M. A. Hink and others, mScarlet: a bright monomeric red fluorescent protein for cellular imaging, Nat. Methods **14**, 53 (2017).

[17] K. S. Sarkisyan, A. S. Goryashchenko, P. V. Lidsky, D. A. Gorbachev, N. G. Bozhanova, A. Y. Gorokhovatsky, A. R. Pereverzeva, A. P. Ryumina, V. V. Zherdeva, A. P. Savitsky and others, Green fluorescent protein with anionic tryptophan-based chromophore and long fluorescence lifetime, Biophys. Jour. **109**, 380 (2015).

[18] K. D. Piatkevich, V. N. Malashkevich, K. S. Morozova, N. A. Nemkovich, S. C. Almo, V. V. Verkhusha, Extended stokes shift in fluorescent proteins: chromophore–protein interactions in a near-infrared tagrfp675 variant, Scientific Reports **3**, 1 (2013).

[19] A. Hense, B. Prunsche, P. Geo, Y. Ishitsuka, K. Nienhaus, G. U. Nienhaus, Monomeric Garnet, a far-red fluorescent protein for live-cell STED imaging, scientific Reports **5**, 1 (2015).

[20] G. Matela, P. Gao, G. Guigas, A. F. Eckert, K. Nienhaus, G. U. Nienhaus, A far-red emitting fluorescent marker protein, mGarnet2, for microscopy and STED nanoscopy, Chem. Comm. **53**, 979 (2017).

[21] S. Karasawa, T. Araki, T. Nagai, H. Mizuno, A. Miyawaki, Cyan-emitting and orange-emitting fluorescent proteins as a donor/acceptor pair for fluorescence resonance energy transfer, Biochemical Journal **381**, 307 (2004).

[22] H. Tsutsui, S. Karasawa, Y. Okamura, A. Miyawaki, Improving membrane voltage measurements using FRET with new fluorescent proteins, Nat. Methods **5**, 683 (2008).

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

[24] R. Heim, Improved green fluorescence, Nature **373**, (1995).

[25] A. M. Bogdanov, A. Acharya, A. V. Titelmayer, A. V. Mamontova, K. B. Bravaya, A. B. Kolomeisky, K. A. Lukyanov, and A. I. Krylov, Turning on and off photoinduced electron transfer in fluorescent proteins by π-stacking, halide binding, and Tyr145 mutations, J. Am. Chem. Soc. **138**, 4807 (2016).

[26] T. Sen, A. V. Mamontova, A. V. Titelmayer, A. M. Shakhov, A. A. Astafiev, A. Acharya, K. A. Lukyanov, A. I. Krylov, and A. M. Bogdanov, Influence of the first chromophore-forming residue on photobleaching and oxidative photoconversion of EGFP and EYFP, Int. J. Mol. Sci. **20**, 5229 (2019).

[27] A. Royant, M. Noirclerc-Savoye, Stabilizing role of glutamic acid 222 in the structure of Enhanced Green Fluorescent Protein, J. Str. Biol. **174**, 385 (2011).

[28] M. H. M. Olsson, C. R. Sondergaard, M. Rostkowski, and J. H. Jensen, PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions, J. Chem. Theory Comput. **7**, 525 (2011).

[29] B. L. Grigorenko, A. V. Nemukhin, I. V. Polyakov, D. I. Morozov, and A. I. Krylov, First-principle characterization of the energy landscape and optical spectra of the green fluorescent protein along A-I-B proton transfer route, J. Am. Chem. Soc. **135**, 11541 (2013).

[30] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. **26**, 1781 (2005).

[31] J.-D. Chai and M. Head-Gordon, Systematic optimization of long-range corrected hybrid density functionals, J. Chem. Phys. **128**, 084106 (2008).

[32] J.-D. Chai and M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom-atom dispersion interactions, Phys. Chem. Chem. Phys. **10**, 6615 (2008).

[33] Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A.T.B.; Wormit, M.; Kussmann, J.; Lange, A.W.; Behn, A.; Deng, J.; Feng, X., et al., Advances in molecular quantum chemistry contained in the Q-Chem 4 program package, Mol. Phys. **113**, 184 (2015).

[34] A. I. Krylov, P. MW. Gill, Q-Chem: an engine for innovation. Wiley Interdisciplinary Reviews: Computational Molecular Science **3**, 317 (2013).

[35] S. Faraji, A. I. Krylov, On the nature of an extended Stokes shift in the mPlum fluorescent protein, J. Phys. Chem. B. **119**, 13052 (2015).

[36] T. H. Rod, U. Ryde, Accurate QM/MM free energy calculations of enzyme reactions: methylation by catechol O-methyltransferase, J. Comput. Chem. **1**, 1240 (2005).

[37] T. H. Rod, U. Ryde, Quantum mechanical free energy barrier for an enzymatic reaction, Physical Review Letters, **94**, 138302 (2005).

[38] A. Warshel, Calculations of enzymic reactions: calculations of pKa, proton transfer reactions, and general acid catalysis reactions in enzymes, Biochemistry, **20**, 3167 (1981).

[39] T. R. Gosnell, Fundamentals of Spectroscopy and Laser Physics, Camb. Univ. Press 3 (2002).

[40] D. B. Hand, The refractivity of protein solutions, J. Biol. Chem. **108**, 703 (1935).

[41] T. L. MacMeekin, M. L. Merton, N. J. Hipp, Refractive indices of amino acids, proteins, and related substances, Advances in Chemistry **44**, 54 (1964).

[42] M. de Wergifosse, C. G. Elles, A. I. Krylov, Two-photon absorption spectroscopy of stilbene and phenanthrene: Excited-state analysis and comparison with ethylene and toluene, J. Chem. Phys. **146**, 174102 (2017).

[43] K. Takaba, Y. Tai, H. Eki, H.-A. Dao, Y. Hanazono, K. Hasegawa, K. Mikia, and K. Takeda, Subatomic resolution x-ray structures of green fluorescent protein, IUCrJ **6**, 387 (2019).

[44] M. Taniguchi, H. Du, J. S. Lindsey, PhotochemCAD 3: diverse modules for photophysical calculations with multiple spectral databases, Photochem. Photobiol. **94**, 277 (2018).

[45] A. V. Mamontova, A. M. Shakhov, K. A. Lukyanov, A. M. Bogdanov, Deciphering the Role of Positions 145 and 165 in Fluorescence Lifetime Shortening in the EGFP Variants, Biomolecules **10**, 1547 (2020).

[46] A. V. Mamontova, I. D. Solovyev, A. P. Savitsky, A. M. Shakhov, K. A. Lukyanov, A. M. Bogdanov, Bright GFP with subnanosecond fluorescence lifetime, Scientific reports **8**, 1 (2018).

Chapter 7: Future Work

The thesis focuses on excited-state photophysics of FPs. We aimed to understand the effect of mutations on radiative and nonradiative relaxation process, which control properties such as excited-state lifetimes, FQY, RB, *e*tc. However, important phenomena such as photobleaching remain largely unexplored. One important example is redding: EGFP is known to undergo oxidative redding in presence of an external oxidants[1]. Understanding such processes is important for understanding photostability

7.1 Understanding the photostability in EGFP mutants

For a given rate of the bleaching process (via photo-oxidation or other photochemical processes), the yield of the bleached forms is smaller for systems with shorter apparent excited-state lifetimes. As photostability is inversely proportional to Y_{bl}, the ratios of $1/Y_{bl}$ can be interpreted as relative photostabilities. Bleaching rates can vary significantly among different proteins, because electron-transfer (ET) pathways and the rates are sensitive to mutations[1,2]. Because of the high cost of such calculations, the effects of mutations on the rates of electron transfer are not fully investigated in the present thesis.

Within the first-order kinetics, the yield of bleaching Y_{bl} is given by:

$$Y_{bl} = \frac{\tau}{\tau_{bl}}, \tag{7.1}$$

where τ is excited-state lifetime, τ_{bl} is related to rate of electron transfer processes. In our study in Chapter 4 and 6, we computed Y_{bl} as a ratio of lifetimes of mutants to that in EGFP.

Table 7.1: Computed and experimental values of relative photobleaching rate (relative to EGFP). Experimental values are in PBS+Ox. (Reproduced from Chapter 4).

Protein	relative photostability (theory)	relative photostability (exp.)
EGFP	1.0	1.0
EGFP-T65G	10.0	17.0
EYFP	2.5	0.4
EYFP-G65T	0.71	6.4

Table 7.1, shows that we fail to achieve qualitative agreement in computed and experimental photostability in mutants[4]. This suggests that indeed ET pathways and the rates are sensitive to mutations.

The oxidative redding in EGFP can be expressed as a series of reactions as follows[1,2]:

$$Chro^{-} \xrightarrow{h\nu} Chro^{*} \xrightarrow{fast,-1e} Chro^{\cdot} \xrightarrow{slow,Chemistry} Redform \tag{7.2}$$

A detailed mechanistic study by Bogdanov *e*t al. focused on the initial step; ET from the chromophore (electron donor) to a nearby aromatic residue (electron acceptor). The rate of ET between different sites are given by Marcus rate expression[5,6]:

$$k_{ET} = |H_{DA}|^2 \frac{1}{\sqrt{4\pi\lambda k_B T}} exp\left\{ -\frac{(\Delta G + \lambda)^2}{4\lambda K_B T} \right\} \tag{7.3}$$

where ΔG, λ and H_{DA} are the free energy change, reorganization energy, and coupling between the electronic states involved in ET. Computation of rate of ET in different mutants, as discussed in Chapter 4 and 6, may provide a new approach towards enhancing phtostability of FPs.

7.2 Exploring the role of a triplet state in oxidative photochemistry in EGFP

Two recently discovered phenomena in EGFP have generated considerable interest. The first one is oxidative redding, which was observed in 2009 by Bogdanov *et al*[7]. The other one is primed photoconversion, which was observed in 2015 by Dempsey *et al.*[8] in some photoconvertible FPs such as Dendra2. Both phenomena have in common that, upon absorption of light, the green chromophore changes its color to red. This happens with low quantum yield. For oxidative redding, the authors favor an excited singlet precursor developing into a radical state that further reacts in the dark although the possibility of the involvement of a triplet state was also briefly mentioned[9]. In contrast, for the primed conversion, Mohr *e*t al.[10] proposed a triplet state with millisecond lifetime that absorbs a second photon forming a higher excited triplet state, serving as a doorway for further chemical transformation. Formation of the triplet state was characterized by phosphorescence emission (time-resolved phosphorescence spectroscopy) and an intermediate with 5 ms lifetime was observed in transient absorption (TA)[11].

The proposed mechanism of that process involves an $S_1 \rightarrow T_1$ intersystem crossing. The chromophore then ends up in a low lying triplet state. In presence of an oxidant, ET transfer takes place from the chromophore to the oxidant. However, very little is known about the

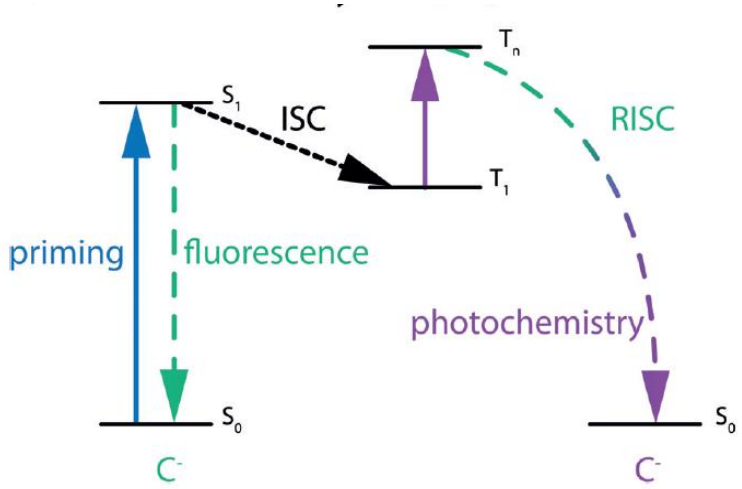


Figure 7.1: Summary of the mechanism of primed conversion: 488 nm excitation or priming of the anionic cis chromophore, C^-. populates the $S_1(C^-)$ state. De-population of the $S_1(C^-)$ state may occur via (i) fluorescence emission or (ii) low-yield intersystem crossing to the lowest triplet state, T_1. Excitation of T_1 with the red conversion beam causes a T_1T_n transition. The ensuing relaxation process to the singlet ground state involves reverse intersystem crossing (RISC) and excited state chemical transformation generating the red species. (Reproduced from Ref. 10)

electronic structure of this triplet state, and the entire mechanism of the process of redding via primed photoconversion.

Because primed photoconversion is used as an alternate way of green to red conversion of EGFP, a mechanistic study of that process would help to understand the photophysics of FPs and hopefully, design better FPs for super resolution imaging.

7.3 Chapter 7 references

[1] A. M. Bogdanov, A. Acharya, A. V. Titelmayer, A. V. Mamontova, K. B. Bravaya, A. B. Kolomeisky, K. A. Lukyanov, A. I. Krylov, Turning on and off photoinduced electron transfer in fluorescent proteins by π-stacking, halide binding, and Tyr145 mutations, J. Am. Chem. Soc. bf 138, 4807 (2016).

[2] A. Acharya, A. M. Bogdanov, K. B. Bravaya, B. L. Grigorenko, A. V. Nemukhin, K. A. Lukyanov, and A. I. Krylov, Photoinduced chemistry in fluorescent proteins: Curse or blessing?, Chem. Rev. **117**, 758 (2017).

[3] R.Y. Tsien, The green fluorescent protein, Annu. Rev. Biochem. **67**, 509 (1998).

[4] T. Sen, A. V. Mamontova, A. V. Titelmayer, A. M. Shakhov, A. A. Astafiev, A. Acharya, K. A. Lukyanov, A. I. Krylov, and A. M. Bogdanov, Influence of the first chromophore-forming residue on photobleaching and oxidative photoconversion of EGFP and EYFP, Int. J. Mol. Sci. **20**, 5229 (2019).

[5] R. A. Marcus, On the theory of oxidation-reduction reactions involving electron transfer. I, J. Chem. Phys. **24**, 966 (1956).

[6] R. A. Marcus, Chemical and electrochemical electron-transfer theory, Annu. Rev. Phys. Chem. **15**, 155 (1964).

[7] A. M. Bogdanov, A. S. Mishin, I. V. Yampolsky, V. V. Belousov, D. M. Chudakov, F. V. Subach, V. V. Verkhusha, S. Lukyanov, K. A. Lukyanov, Green fluorescent proteins are light-induced electron donors, Nat. Chem. Biol. **5**, 459 (2009).

[8] W. P. Dempsey, L. Georgieva, P. M. Helbling, A. Y. Sonay, T. V. Truong, M. Haffner, P. Pantazis, In vivo single-cell labeling by confined primed conversion, Nat. Method **12**, 645 (2015).

[9] R. B. Vegh, K. B. Bravaya, D. A. Bloch, A. S. Bommarius, L. M. Tolbert, M. Verkhovsky, A. I. Krylov, K. M. Solntev, Chromophore photoreduction in red fluorescent proteins is responsible for bleaching and phototoxicity, J. Phys. Chem. B. **118**, 4527 (2014).

[10] M. A. Mohr, A. Y. Kobitski, L. R. Sabater, K. Nienhaus, C. J. Obara, J. Lippincott-Schwartz, G. Nienhaus, P. Pantazis, Rational Engineering of Photoconvertible Fluorescent Proteins for Dual-Color Fluorescence Nanoscopy Enabled by a Triplet-State Mechanism of Primed Conversion, Angew. Chem. Int. Ed. **56**, 11628 (2017).

[11] M. Byrdin, C. Duan, D. Bourgeois, K. Brettel, A long-lived triplet state is the entrance gateway to oxidative photochemistry in green fluorescent proteins, J. Am. Chem. Soc. **140**, 2897 (2018).