Theoretical study on the photoswitching mechanism of negative reversibly photoswitchable fluorescent proteins

by

Bruno Torcal Embeita

Supervisors:
Dr. Marius Wanko
Prof. Angel Rubio

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy
in the
Faculty of Physics and Materials Science

University of the Basque Coutry

Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

image © Rob Rey

September 2016

Universidad del País Vasco
Euskal Herriko Unibertsitatea

University of the Basque Country

Theoretical study on the photoswitching mechanism of negative reversibly photoswitchable fluorescent proteins

by

Bruno Torcal Embeita

Supervisors:
Dr. Marius Wanko
Prof. Angel Rubio

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
Faculty of Physics and Materials Science

September 2016

Abstract

Reversibly photoswitchable fluorescent proteins (RSFPs) are genetically engineered proteins that can be switched by light absorption between a fluorescent ON state and a dark OFF state. Among other applications they allow to increase the resolution beyond the diffraction limit in cell imaging by fluorescence microscopy. RSFPs have extended the possibilities of fluorescence microscopy and other biotechnological tools, but the development of their properties is still far from being rationally designed. Thus, there might be much room for improvement if we manage to understand the switching mechanisms. The switching mechanisms in several negative RSFPs is being elucidated but still under debate. In this thesis I extend the theoretical knowledge about the photoswitching of negative RSFPs by studying the excited-state potential energy surface of both the ON- and the OFF-state. I compare three RSFPs, namely IrisFP, Dronpa and a fast switching single mutant of Dronpa called Dronpa2 to search for the origin of their different switching quantum yields in the ON-state. For the OFF-state, the results of the combined quantum mechanics–molecular mechanics (QM/MM) calculations show that chromophore photoisomerization happens in its neutral form and is followed by ground state deprotonation. This is in agreement with a very recent ultrafast absorption spectroscopy study for IrisFP, and studies on Dronpa and other negative RSFPs. Although the experimental results in both proteins show the same steps, I found that they have different processes at the atomic level due to structural and electrostatical differences, but leading to the same intermediates. This is in contrast to the ON-state where I get the same picture for the three proteins studied, identifying the conical intersection that quenches the fluorescence and controls the photoswitching quantum yield. The major difference between the three proteins in terms of fluorescence and photoswitching characteristics comes from the different sterical environment produced by the residue 159, which is a methionine in different isomers in IrisFP and Dronpa and a smaller threonine that allows a faster isomerization of the chromophore in Dronpa2.

Acknowledgements

Firs of all I would like to thank my supervisors. Marius Wanko, from whom I have learned a lot and who gave me priceless help until the very end steps of this thesis. (Also with my horrible english, noticeable in this acknowledgements that no one corrected). Angel Rubio, for the opportunity to work in this fabulous group and the support. To all my friends who where concerned with the big effort necessary to accomplish this project, and specially to the ones that helped me to keep going on and maintain a little bit of mental health in the last months. Juan, your push and your cooking where decisive. Ohiana, your emotional support with the "issues external to the thesis" where vital. Basket-pote people, Quim, roommates, cuadrilla friends, friends of friends and even sisters of friends. Miguel, for the "better finished than perfect" insight. To all of you thank you. Finally I would like to thank my family, specially my parents for the unconditional support they always give me no matter what I do; to my cousin Clara and her little Juli and Titin, very concerned with how many pages was I writing; and of course Alvarito, a brother, a friend. Ah! and I don't have to forget to thank the deadlines! without whom non of this would have been possible. (I'm probably forgetting many important people, but I know that you know me and you know how important are you for me).

List of Publications

- Theoretical study on the photoswitching mechanism of Dronpa and IrisFP B. Torcal–Embeita, A. Rubio, M. Wanko (under review)
- What limits reversible photoswitching? A comparative theoretical study of the ON-state of Dronpa, Dronpa2, and IrisFP. B. Torcal–Embeita, A. Rubio, M. Wanko (preparing for submission)

Contents

Abstract iii

Acknowledgements iv

List of Publications v

List of Figures xi

List of Tables xvii

Abbreviations xix

1 Introduction 1
- 1.1 Fluorescent Proteins . . . 1
- 1.2 Reversibly Photoswitchable Fluorescent Proteins . . . 5
- 1.3 Most Recent Developments . . . 6

2 Theory & Methods 13
- 2.1 Photochemical Processes . . . 13
  - 2.1.1 Radiationless Decay at Conical Intersection . . . 14
  - 2.1.2 Fluorescence . . . 15
- 2.2 Quantum Mechanical Methods . . . 16
  - 2.2.1 Hartree–Fock Theory . . . 18
  - 2.2.2 Density Functional Theory . . . 20
  - 2.2.3 Configuration Interaction and Coupled Cluster . . . 24
  - 2.2.4 Multi-Reference Methods . . . 25
- 2.3 Selected Ground State Methods . . . 27
  - 2.3.1 SCC-DFTB . . . 28
  - 2.3.2 MP2 . . . 28
- 2.4 Selected Excited-State Methods . . . 29
  - 2.4.1 CC2 . . . 29
  - 2.4.2 OM2/MRCI . . . 30
- 2.5 Quantum Mechanics–Molecular Mechanics approach (QM/MM) . . . 32

2.5.1 Force Fields . . . 32
2.5.2 Embedding Scheme . . . 33
2.5.3 Cutting Covalent Bonds . . . 34
2.5.4 Software . . . 36
2.6 Molecular Dynamics . . . 36
2.7 Fluorescent Protein Setup . . . 37

3 Photoswitchable Fluorescent Proteins: OFF-state 41
3.1 Introduction . . . 41
3.2 Methods . . . 44
Setup. . . . 44
QM region. . . . 45
Pathway calculations. . . . 46
3.3 Results and Discussion . . . 47
3.3.1 Isomerization of the neutral chromophore . . . 47
3.3.2 Concerted proton transfer and isomerization. . . . 52
3.3.3 The Role of Glu212 and hydrogen-bonds to the chromophore 59
Comparison of the relaxed 2D scan. . . . 60
Single dihedral constrained pathways . . . 61
Correlation between isomerization and proton transfer. 61
3.4 Conclusions . . . 65

4 Photoswitchable Fluorescent Proteins: ON-state 69
4.1 Introduction . . . 69
4.2 Methods . . . 71
Setup. . . . 71
QM Region. . . . 71
Pathway calculations. . . . 72
4.3 Methodological Test . . . 72
IrisFP ON-state: Isomerization of the anionic CRO. . . 72
4.4 Results and Discussion . . . 73
4.5 Conclusions . . . 81

Conclusions 83

A Photoswitchable Fluorescent Proteins: OFF-state 85
A.1 Protonation State Selection . . . 85
A.2 Results with IrisFP model B . . . 87
A.3 QM method for PT (test in IrisFP) . . . 89
A.4 OM2/MRCI τ-constrained isomerization pathway in IrisFP . . . 90
A.5 Active space (AS) selection . . . 90
A.6 Results with Dronpa model Y1 . . . 91

Bibliography **95**

List of Figures

1.1 (a)Negative photoswitching behaviour (Example: Dronpa). (b)Positive photoswitching behaviour (Example: Padron).
Figure from Zhou et al. [1] 5

1.2 (a)ON-state chromophore pocket and hydrogen bonding network around it (IrisFP example and numbering). (b)OFF-state chromophore pocket and hydrogen bonding network around it (IrisFP). (c)View of the whole protein, cartoon style + small balls and sticks + the chromophore conjugated system in thick green sticks. Figure adapted from Duan et al. [2] 7

2.1 Excitation by absorption of a photon of the right energy $h\nu$ to the Franck-Condon region, and posterior de-excitation processes (black dotted lines) between the first excited state S_1 (red) and the ground state S_0 (blue). On the left, radiationless decay through the CI. On the right, fluorescence emission of a photon of lower energy $h\nu'$. . . 14

2.2 QM/MM concept 35

3.1 Chromophore in the ON- and OFF-states. 41

3.2 IrisFP in fluorescent (green) and dark (black) states; Dronpa in fluorescent (lime) and dark (gray) states. Delimited by a red dotted line a region with accumulation of structural differences between the two proteins, coming from the different conformation of Met159 and the substitutions Ser/Phe and Ile/Val of residues 173 and 157 respectively for IrisFP/Dronpa. 42

3.3 QM regions: a) QM1 including the whole CRO and two backbone atoms from each of the two neighboring residues. b) QM2 including just the conjugated CRO atoms c) QM3, includes Glu144 side chain d) QM4, includes H-bond mediating water267 for Dronpa e) conjugated CRO atoms (green), Glu144 side chain (tan), QM2 AR (blue), extra atoms to the AR for QM3 (cyan). In red the rest of the atoms of the residues having just some of them inside the AR (example for IrisFP). 45

3.4 In black the atoms involved in the dihedrals constrained to do the τ-ϕ 2D scanning. ϕ and τ dihedrals in yellow and gray respectively. 46

3.5 IrisFP CRO-Glu144 distance in IrisFP (H-bond in yellow, black structure) and in Dronpa (H-bond in blue, gray structure). In red IrisFP's Glu211-CRO H-bond. 48

3.6 **IrisFP** relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The white dot is the FC point. 48

3.7 **IrisFP** ES-GS energy difference of the relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The lower the value the nearer to a CI seam. 48

3.8 **Dronpa** relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The white dot is the FC point. 49

3.9 **Dronpa** ES-GS energy difference of the relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The lower the value the nearer to a CI seam. 49

3.10 **IrisFP** relaxed (ES-optimized) energy profile of the τ-constrained (a) and ϕ-constrained (b) isomerization pathway for the neutral CRO. Zero energy is set to the FC point. 50

3.11 **Dronpa** relaxed (ES-optimized) energy profile of the τ-constrained (a) and ϕ-constrained (b) isomerization pathway for the neutral CRO. Zero energy is set to the FC point. 50

3.12 **IrisFP** structure at the CI point for negative τ isomerization (purple) compared with the FC structure (black). 51

3.13 **Dronpa** structure at the CI point for positive τ isomerization (purple) compared with the FC structure (gray). 51

3.14 **IrisFP** ES energy profile along the λ-constrained PT pathway. CC2 is used as QM method. Zero energy is set to the FC point. 53

3.15 **Dronpa** ES energy profile along the λ'-constrained PT pathway, except for 144-D, where the constrain is λ. CC2 is used as QM method. Zero energy is set to the FC point. 53

3.16 **IrisFP** τ and ϕ dihedrals along the λ-constrained PT pathway (CC2 is used as the QM method). The upper and lower horizontal lines mark the τ and ϕ values at the CI, respectively. 54

3.17 **Dronpa** τ and ϕ dihedrals along the λ'-constrained PT pathway (CC2 is used as the QM method). 54

3.18 **IrisFP** (a) Energy profile along the τ-constrained isomerization pathway with QM2 (light blue), with QM3 and optimizing from neutral CRO (dark blue) and with QM3 and frozen λ to anionic CRO (magenta) using CC2 as the QM method. The OM2/MRCI profile for QM2 and neutral CRO is shown for comparison (gray). (b) The corresponding λ values. Zero energy is set to the FC point. 56

3.19 **Dronpa** (a) Energy profile along the τ-constrained isomerization pathway with QM2 (light blue), with QM4 and optimizing from neutral CRO (dark blue) and with QM4 and from anionic CRO (magenta) using CC2 as the QM method. The OM2/MRCI profile for QM2 and neutral CRO is shown for comparison (gray). (b) The corresponding λ' values. Zero energy is set to the FC point. 56

3.20 **Dronpa** (model Z). In purple the hydrogens and H-bonds involved in the PT. It is a water mediated H-bond. Glu211 is deprotonated. 59

3.21 Dronpa model **Y1**. In purple the hydrogens and H-bonds involved in the PT. It is a water mediated H-bond. Glu211 is protonated (orange). 59

3.22 Dronpa model **Y2**. In purple the hydrogens and H-bonds involved in the PT. It is a direct H-bond. In orange the Glu211-CRO H-bond. 60

3.23 **IrisFP**. In purple the hydrogens and H-bonds involved in the PT. It is a direct H-bond. In orange the Glu211-CRO H-bond. 60

3.24 IrisFP, Dronpa Y1 and Dronpa (Z) relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The white dots are the FC points. 60

3.25 **IrisFP**. In red *tau*-constrained isomerization pathway's unfavoured CI point. FC structure in black for comparison. 61

3.26 **IrisFP**. In purple *tau*-constrained isomerization pathway's preferred CI point. FC structure in black for comparison. 61

3.27 IrisFP and Dronpa models Z and Y2 relaxed (ES-optimized) energy profiles of the τ-constrained (a) and ϕ-constrained (b) isomerization pathway for the neutral CRO. Zero energy is set to the FC point. . 62

3.28 τ and ϕ dihedrals along the λ/λ'-constrained PT pathway for IrisFP, Dronpa (Z) and Drompa model Y2 (CC2 is used as the QM method). The upper and lower horizontal red lines mark the τ and ϕ values at the lower energy CI, respectively (For IrisFP). 63

3.29 Energy profile along the λ/λ'-constrained PT pathway for IrisFP, Dronpa (Z) and Drompa model Y2 (CC2 is used as the QM method). 63

3.30 (a) Energy profile along the τ-constrained isomerization pathway with QM2 (light blue), with QM3 and neutral CRO (dark blue) and with QM3 or QM4 and anionic CRO (magenta or green respectively) using CC2 as the QM method. The OM2/MRCI profile for QM2 and neutral CRO is shown for comparison (gray). (b) The corresponding λ values. In Dronpa model Y2, λ values of -2 are reached because the Glu144-CRO H-bond changes from direct to water mediated. Zero energy is set to the FC point. 64

4.1 **Dronpa2**. In black the H-bonds to the CRO phenol. Residue 159 is a Threonine. 70

4.2 **Dronpa**. In black the H-bonds to the CRO phenol. Residue 159 is a Metionine, in a different position than IrisFP. 70

4.3 **IrisFP**. In black the H-bonds to the CRO phenol. Residue 159 is a Metionine, in a different position than Dronpa. 70

4.4 ON-state: Relaxed (ES-optimized) energy profile along the τ-constrained (a) and ϕ-constrained (b) isomerization pathway. The dashed line denotes the FC point. 73

4.5 **IrisFP** relaxed 2D (τ, ϕ) scan of the ES potential energy surface (2D projection). 74
4.6 **IrisFP** ES-GS energy difference of the relaxed 2D (τ, ϕ) scan of the ES potential energy surface (2D projection). 74
4.7 **Dronpa2** relaxed 2D (τ, ϕ) scan of the ES potential energy surface (2D projection). White line: manually optimized minimum energy pathway to the CI. 74
4.8 **Dronpa2** ES-GS energy difference of the relaxed 2D (τ, ϕ) scan of the ES potential energy surface (2D projection). 74
4.9 **Dronpa** relaxed 2D (τ, ϕ) scan of the ES potential energy surface (2D projection). 75
4.10 **Dronpa** ES-GS energy difference of the relaxed 2D (τ, ϕ) scan of the ES potential energy surface (2D projection). 75
4.11 Energy profile along the hula-twist coordinate. 76
4.12 Close up of Figure 4.11. 76
4.13 Dronpa (blue), IrisFP (magenta), and Dronpa2 (green). **a)** hula-twist from $\tau = 70°$ (translucent) to 85° (opaque, the CI point). No change apart from the dihedrals. **b)** hula-twist from $\tau = 85°$ (translucent) to 95° (opaque). The CRO phenolate rearranges in Dronpa2 (Water1 follows its motion) and loses the H-bond to Water2. **c)** hula-twist from $\tau = 95°$ (translucent) to 105° (opaque). In IrisFP the CRO starts to push against Met159 and loses the H-bond to Water2. **d)** hula-twist from $\tau = 105°$ (translucent) to 150° (opaque). In IrisFP the CRO pushes Met159 further. Water 2 is W278/W17/W320 in the PDB file of Dronpa/IrisFP/Dronpa2. Water 1 is W188/W319 in the PDB file of IrisFP/Dronpa2. 78
4.14 **Dronpa2** energy profiles along the hula-twist coordinate for each possible combination of waters 1 and 2 deleted. 79
4.15 **Dronpa** energy profiles along the hula-twist coordinate. It has only water 2 (see Figure 4.13a). Curves with and without it. An other water H-bond to water 2 (not shown) was also deleted not the replace water 2 effect form a bit further. 79
4.16 **IrisFP** energy profiles along the hula-twist coordinate for each possible combination of waters 1 and 2 deleted. 79

4.17 Hula-twist of normal Dronpa (blue), and IrisFP (magenta) and Dronpa2 (green) with waters 1-2 deleted. 81

A.1 OFF-state: Relaxed (ES-optimized) 2D (τ, ϕ) scan of the ground- and ES potential energy surfaces for the neutral CRO. 87
A.2 OFF-state: Relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). 87
A.3 OFF-state: Relaxed (ES-optimized) energy profile of the τ-constrained (a) and ϕ-constrained (b) isomerization pathway for the neutral CRO. 87
A.4 OFF-state energy profile along τ (red), and ϕ (purple), constraint isomerization pathway for the Anionic CRO. 88
A.5 OFF-state: Energy profile along the λ-constrained PT pathway. CC2 is used as QM method. 88
A.6 OFF-state: τ and ϕ dihedrals along the λ-constrained PT pathway (CC2 is used as the QM method). The upper and lower horizontal lines mark the τ and ϕ values at the CI, respectively. 88
A.7 OFF-state: OM2/MRCI and MP2 relaxed GS energy profiles along the λ-constrained PT pathway. The ϕ and τ dihedrals were fixed to the same GS values during optimization. 89
A.8 OFF-state: OM2/MRCI and CC2 energy profiles along the ES-relaxed λ-constrained PT pathway. The ϕ and τ dihedrals were fixed to the same GS values during optimization. 89
A.9 OFF-state: (a) Energy profile along the τ-constrained isomerization pathway with QM3 and OM2/MRCI as the QM method. Optimization with free λ (orange) or with λ constrained to describe a neutral (black) or anionic (blue) CRO. (b) The corresponding λ values. . . 90
A.10 **Dronpa model Y1** relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The white dot is the FC point.

. . . . 91
A.11 **Dronpa model Y1** ES-GS energy difference of the relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The lower the value the nearer to a CI seam. . . . 91
A.12 **Dronpa model Y1** relaxed (ES-optimized) energy profile of the τ-constrained (a) and ϕ-constrained (b) isomerization pathway for the neutral CRO. Zero energy is set to the FC point. 92
A.13 **Dronpa model Y1** ES energy profile along the λ'-constrained PT pathway, except for "mean3 Direct", where the constrain is λ. CC2 is used as QM method. Zero energy is set to the FC point. 92
A.14 **Dronpa model Y1** τ and ϕ dihedrals along the λ'-constrained PT pathway (CC2 is used as the QM method).
. . . . 92

A.15 **Dronpa model Y1** (a) Energy profile along the τ-constrained isomerization pathway with QM2 (light blue), with QM4 and optimizing from neutral CRO (dark blue) and with QM4 and from anionic CRO (magenta) using CC2 as the QM method. The OM2/MRCI profile for QM2 and neutral CRO is shown for comparison (gray). (b) The corresponding λ' values. Zero energy is set to the FC point. 93

List of Tables

4.1 CI point Energies from the minimum (eV) 77
4.2 Fluorescence quantum yields 77

A.1 **ON-state** protonation models 86
A.2 **OFF-state** protonation models 86
A.3 AS(m,n) 91

Abbreviations

AS	**A**ctive **S**pace
BO	**B**orn–**O**ppenheimer approximation
CC	**C**oupled **C**luster
CC2	second order approximate **C**oupled **C**luster
CCSD	**C**oupled **C**luster **S**ingles **D**oubles
CI	**C**onical **I**ntersection *or* **C**onfiguration **I**nteraction
CISD	**C**onfiguration **I**nteraction **S**ingles **D**oubles
CRO	Ch**RO**mophore
DFT	**D**ensity **F**unctional **T**heory
DFTB	**D**ensity **F**unctional based **T**ight **B**inding
ES	**E**xcited **S**tate
ESPT	**E**xcited **S**tate **P**roton **T**ransfer
FC	**F**ranck–**C**ondon
FF	**F**orce **F**ield
FP	**F**luorescent **P**rotein
FTIR	**F**ourier **T**ransform **I**nfrared spectroscopy
GFP	**G**reen **F**luorescent **P**rotein
GS	**G**round **S**tate
H-bond	**H**ydrogen bond
HF	**H**artree–**F**ock
HOMO	**H**ighest **O**ccupied **M**olecular **O**rbital
KS	**K**ohn–**S**ham
LUMO	**L**owest **U**noccupied **M**olecular **O**rbital
MD	**M**olecular **D**ynamics

MM	**M**olecular **M**echanics
MP	**M**oller–**P**lesset perturbation theory
MP2	second order **M**oller–**P**lesset perturbation theory
MRCI	**M**ulti **R**eference **C**onfiguration **I**nteraction
OM2	**O**rthogonalization **M**ethod 2
PT	**P**roton **T**ransfer
QM	**Q**uantum **M**echanics
QM/MM	**Q**uantum **M**echanics approach
QY	**Q**uantum **Y**ield
RHF	**R**estricted **H**artree–**F**ock
ROHF	**R**estricted **O**pen-shell **H**artree–**F**ock
RSFP	**R**reversibly **S**witchable **F**luorescent **P**rotein
SCF	**S**elf **C**onsistent **F**ield

Chapter 1

Introduction

1.1 Fluorescent Proteins

"Proteins are large biomolecules, or macromolecules, consisting of one or more long chains of amino acid residues. Proteins perform a vast array of functions within organisms, including catalysing metabolic reactions, DNA replication, responding to stimuli, and transporting molecules from one location to another. Proteins differ from one another primarily in their sequence of amino acids, which is dictated by the nucleotide sequence of their genes, and which usually results in protein folding into a specific three-dimensional structure that determines its activity." (Wikipedia)

The importance of proteins for the cell functioning is undeniable, they constitute the main machinery and are an important part of the building blocks together with other ones like the lipids that form the membranes or DNA and RNA for information storage and transmission. Many diseases, like for instance HIV, cancer, heart disease or malaria, can be studied and better understood by studying the effect of related proteins in each case. Monitoring them provides clues to their function and regulation, and thus the ability to tag and track them is a priceless tool for medicine and biology.

The research and development of new imaging probes and advanced imaging technologies that make it possible to track and measure the properties of biological molecules triggered a flourishing epoch of biological research. Watching molecular behaviour in real time, at a nanometre scale, in living cells and organelles, opened the possibility to unravel the underlying complexities and regulatory mechanisms of biological systems. There are already many standard ways to label a protein inside a cell for fluorescent imaging, but the most common and conventional method is to use intrinsically fluorescent proteins (FPs) related in structure or sequence to green fluorescent protein (GFP) [3]. The FP's DNA sequence can be genetically fused to the sequence of the target protein. In such a mutant, the target protein is expressed along with the FP, and thus covalently bonded to it.

The engineering of fluorescent proteins emerged in the late 1990's, and its story began with GFP, the green fluorescent protein from Aequorea victoria. Osamu Shimomura, while studying this marine Hydrozoa species of jelly fish (he was studying it for 40 years), purified GFP from it. But it wasn't until 1992, when Douglas Prasher cloned and sequenced its gene [4] that it captured biologists' attention. Soon after GFP was expressed within E. coli and C. elegans cells [5] and the green fluorescence was successfully observed by UV light illumination. From then on, FP engineering started, with the group of Roger Y. Tsien being one of the firsts in exploring ways to improve GFP by mutagenesis. In 1995 this group developed the first truly usable GFP mutant (GFP-S65T [6]) that was improving the main problems of the wild type of GFP by changing the excitation peak from UV to visible and increasing the fluorescence brightness and photostability [7]. The field started to develop, and FPs quickly became indispensable imaging tools across large areas of biological research, and were used as a gene expression marker and virus infection reporter among other applications [5, 8, 9].

Since then, fluorescent proteins from other organisms where found and engineered, resulting in new colors and properties. Tsien and collaborators and some other groups continued with the GFP mutagenesis experiments aiming at obtaining proteins with better brightness and different colors. This way GFP was modified via directed evolution to fluoresce brighter and at a broad spectrum of wavelengths

[10–12]. The work on Aequorea GFP resulted in a variety of fluorescent proteins in the visible region from blue to yellow. But attempts to obtain a stable and bright red version of GFP where unsuccessful. This apparent dead end fueled the search for fluorescent proteins in other organisms, in case red variants already existed in nature. At the same time, the first GFP mutants and wild type GFP crystal structures where published (GFP-S65T [13] and wild type GFP [14]). The discovery of new naturally occurring fluorescent proteins was the first major expansion of FP diversity. FPs with altered chromophore environments where found. They absorbed and emitted light over a broad spectrum of colors, ranging from cyan to red [15–17]. The resolved structures allowed researchers to identify the chromophore that is responsible for fluorescence as well as its environment and interactions with neighbouring residues. The variety of colors and environments and the resolution of crystal structures encouraged molecular biologists and protein engineers to try a more rational approach than directed evolution and modify FP spectra by mutations to the chromophore and residues in the immediate chromophore environment. It was a successful approach and further expanded the color palette to include deep blue and far-red FPs [18–23], opening the way to rational design. Even today, we are far from understanding all residue's functionalities in the FPs, and rational engineering of FPs is usually very difficult and brings unexpected consequences. Even when the desired effect has been achieved, several rounds of random mutagenesis are usually needed after the site directed mutations to fix undesired side effects. During the discovery of this diversity of FPs with different spectra, a broad range of attributes were discovered and engineered for practical applications including halide detection, pH measurement and phototransformable fluorescence [24–29]. FPs with the latter property, phototransformable FPs (PTFPs), more precisely, reversibly photoswitchable FPs (RSFPs), are the object of study of the present thesis. RSFPs can be reversibly photo-transformed between two different fluorescent states, a bright ON-state and a dark OFF-state. Two other types of PTFPs exist that will not be investigated here, photoactivatable FPs (PAFPs), which are naturally dark and can irreversibly become fluorescent when exposed to light, and photoconvertible FPs (PCFPs), which can be non-reversibly transformed between two fluorescent states with different emission colors.

Photoactivatable fluorescent proteins, in particular RSFPs, extended the possibilities of fluorescence microscopy by some of its most successful techniques. Single molecule localization microscopy, with techniques such as stochastic optical reconstruction microscopy (STORM [30]), are based on the stochastic switching of RSFPs under light; "false-multicolor" photo activated localization microscopy (PALM [31]), distinguishes RSFP variants of similar colors by properties such as different photoswitching rates. Super-resolution optical fluctuation imaging (SOFI), has been improved thanks to the possibility of precisely controlling the switching of RSFPs, developing photochromic SOFI (pcSOFI [32]). They can also be advantageously used in nonlinear imaging with techniques such as reversible saturable optical fluorescence transitions (RESOLFT [33]), which works as stimulated emission depletion microscopy (STED), but replacing the laser driven depletion of standard fluorophores by RSFP's photoswitching, this way reducing the power density required; similarly, saturated structured illumination microscopy (SSIM [34]) can use very weak illumination intensities when combined with RSFPs.

Research on FPs was awarded with two Novel Prices:

O. Shimomura, M. Chalfie, R. Tsien have been awarded the Nobel Prize in Chemistry 2008 for the discovery, use and development of the GFP.

Eric Betzig, Stefan W. Hell, William E. Moerner have been awarded the Nobel Prize in Chemistry 2014 for the development of super-resolved fluorescence microscopy, such a successful technique thanks to taking advantage from the properties of RSFPs.

1.2 Reversibly Photoswitchable Fluorescent Proteins

The first observation of reversible photoswitching of a fluorescent protein was made with yellow derivatives of Aequora victoria GFP, at the single molecule level [35] and at ensemble level and room temperature [36–38], but efficient photoswitching was initially developed on FPs with Anthozoan origin. The first reported case on Anthozoans was the weakly fluorescent and tetrameric asFP595, from the sea anemone Anemonia sulcata. It was weakly fluorescent in its natural state and could be switched on (brightened) by the same light that excites fluorescence (568 nm) and quenched by illumination with blue light (450 nm) [39]. This behaviour is called positive photoswitching (Figure 1.1b). After that, the first RSFP with suitable characteristics for practical biological experiments was Dronpa, engineered form a Pectiniidae coral FP [40], now widely used. The photoswitching of Dronpa is in the opposite direction than that of asFP595. It is naturally in its bright state and can be switched from ON to OFF by the same light that excites fluorescence (490 nm) and recovered by violet light (405 nm). This behaviour is called negative photoswitching(Figure 1.1a). As these two proteins were the first ones of each class of photoswitching, a large number of mechanistic investigations were performed on them [41–57].

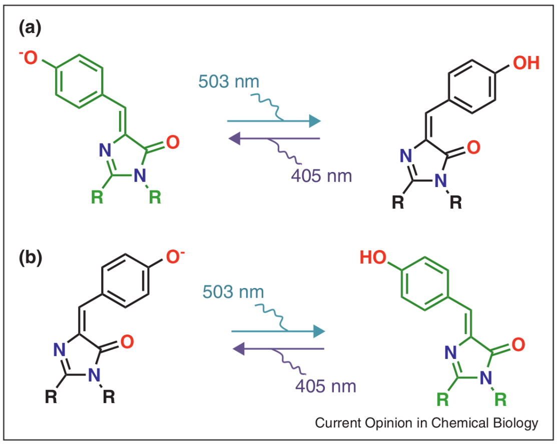

Figure 1.1: (a)Negative photoswitching behaviour (Example: Dronpa). (b)Positive photoswitching behaviour (Example: Padron). Figure from Zhou et al. [1]

The greatest success of RSFPs is that they can be advantageously used in fluorescence microscopy to increase the resolution beyond the diffraction limit by using the technique of single molecule localization microscopy or nonlinear imaging [30–34]. This is called super-resolution microscopy and has been a big motivation to develop variants with improved properties such as increased photoresistance, red-shifted emission, higher fluorescence brightness, enhanced switching contrast or tunable switching quantum yields.

Surprisingly little structural information is available on RSFPs evolved from GFP derivatives. In this thesis I will focus on the variety of RSFPs from Anthozoan origin, for which rich structural information is available, from crystal structures of both ON- and OFF-states to spectroscopical characterization.

FPs are composed of a $\sim$ 240 amino acids long chain ($\sim$ 27 kD), which folds as a 11-stranded β-barrel, enclosing a 4-(p-hydroxybenzylidene)-5-imidazolinone (p-HBI) chromophore (Figure 1.2c). One of the advantages of FPs is that the chromophore is autocatalytically formed, and only needs oxygen as an external cofactor to mature. Therefore, it is not necessary to inoculate the cell with any external cofactor. The chromophore is formed by the three central amino acids of an α-helix that goes through a β-barrel that has a length of 42 Å and a diameter of 24 Å. This places the chromophore in the middle of the barrel and and shields it from the solvent, making the FPs relatively small, compact and chemically inert proteins. The β-barrel provides the chromophore with a rich and chemically complex environment responsible for the diverse characteristics of different FPs, including the ability to be reversibly photoswitched.

1.3 Most Recent Developments

When I started this thesis the knowledge about the precise mechanism for the photoswitching was far less clear than it is today. It was only known that it was involving an structural and protonation state change.

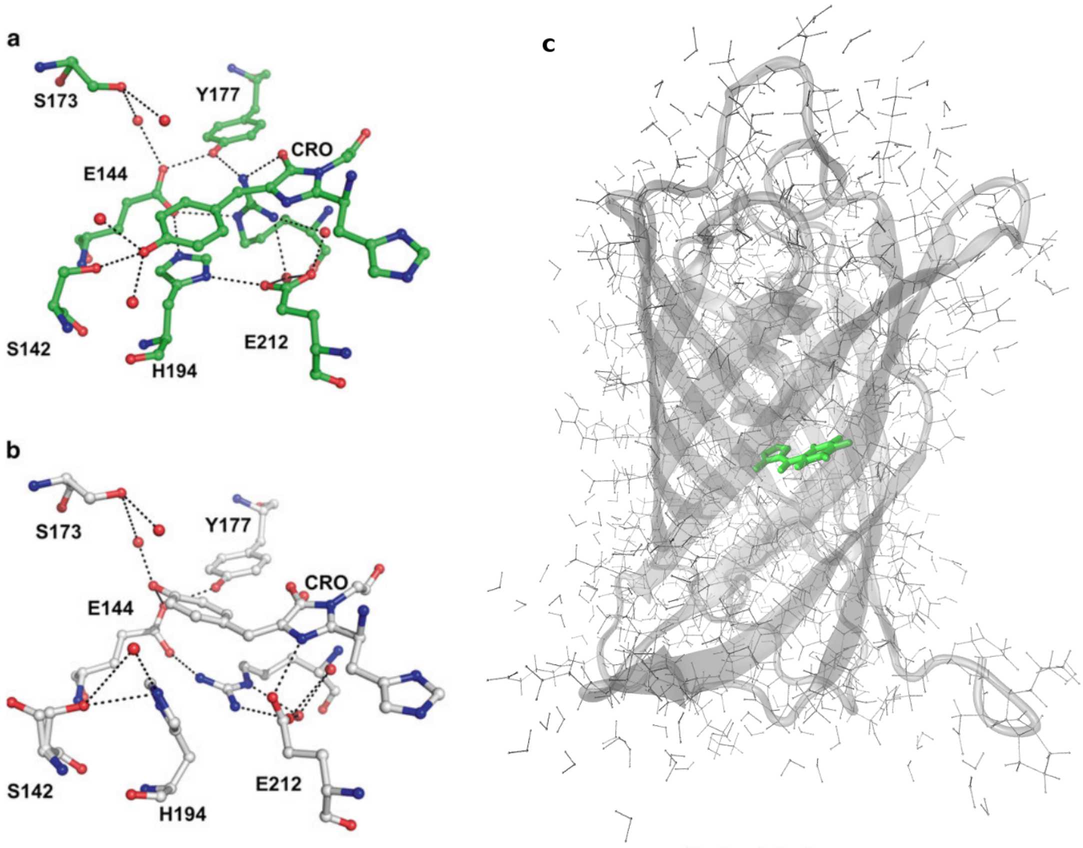

FIGURE 1.2: (a)ON-state chromophore pocket and hydrogen bonding network around it (IrisFP example and numbering). (b)OFF-state chromophore pocket and hydrogen bonding network around it (IrisFP). (c)View of the whole protein, cartoon style + small balls and sticks + the chromophore conjugated system in thick green sticks. Figure adapted from Duan et al. [2]

Crystalographic structures of the ON- and OFF-states of the negative photoswitching RSFPs Dronpa [47, 48], mTFP0.7 [58] and IrisFP [59]) show a cis chromophore in the ON-state and a trans configuration in the OFF-state. A substantial conformational change of the chromophore pocket accompanies the isomerization, and both the ON- and OFF-states are stabilized by a switch in the hydrogen bonding network around the chromophore (Figure 1.2 a and b). The tightly H-bonded triad Glu144-His193-Glu211 in the cis configuration is replaced by the Glu144-Arg66-Glu211 triad in the trans configuration, with either His193 or Arg66 stabilizing the chromophore by $\pi-\pi$ stacking and π-cation interactions with the hydroxybenzylidene moiety, respectively (the residues are named according to the numbering in Dronpa). Ser142 plays also a key role, maintaining a strong H-bond with the hydroxybenzylidene moiety in the cis state and finding another

H-bonding partner once the chromophore has isomerized [59]. The cis ON-state is anionic, while the trans OFF-state is neutral, and the difference in protonation state was explained by the substantial change in the electrostatic environment [43].

The reversible light-induced cis-trans isomerization of the chromophore and/or chromophore protonation were hypothesized to be most likely reasons for the observed switching, but it was not clear whether one of the two processes triggers the other or whether they happen in a concerted manner. The hypothesis was coming from mechanistic investigations of some of the RSFPs and was generally based on a combination of crystallographic, spectroscopic, and molecular dynamics investigations [42–44, 58–61].

This thesis is focused on the study of these two processes (the isomerization and the protonation state change), the relation between the two, and the differences between three different RSFPs: IrisFP, Dronpa and Dronpa2, regarding to the chromophore pocket (Dronpa2 will not be studied in the OFF-state as there is no crystal structure available for it). I performed ground state calculations with the well established quantum mechanical/molecular mechanical methodology (QM/MM) [62, 63], which allows to treat the fluorescent chromophore at the level of quantum mechanics (QM) and the rest of the protein and solvent by using molecular mechanics description (MM). The reliable and efficient approximation to density functional theory (DFT) called DFTB will be used as the QM method for ground state optimizations and molecular dynamics. The fast semiempirical OM2/MRCI method and the CC2 approximation to the ab-initio coupled cluster method will be used for the excited state optimizations. The goal is a better understanding of the photoswitching process and answering the following questions related to it.

Some of the open questions at the time I started the thesis in 2012 were stated in the RSFP review by Dominique Bourgeois and Virgile Adam [64]:

- What makes the trans state of the chromophore dark and the cis state bright? In all known RSFPs photoswitching involves cis-trans isomerization and the dark state has the trans chromophopre, but the trans state of the chromophore can be fluorescent as well, as it is in other FPs [50, 65]. The relation between the fluorescence and the isomeric state was not obvious and chromophore planarity, chromophore rigidity, chromophore protonation state and possibly excited state proton transfer (ESPT) where suggested as possible explanations.

- Which mechanism triggers the OFF to ON, and the opposite switching process? Single-molecule [41] and fluorescence correlation spectroscopy [45] demonstrated that switching occurs on rapid time-scales (ms to μs) and found that the light-induced OFF-state corresponds to a protonated state of the chromophore that differs from the pH-induced protonated state (the ON-state was known to be anionic, and crystal structures further confirmed this). For the ON to OFF switching, the hypothesis that a photoinduced protonation is gating the process was ruled out in view of the fact that the primary response of a free chromophore to photoabsorption is to rotate around its methylene bridge bonds together with the fast photoswitching time-scales found. Instead, there where two alternative hypotheses: The isomerization of the chromophore followed by protonation [66], or a concerted isomerization/protonation mechanism, where the coupling would arise from the fact that drastic changes in chromophore pKa occur during the photoisomerization [67], in line with ab-initio calculations [68]. For the OFF to ON switching, ESPT to an intermediate state "I" was proposed to account for the high-yield [41], which was thought to be confirmed by ultrafast absorption spectroscopy and kinetic isotope effect measurements using deuterated samples [46]. For Dronpa also a different scenario was proposed. By employing NMR, Mizuno et al. [44] proposed, in contrast to the crystallographic results described above, that on-off switching in Dronpa results from a disorder of the chromophore geometry, but not necessarily from cis-trans isomerization. The study pointed out that the importance of the protein oligomerization state (found in the crystal) in possibly modifying the structural flexibility necessary for switching. Chromophore disordering would be initiated by light induced protonation in the cis triplet state, after intersystem crossing. The hypothesis was consistent with spectroscopic studies carried out at cryo-temperature which, by blocking the conformational rearrangements necessary for isomerization, strongly suggested that photo-induced protonation without isomerization are feasible processes [69].

- The precise isomerization pattern (single bond flip, or the more space conserving hula-twist) was also unknown, with theoretical works on asFP595 [54, 55] and HcRed [70] (a FP that can be photoisomerized but without enough contrast to be considered a RSFP) giving contradictory results.

In February 2013, Warren et al. published a time-resolved infrared spectroscopy and Fourier transform infrared (FTIR) study on the OFF-state of Dronpa. Pump-probe measurements directly resolve the spectral features of the primary photoproduct, and the comparison of the long-lived component with the static FTIR difference spectra excluded the possibility of deprotonation. In particular, the upshifted C=C stretching and phenol frequencies seen in the static FTIR difference spectra were strikingly absent from the long-lived photoproduct spectrum, indicating a cis protonated ground state intermediate, and strongly suggesting that deprotonation of the chromophore occurs in the electronic ground state as a step subsequent to isomerization [51]. This conclusion was further supported by Yadav et al. in October 2014, through the study of the same intermediate [57], this time comparing the UV-visible transient absorption spectra with the steady-state one. The latter work was studying Dronpa2 in parallel with identical results. Dronpa2 is a fast photoswitching single mutant of Dronpa discovered in 2007 [48], whose ON-state crystal structure was obtained in December 2014 [71]. The greater photoswitching quantum yield of Dronpa2 makes it a good candidate to study the process, not only in the OFF but also in the ON-state. In September 2013, Lukacs et al. made the first time resolved spectroscopy study on both states. They performed time-resolved infrared (TRIR) spectroscopy and clearly discarded excited state proton transfer [72]. In november 2014, Kaucikas et al. also studied Dronpa2 in the ON and OFF-states, this time comparing the time-resolved infrared spectra and the static FTIR difference spectra. For the OFF-state their results supported the same process as the results of Warren et al. and Yadav et al. for Dronpa [51, 57]. For the ON-state they conclude a fast excited state isomerization followed by ground state protonation in the nano- to millisecond time scale. In November 2015 a theoretical study on the ON to OFF switching of Dronpa2 was published by Morozov et al. [73]. They proposed that a sub population of fluctuations in GS with certain hydrogen bonds to the chromophore phenolate broken, was the responsible of the photoisomerization. Finally, in February 2016 Colletier et al. made a time resolved spectroscopy study on IrisFP's OFF-state by ultrafast transient absorption techniques with identical results as the ones obtained for Dronpa and Dronpa2 [51, 57, 74].

Altogether, the gathered experimental evidence shows that in both, ON to OFF and OFF to ON photoswitching processes excited state photosiomerization is followed by ground state protonation or deprotonation of the chromophore respectively. But there is still little theoretical work explaining why it is so and how it works in microscopical detail. This thesis is a theoretical work to achieve a deeper understanding of the process beyond the issue about the order of the two processes in orther to support the rational design of RSFPswith the characteristics required for practical applications, like photoisomerization rates and quantum yields.

Chapter 2

Theory & Methods

2.1 Photochemical Processes

In the study of fluorescent proteins, we are directly involved with excited-state properties. Without describing how do we reach the excited-state, i.e., how the protein absorbs the photons, the scope of this thesis is on how does the system evolve afterwards. When the energy difference between two electronic states coincides with the energy of an incident photon, the photon can be absorbed promoting an electron from the lower to the higher energy level. Due to the higher energy and the interaction with the environment, the excited-state is just metastable in the sense of not being a true eigenstate, and it will relax back to the ground state. The relaxation process can happen through three different mechanisms. The fastest one is the radiationless decay at a conical intersection, on time scales of pico- or even femtoseconds. When this is not easily accessible the radiative relaxation takes place, by fluorescence within nanoseconds or, even slower, by phosphorescence, which can happen from milliseconds to hours. I am not going to discuss phosphorescence, which unlike fluorescence, involves transitions between singlet and triplet states due to spin–orbit coupling, as it is not relevant for the systems under investigation in this thesis. You can see radiationless decay and fluorescence depicted in Fig. 2.1.

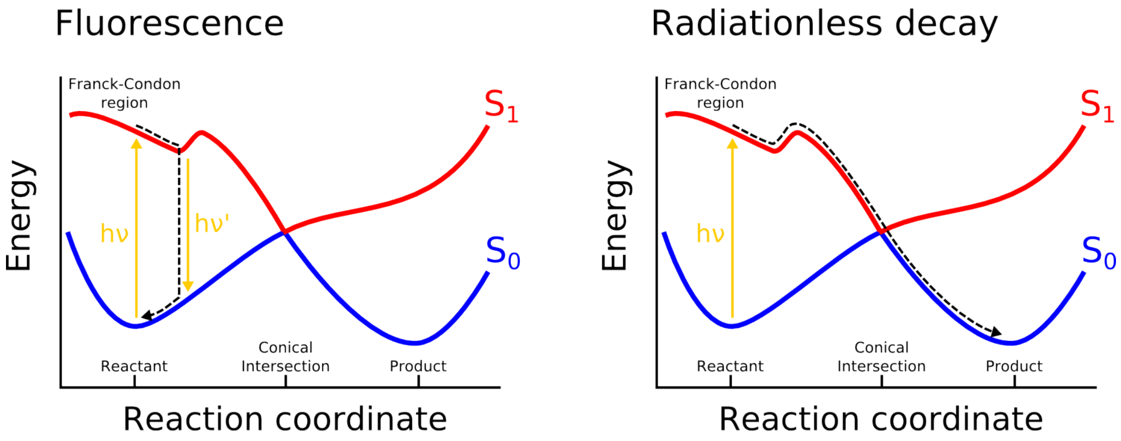

FIGURE 2.1: Excitation by absorption of a photon of the right energy $h\nu$ to the Franck-Condon region, and posterior de-excitation processes (black dotted lines) between the first excited state S_1 (red) and the ground state S_0 (blue). On the left, radiationless decay through the CI. On the right, fluorescence emission of a photon of lower energy $h\nu'$.

2.1.1 Radiationless Decay at Conical Intersection

A conical intersection (CI) is a region in the nuclear configuration space where two electronic states cross, i.e., they are degenerate. In these intersection regions, molecules can effectively and quickly cross from one electronic state to the other and therefore are able to convert the electronic excitation energy to vibrational energy completely without any emission of radiation. The topology of the CI and its location along a reaction coordinate determines whether either the initial ground state configuration is restored, or a chemically different species is formed (Fig. 2.1). If this kind of de-excitation through a CI would be the main process, then the fluorescent protein (FP) would be dark, and can be at least partially responsible of photoshwitchable FPs' OFF-state.

All the quantum mechanical calculations we do thorough this thesis were done within the Born–Oppenheimer approximation (BO), which decouples the electron and nuclear equations based on the much faster movement of electrons [75]. It is important to note the the BO approximation is broken near the CI, as the energy gaps between different vibrational levels of the nuclear wave function are comparable to the gap between the electronic states near the CI. Therefore, nuclear and electronic wave functions might resonantly interfere, and slight changes in the nuclear positions can induce transitions between the different electronic surfaces. Some methods using the BO approximation can get the right shape of the CI [76], but the so called non-adiabatic coupling means that the change from one electronic state to the other can happen before their potential energy surfaces actually cross, as soon as they start to get closer. For example, a hopping algorithm can be based on the one-dimensional Landau–Zener equation, which relates the probability of a transition between two electronic states S_1 and S_2 to the non-adiabatic coupling, via

$$P_{1\to 2} = exp(-\frac{1}{4}\pi\xi). \tag{2.1}$$

In this equation ξ is the Massey parameter, defined as [77]

$$\xi = \frac{\Delta E}{\hbar\frac{\partial Q}{\partial t} \cdot g(Q)}, \tag{2.2}$$

where ΔE is the energy gap between the adiabatic states, Q represents a one dimensional nuclear reaction coordinate, and $g(Q)$ is the non-adiabatic coupling.

2.1.2 Fluorescence

When the molecule is excited by the photon, it does a vertical transition from the ground state minimum to a region of the excited-state potential energy surface (the so called Franck–Condon region, Fig. 2.1). It afterwards relaxes to a nearby local minimum. If the energy barriers towards the surrounding CIs are to high to be crossed within nanoseconds there is time for an spontaneous emission of a photon, it fluoresces returning to the ground state (Fig. 2.1 left). This is the main process when I refer to a fluorescent or bright state of the FPs. Brightness is actually dependent on other factors like absorptivity, but they will not be discussed here. The fast relaxation into a nearby energy minimum before emitting the photon, is responsible for its typically red-shifted energy with respect to the absorbed one. This effect is called Stokes shift, $Stokes\ shift = \lambda^{em}_{max} - \lambda^{abs}_{max}$.

2.2 Quantum Mechanical Methods

The quantum mechanical (QM) methods used for excited-state calculations in this thesis, are mainly the second order approximation to coupled cluster (CC2) and a semi empirical method called OM2/MRCI. For ground-state calculations the density functional based tight binding approximation (DFTB) is the main method of choice. Before discussing these methods, the basics underlying them will be briefly introduced.

All the QM methods I use, are a way to solve the time-independent "non-relativistic" Schrödinger equation of N identical fermions in the potential v_i^{e-n} of classical nuclei (BO approximation):

$$\begin{aligned} &\hat{H}\Psi = E\Psi\,; \\ &\hat{H} = \sum_{i=1}^{N}\hat{T}_i + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{i=1}^{N} v_i^{e-n}; \\ &\hat{T}_i = -\frac{1}{2}\nabla_i^2, \qquad\qquad v_i^{e-n} = -\sum_{A}\frac{Z_A}{|r_i - R_A|}. \end{aligned} \tag{2.3}$$

Atomic units are used thorough this chapter.

The exact solution to Eq. 2.3 is given by an antisymmetric N-particle wave function $\Psi(\mathbf{x}_1, ..., \mathbf{x}_N)$ of spin and spatial coordinates $\mathbf{x}_i = \{\mathbf{r}_i, \sigma\}$. Any explicit algebraic approach to solve Eq. 2.3 projects Ψ from the N-particle Hilbert space $\mathcal{H}^N$ into a finite subspace of it. Except for explicitly-correlated r_{12} methods [78], these approaches expand Ψ in a finite set of products of single-particle functions $\psi_i(x)$, also called orbitals:

$$\Psi(\mathbf{x}) = \sum_{pq...s} C_{pq...s}\psi_p(\mathbf{x}_1)\psi_q(\mathbf{x}_2)...\psi_s(\mathbf{x}_N). \tag{2.4}$$

As the Pauli postulate requires an antisymmetric wave function for fermions, Ψ can be re-written as a linear combination of Slater determinants

$$\Psi(\mathbf{x}) = \sum_{pq...s} C'_{pq...s}\Phi_{pq...s}\,, \qquad \Phi_{pq...s} = \frac{1}{N!}det(\psi_p(\mathbf{x}_1)\psi_q(\mathbf{x}_2)...\psi_s(\mathbf{x}_N)). \tag{2.5}$$

The one-particle functions ψ themselves are represented on a grid in real-space methods, or expanded into plane waves that satisfy periodic boundary conditions at the border of a finite box, or expanded in a basis of localized analytic functions, mostly atom-centered Gaussian- or Slater-type orbitals. Doing the latter:

$$\psi_i = \sum_j c_i^j \phi_j, \tag{2.6}$$

we get

$$\Psi(\mathbf{x}) = \sum_{pq...s} C'_{pq...s}\frac{1}{N!}det\left(\left(\sum_j c_p^j\phi_j(\mathbf{x}_1)\right)\left(\sum_l c_q^l\phi_l(\mathbf{x}_2)\right)...\left(\sum_f c_s^f\phi_f(\mathbf{x}_N)\right)\right). \tag{2.7}$$

Different quantum chemical approaches are distinguished by their strategy to determine the wave function and orbital coefficients C and c, respectively, and estimate the eigenvalues of $\hat{H}$ based on approximate guesses of these coefficients.

The exact solution of Eq. 2.3 in a given subspace is equivalent to the task of finding the stationary points of the Rayleigh quotient:

$$\bar{E} = \frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle}. \tag{2.8}$$

Methods that predict ground-state or excited-state energies based on the variational or Hylleraas–Undheim–MacDonald theorem [79, 80] respectively to find the stationary points of the Rayleigh quotient, are called variational methods. They have the advantage that they provide an explicit guess of the many-body wave function whose quality correlates with the deviation of the energy from the exact solution. The main drawback of these methods is that energies and energy differences are systematically overestimated (they are upper bounds to the exact values) and that the errors are usually larger than those of non-variational approaches at the same computational expense [81].

2.2.1 Hartree–Fock Theory

In Hartree–Fock (HF), the N-electron ground-state (GS) wave function is approximated by a single Slater determinant Φ of N spin-orbitals (thus we do not have wave function coefficients C and only the orbital coefficients c are calculated). As a result of this approximation, the interacting many electron problem is approximated to a non-interacting one, and we can solve it through a one-electron Hamiltonian with the other electrons' effect integrated as a potential. This approach can look like a mean field approximation, but it is actually more complex as the effective potential has a non-local part.

The HF approximation leaves the positions of the electrons uncorrelated, that is, the probability of finding one electron in a position depends on the average potential created by the rest of the electrons, but does not depend on the precise positions of them. Nevertheless there is a small correlation contribution coming from the Pauli exclusion effect introduced by the antisymmetry of the Slater determinant.

The HF equations are obtained by minimizing the energy corresponding to the mentioned single Slater determinant (for example by Lagrange's method of undetermined multipliers) getting the mentioned one-electron equations:

$$
\begin{aligned}
&(-\frac{1}{2}\nabla_1^2)\psi_i(\mathbf{x}_1) + v_1^{e-n}\psi_i(\mathbf{x}_1)\\
&+\sum_{j\neq i}\Big[\int d\mathbf{x}_2|\psi_j(\mathbf{x}_2)|^2 r_{12}^{-1}\Big]\psi_i(\mathbf{x}_1)\\
&-\sum_{j\neq i}\Big[\int d\mathbf{x}_2\psi_j^*(\mathbf{x}_2)\psi_i(\mathbf{x}_2)r_{12}^{-1}\Big]\psi_j(\mathbf{x}_1) = \epsilon_i\psi_i(\mathbf{x}_1).
\end{aligned}
\tag{2.9}
$$

In the second line of the previous equation we have what is defined as the Hartree potential v_H, and in the third a more tricky exchange potential $\hat{v}_x$ can be defined to rewrite equation 2.9 as:

$$
\Big[(-\frac{1}{2}\nabla_1^2) + v_1^{e-n} + v_H + \hat{v}_x\Big]\psi_i(\mathbf{x}_1) = \epsilon_i\psi_i(\mathbf{x}_1). \tag{2.10}
$$

The addend in the third line is non-local, because it introduces the ψ_j orbitals in the definition of $\psi_i(\mathbf{x}_1)$ (for $j \neq i$) without integrating them out through $\mathbf{x}_2$. A permutation operator $\hat{P}_{12}$ is necessary in order to define $\hat{v}_x$ and take $\psi_i(\mathbf{x}_1)$ as common factor. This allows the definition of the Fock operator $\hat{f}$ and to rewrite equation 2.9 as:

$$
\hat{f}(\mathbf{x}_1)\psi_i(\mathbf{x}_1) = \epsilon_i\psi_i(\mathbf{x}_1). \tag{2.11}
$$

Finally, using the basis expansion Eq. 2.6 we obtain the Roothaan equations [82, 83] :

$$
\sum_\nu F_{\mu\nu}c_i^\nu = \epsilon_i\sum_\nu S_{\mu\nu}c_i^\nu, \tag{2.12}
$$

where we have introduced the matrix element notation:

$$
\begin{aligned}
F_{\mu\nu} &= \int d\mathbf{x}_1\phi_\mu^*(\mathbf{x}_1)\hat{f}(\mathbf{x}_1)\phi_\nu(\mathbf{x}_1),\\
S_{\mu\nu} &= \int d\mathbf{x}_1\phi_\mu^*(\mathbf{x}_1)\phi_\nu(\mathbf{x}_1),
\end{aligned}
\tag{2.13}
$$

and hence they can be treated as matrices:

$$
\mathbf{FC} = \epsilon\,\mathbf{SC}. \tag{2.14}
$$

The Roothaan equations with a basis transformation gives the usual eigenvalue problem. It is easy to see at equations 2.14 and 2.13 that the HF equations depend on their own resulting orbitals, as they are needed to calculate the integrals. In the self-consistent field (SCF) approach, the one-electron equations are solved iteratively, starting from an initial guess and feeding the resulting new “electronic potential” (the integrals in 2.13) in each step until convergence is achieved.

HF gives us a set of one-electron wave functions (eigenfunctions) with their corresponding energies (eigenvalues) as large as two times the number of functions in the selected basis set (due to the two possible spin states). With the N ones of lower energy we get the HF GS wave function (and energy), but we also have the rest of them, the “unoccupied” ones, that can be used for post HF methods. These methods are usually discussed as different ways to recover the missing correlation in HF, and they are multi-configurational [84], in the sense that the resulting wave function consists of a linear combination of more than one slater determinant, more than one configuration.

2.2.2 Density Functional Theory

Density Functional Theory (DFT) development started as an approximation to HF, but the Hohenberg–Kohn theorems [85] showed that DFT has a bigger potential, it is an exact theory. Its strength relies in being a formalism that introduces electron correlation with electron density ρ (equation 2.15) as basic variable and not requiring any information about the many-body wave function.

$$\rho(\mathbf{x}_1) = N \int d\mathbf{x}_2 ... d\mathbf{x}_N |\Psi(\mathbf{x}_1, \mathbf{x}_2 ... \mathbf{x}_N)|^2 \tag{2.15}$$

The Hohenberg–Kohn theorems prove that for a given many-particle stationary system, the correspondence between the external potential to which it is subjected, its GS wave function, and the GS density is one to one. As a consequence we can deduce that it exists a functional of the density that gives the wave function $|\Psi[\rho]\rangle$ (nevertheless it must have a really complex form) and hence any property of the system can be found as a functional of the density, in principle with no explicit information of the wave function needed. This is not true in general when degenerate states come to play, but it can be demonstrated that for the energy it still holds. We can write the energy:

$$E[\rho] = \langle\Psi[\rho]|\hat{H}|\Psi[\rho]\rangle = \langle\Psi[\rho]|\sum_{i=1}^{N}\hat{T}_i + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{i=1}^{N}v_i^{e-n}|\Psi[\rho]\rangle, \tag{2.16}$$

where $F[\rho] = \langle\Psi[\rho]|\sum_{i=1}^{N}\hat{T}_i + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}}|\Psi[\rho]\rangle$ is a universal functional of ρ, independent of the external potential, which can be the usual electron interaction with the classical nuclei v_i^{e-n} or any more general v_{ext}. And hence we have:

$$E[\rho] = F[\rho] + \int d\mathbf{x}\rho(\mathbf{x})v^{e-n}(\mathbf{x})d\mathbf{x}. \tag{2.17}$$

The Hohenberg–Kohn theorems also demonstrate the existence of a minimum principle:

$$E[\rho_0] < E[\rho_0'] \qquad \Longleftrightarrow \qquad E_0 = \underset{\rho\epsilon\mathcal{N}}{min}E[\rho], \tag{2.18}$$

with ρ_0 and ρ_0' two GS densities from $\mathcal{N}$, which represents all GS densities which are obtained by solution of the Schrödinger equation.

However, the exact functional $F[\rho]$ is unknown, and it is the holy grail of DFT to find that functional.

Kohn and Sham introduced a scheme to approximate the energy functional [86]. They proved that for each interacting system there is an exact mapping to a non-interacting one with the same density (within some mathematical conditions),so we have both systems connected by:

$$\rho_0(\mathbf{x}) \equiv \rho_0^{KS}(\mathbf{x}), \tag{2.19}$$

where $\rho^{KS}(\mathbf{x})$ of the one particle wave functions ψ_i^{KS} of the Kohn–Sham system is:

$$\rho^{KS}(\mathbf{x}) = \sum_i |\psi_i^{KS}|^2, \tag{2.20}$$

and the subscript 0 indicates de GS minimum.

The strength of using a non-interacting system is, on the one hand, that for the non-interacting Kohn–Sham system $F^{KS}[\rho] \equiv T^{KS}[\rho] = \sum_{i=1}^{N} \langle \psi_i^{KS} | \hat{T}_i | \psi_i^{KS} \rangle$, which is universal for all possible Kohn–Sham systems and can be explicitly constructed from the auxiliary orbitals ψ_i^{KS}. Rewriting the interacting system energy in terms of the non-interacting kinetic energy:

$$E[\rho] = T^{KS}[\rho] + E_H[\rho] + E^{e-n}[\rho] + E_{xc}[\rho], \tag{2.21}$$

where

$$\begin{aligned} E_H[\rho] &= \frac{1}{2} \int \int d\mathbf{x}_1 d\mathbf{x}_2 \rho(\mathbf{x}_1) \rho(\mathbf{x}_2) r_{12}^{-1}, \\ E^{e-n}[\rho] &= \int d\mathbf{x}_1 \sum_i^N v_i^{e-n} \rho(\mathbf{x}_1), \end{aligned} \tag{2.22}$$

and $E_{xc}[\rho]$ is defined by Eq. 2.21, and is the term that absorbs all the complicated many-body effects not contained in the others. It is a functional as the others also are (T^{KS} is the less obvious but remember that it is $F[\rho]$ of the non-interacting system) and universal for the same reason.

On the other hand we know how to solve the non-interacting system. From here, imposing the condition that the KS orbitals minimize the energy, we find the Kohn-Sham equations to which KS orbitals ψ_i^{KS} obey:

$$\Big[(-\frac{1}{2} \nabla_1^2) + \hat{v}_{ext}^{KS} \Big] |\psi_i^{KS}\rangle = \epsilon_i |\psi_i^{KS}\rangle, \tag{2.23}$$

where the effective potential or the v_{ext}^{KS} is:

$$v_{ext}^{KS} = v_1^{e-n} + v_H + \hat{v}_{xc}, \tag{2.24}$$

with v_1^{e-n} and v_H as where defined in previous sections (equation 2.3 and equations 2.9 and 2.10 respectively) and

$$\hat{v}_{xc} = \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{x}_1)}, \tag{2.25}$$

note that unlike HF's $\hat{v}_x$ non-local potential this one is local.

We have all the terms to solve Eq. 2.23 except for the exchange correlation part. As we said before, the Holy grail and the reason why we have different DFT methods is how to approximate the energy functional (until here the equations are exact), to be more precise, the $E_{xc}[\rho]$ part of it.

To summarize, we know how to solve the equations of a non-interacting particle system getting the single-particle wave functions, like we did in the HF approximation. In HF we derive the single-electron equations from the single determinant assumption. This results in something similar to a non-interacting electron system with the electron-electron interaction effect introduced in an "average manner" in an effective potential. In the Kohn–Sham scheme the single determinant is not an assumption but the result of solving the problem through a non-interacting toy system. We also derive single-electron equations in an effective potential (which has a non-local part), but as Kohn and Sham proved thanks to the Hohenberg–Kohn theorems, the constructed toy system gives the exact solution of the real one in contrast to HF that is an approximation. This ability to solve (in principle exactly) any complex system through a much simpler non-interacting one is at a price, we do not get the interacting system's wave function, we just get the energy and the density. With the wave function we know how to evaluate the different observables, in DFT in principle all observables are a functional of the density (Hohenberg–Kohn theorems) but just the energy functional has been approximated. The non-interacting system, called the Kohn–Sham system, is not obtained as an approximation to the interacting one, but it is a mathematical tool that gives a set of orbitals that can be used to calculate the ground state density of the interacting system, and from this density, the energy.

2.2.3 Configuration Interaction and Coupled Cluster

The general expansion of the many-body wave function Eq. 2.4 for a finite single-particle basis set can be re-written as an expansion based on the HF ground-state wave function $|0\rangle$, which serves as a zeroth-order reference (in the spirit of perturbation theory):

$$|\psi\rangle = \left\{ \sum_{ia} C_{ia} \hat{E}_{ia} + \sum_{ia,jb} C_{ia,jb} \hat{E}_{ia} \hat{E}_{jb} + ... \right\} |0\rangle, \tag{2.26}$$

where the excitation operator $\hat{E}_{ia}$ replaces occupied by virtual orbitals in the HF determinant:

$$\hat{E}_{ia} = \sum_{p=1}^{N} |\psi_a(p)\rangle\langle\psi_i(p)| = \hat{a}^{\dagger}_{a\alpha}\hat{a}_{i\alpha} + \hat{a}^{\dagger}_{a\beta}\hat{a}_{i\beta}. \tag{2.27}$$

Or, if we define $\hat{I}_n$ as the operator of all n^{th} excitations (for example $\hat{I}_2 = \sum_{ia,jb} C_{ia,jb} \hat{E}_{ia} \hat{E}_{jb}$),

$$|\psi\rangle = \left\{ \hat{I}_1 + \hat{I}_2 + ... \right\} |0\rangle. \tag{2.28}$$

With this multi-configurational wave function, minimization of the Rayleigh quotient (2.8) under variation of the coefficients C of the different configurations with a fixed set of orbitals is called configuration interaction (CI), or full CI. If the expansion 2.28 is truncated after a certain class of excitations, the method is called configuration interaction singles, doubles, etc. (CIS, CISD...).

In coupled cluster theory (CC) we define the cluster operator $\hat{I} = \hat{I}_1 + \hat{I}_2 + \hat{I}_3 + ...$, and instead of applying it directly to $|0\rangle$, as in Eq. 2.28, we use an exponential ansatz of the wave function:

$$|\psi\rangle = e^{\hat{I}}|0\rangle. \tag{2.29}$$

The exponential operator $e^{\hat{I}}$ may be expanded as a Taylor series, and as in CI, $\hat{I}$ can be truncated after a certain class of excitations in the operator. If we consider only the $\hat{I}_1$ and $\hat{I}_2$ cluster operators of $\hat{I}$ (CCSD), we can write

$$e^{\hat{I}} = 1 + \hat{I} + \frac{1}{2!}\hat{I}^2 + \cdots = 1 + \hat{I}_1 + \hat{I}_2 + \frac{1}{2}\hat{I}_1^2 + \hat{I}_1\hat{I}_2 + \frac{1}{2}\hat{I}_2^2 + \cdots \tag{2.30}$$

In Eq. 2.30 we can see that applying the exponential cluster operator to the wave function, one can then generate more than doubly excited determinants due to the various powers of $\hat{I}_1$ and $\hat{I}_2$ that appear in the resulting expressions.

Both, CI and CC have an unfortunate scaling with the system size (approximately $\mathcal{O}(N^{2m+2})$, for excitation level m, higher than singles doubles level (CISD or CCSD) are rarely applied. But CCSD already offers a computationally affordable method that performs better than MP2 and CISD. Actually, CCSD(T), which takes into account triples in a perturbative way, has the same scaling as CCSD being much more precise.

In CI, The convergence of the correlation energy with respect to the dimension of the CI expansion is very slow, and it also has the problem of truncated CI methods not being size extensive. CC theory is an approach to achieve size extensivity, and highly superior to CI in recovering the so called dynamic correlation. The CC expansion converges to full CI, however, convergence is slow and multireference methods, including multireference CI, (see section 2.2.4 below) are considerably more efficient in the description of the so called static correlation. The distinction between dynamic and non-dynamic correlation is to some extent artificial and arbitrary. Roughly speaking, on the one hand we have the dynamic correlation, that refers to the correlation missing when the system is well represented by the HF determinant. The static correlation on the other hand, is somehow an extra correlation missing when the wave function is badly represented by a single determinant and hence, not only dynamic correlation is needed. The frontier between the two is blurry, as two strategies, each to recover one of those correlations, applied to the same system, should converge to the same result.

2.2.4 Multi-Reference Methods

The multi-configurational methods discussed above, when taken to practice, all have truncated expansions to be computationally affordable. The truncations are taking into account only the configurations that share most of the orbitals with the HF solution, the one they take as reference (in CC all configurations are taken into account but the coefficients solved are just the singles and doubles ones). For states that are well approximated by a single Slater determinant, i.e., for which HF lacks merely dynamic correlation, this is a good approach, but when this is not the case, when static correlation is important, i.e., when the HF wave function is not a good approximation, the truncation leaves out important configurations for the description of the wave function. Multi-reference methods solve this lack of static correlation by applying the truncated CI or CC expansion to a wave function including not only the HF determinant, but a number of other excited determinants also. For example, when studying a chemical reaction where one covalent bond breaks, if we want to calculate the transition state, two near degenerate Slater determinants are going to appear (the one that would be the HF solution if the bond is formed and the one that would be the solution if it is broken) and both should be taken as reference, even if one is going to be "excited" respect to the HF solution. Imagine that the near degenerate "excited" reference could be constructed by a single excitation from the HF solution. A MRCISD including those two references would in practice include some key triple and quadruple excitations, even if it is a singles doubles.

Making the method multi-reference we solve the static correlation problem without increasing the computational scaling, but the singles doubles expansion to recover the dynamic correlation might still be too heavy. To reduce it without sacrificing double excitations, it is usually applied the active space (AS) method inherited from multi-configurational SCF (MCSCF) methods as the complete active space SCF (CASSCF) method.

In most MCSCF methods an AS is defined as some excitations are more relevant than others (the excitation of a core electron for example is really rare). Only the electrons at a subset of the occupied molecular orbitals are susceptible of being excited to a subset of unoccupied or virtual molecular orbitals, the sum of both subspaces being the active space (in general the highest energy occupied orbitals and the lowest energy virtual ones). In the mentioned CASSCF method for example, all possible excitations inside the AS are taken into account. MCSCF methods search for the total energy minimum of a multi-configurational wave function with respect to both orbital and CI coefficients at the same time, and as they are taking into account multiple excitations within the AS they recover the static correlation, but the computational cost increases factorial with the number of active orbitals (m) and electrons (N), limiting the size of the AS and hence the recovery of the dynamic correlation.

In multi-reference CI (MRCI) the non-dynamic correlation can easily be described using a relatively small set of configurations, which may include highly-excited configurations. There are different ways to select the references, depending on the nature and knowledge about the problem. Then a CI expansion is applied to each of them, usually up to singles, or single and double excitations and inside an AS, because we need to limit the computational cost as it was explained before. This way MRCI achieves to incorporate both static and dynamic correlation in an efficient CI expansion to obtain quantitatively accurate results for properties of excited-states or highly correlated ground states that are badly described by the HF configuration.

2.3 Selected Ground State Methods

In this work the focus is on the excited-state properties and processes of fluorescent proteins, but I need reasonable ground state structures as starting point and also to determine the protonation states of the x-ray structures. Therefore I need ground state optimizations and even molecular dynamics (MD) calculations to determine them. For these I mainly use SCC-DFTB, a semi-empirical approximation to DFT, about three orders of magnitude faster than DFT methods, while describing the properties of chromophores with comparable accuracy [87]. We also perform MP2 calculations to get energies that can be compared with excited-state calculations at the CC2 level of theory.

2.3.1 SCC-DFTB

The Density Functional based Tight binding method is based on a second-order expansion of the Kohn-Sham total energy in Density-Functional Theory (DFT) with respect to charge density fluctuations [88]. It starts by initially considering that the system's density ρ_0 is composed of atomic densities, as if atoms in the system were free and neutral, and then including a fluctuation of it, $\rho_{min} = \rho_0 + \delta\rho$, and developing the following equations:

$$
\begin{aligned}
&E[\delta\rho] \approx \sum_i \langle \psi_i^{KS} | (-\frac{1}{2}\nabla_1^2) + v_1^{e-n} + v_H[\rho_0] + \hat{v}_{xc}[\rho_0] | \psi_i^{KS} \rangle \\
&+ \frac{1}{2}\int\int \left(\frac{\delta^2 E_{xc}[\rho_0]}{\delta\rho(\mathbf{x}_1)\delta\rho(\mathbf{x}_2)} + \frac{1}{|\mathbf{x}_1 - \mathbf{x}_2|} \right) \delta\rho(\mathbf{x}_1)\delta\rho(\mathbf{x}_2) - \frac{1}{2}\int v_H[\rho_0](\mathbf{x}_1)\rho_0(\mathbf{x}_1) \\
&+ E_{xc}[\rho_0] - \int \hat{v}_{xc}[\rho_0](\mathbf{x}_1)\rho_0(\mathbf{x}_1).
\end{aligned}
\tag{2.31}
$$

The zeroth order approach is equivalent to a common standard non-self-consistent (TB) scheme, and is already a good approach as the linear terms in the density fluctuations cancel. At second order a readily calculable expression for generalized Hamiltonian matrix elements can be derived. These are subjected to a 2-center approximation where the density fluctuations are represented by self consistent redistribution of Mulliken charges (SCC).The matrix elements (LCAO basis, Slater functions) are calculated with the DFT method it is approximating.

2.3.2 MP2

Mller and Plesset (MP) developed an scheme to apply the perturbation theory to the HF wave function [89]. In fact, the definition of the unperturbed Hamiltonian in this scheme (a sum over Fock operators, see Eq. 2.10 and 2.11) makes the HF wave function the zeroth-order wave function and the HF energy the first-order energy, so the perturbation of first order is already incorporated in HF. For small and medium-sized molecules the computational expense of the second-order MP treatment (MP2) is comparable to HF, giving much better results as it includes correlation energy (but it is not variational). The second-order perturbation expansion employs the same configuration space as CISD, but scales merely $\mathcal{O}(N^5)$ with the system size, and the resolution of the identity approximation reduces the scaling to $\mathcal{O}(N^4)$. The MP ansatz is, thus far, the only perturbation-theory-based ansatz to result in a size-extensive method, which is one of the reasons why it is used, although the perturbation is not small in it.

2.4 Selected Excited-State Methods

For the excited-state calculations I selected two methods quite different in nature, CC2 is an approximated coupled cluster ab-initio method, it is single reference and is not well suited when approaching a conical intersection, while OM2/MRCI is a semi-empirical method, multi-reference and adequate in the conical intersection region (except for the problems with the Born-Oppenheimer approximation, see section 2.1.1). This way a qualitative agreement of the two is rather unlikely a coincidence.

The two methods are also different in speed. OM2/MRCI is much faster and is my default option to explore the different problems, and CC2 is heavier but more established in the quantum chemistry community and, as was said before, complementary in its nature.

2.4.1 CC2

CC2 is an approximation to CCSD [90]. It takes into account all the single excitations as CCS does but unlike CCSD it does not take directly into account all the double ones. CC2 approximates the doubles equations to the form occurring in first order but with the singles retained to provide an approximate description of orbital relaxation. This gives an energy comparable to MP2, but with the ability to do ES calculations (MP2 is just for GS), and we choose it as a better compromise than CCSD for its scaling as N^5 instead of N^6. We have to be careful with CC2 as it is not reliable when approaching the conical intersections. It fails because it is a non Hermitian theory. In QM, all observables are represented by an Hermitian operator, but CC theories (and also CC2 approximation) sacrifice its hermiticity. As it is described by Andreas Köhn and Attila Tajti [91], this lack of hermiticity can result in a wrong topology of the conical intersections (degenerate eigenvalues) due to the breaking of real degeneracies and the appearance of accidental ones. For the numerical example they do with CH_2O, they show that the CC2 artifacts are for HOMO-LUMO energy differences lower than 0.02 eV.

2.4.2 OM2/MRCI

OM2/MRCI is based in the Orthogonalization Model 2 (OM2) [92], an orthogonalization correction to MNDO, which is a Neglect of Diferential Diatomic Overlap (NDDO) semiempirical method. The idea of semiempirical methods is to empirically parametrize an effective one-particle Hamiltonian and fit the parameters to experimental data (or calculations that treat correlation explicitly). Therefore it is formally an approximation to HF, but parametrized to include dynamic correlation. In contrast to other semiempirical methods, although no excited-state properties are used in the parameter fitting, OM2 gives sufficiently accurate single-particle energies for excited-state calculations. OM2/MRCI combines this effective one-particle Hamiltonian with the multi-reference configuration interaction method (MRCI), and yields accurate results for both ground and excited-state applications.

With the software we use, the molecular orbitals for the OM2/MRCI calculations can be determined using the restricted Hartree–Fock (RHF) or the restricted open-shell Hartree–Fock (ROHF). In RHF method the spatial parts of both spin states are restricted to be the same, with the electrons strictly paired. ROHF is more flexible, the spatial parts of both spin states are still restricted to be the same, but it includes some unpaired electrons. When selecting RHF to calculate the orbitals in OM2/MRCI the SCF is done for the closed shell wave function. When selecting ROHF, the SCF is done for a wave function with an electron excited from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO). ROHF is in principle designed to treat systems with odd number of electrons, but also systems with even number of electrons can be open shell in excited state (they are open shell for the first excited state we are interested on). In principle these ROHF orbitals would be better suited than the RHF ones for excited state calculations with truncated CI methods. Anyway, as we increase the AS the results given by both RHF and ROHF orbitals converge to the same one.

To select the references I took two things into account. In one hand that including just three references, or including up to eleven, gives very little increase in the incorporation of the purely non-dynamic correlation as Wanko et al. found [93].On the other hand that our reason to do a MR approach is that we are doing excited-state calculations (not that we are dealing with a ground state not well described by the HF determinant). Consequently, I chose to do the calculations with three references, the HF determinant and the single and double excitations from the HOMO to the LUMO.

The CI expansion was truncated after double excitations, and the AS was chosen case by case, always to be as large as possible within memory limitations, but with additional constraints in some cases. For example, for isomerization analysis calculations, taking care to avoid the active/inactive pair mixing that can occur at some intermediate along the reaction coordinate, when orbitals become nearly degenerate. In some cases, the active space was chosen smaller to reduce SCF convergence problems.

2.5 Quantum Mechanics–Molecular Mechanics approach (QM/MM)

The kind of fluorescent proteins we are studying, basically a compact cylinder made up by a 11-stranded β-barrel enclosing the autocatalytically formed chromophore, around 4000 atoms each, are too big to deserve the enormous effort of treating all their atoms QM. The key element we want to study is the chromophore inside the barrel, which is responsible for the fluorescence. Treating the chromophore, and maybe a few water molecules and sidechains nearby, quantum mechanically is enough to describe its functioning and characteristics. For this reason we are going to use the quantum mechanical/molecular mechanical (QM/MM) methodology [62, 63], that allows to treat the key part of the system QM and the rest with molecular mechanics.

2.5.1 Force Fields

In molecular mechanics (MM) molecules are described by a classical force field (FF), which is a parametrized model of the interactions between the atoms. It is usually an additive potential for the description of the total energy of the system as a function of the nuclear coordinates R_n , which commonly is of the analytic form

$$\begin{aligned}
&E_{MM} = E_{bonded} + E_{nonbonded} \\
&E_{bonded} = \sum_{bonds} K_d (d - d_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\phi [1 + cos(n\phi - \phi_0)] \\
&E_{nonbonded} = \sum_{m<n} \Big\{ \epsilon_{mn} \Big[\Big(\frac{\sigma_{mn}}{R_{mn}} \Big)^{12} - \Big(\frac{\sigma_{mn}}{R_{mn}} \Big)^{6} \Big] + \frac{Q_m Q_n}{R_{mn}} \Big\}.
\end{aligned} \tag{2.32}$$

The empirical parametrization of this potential function, in terms of force constants K_d, K_θ, K_ϕ, equilibrium coordinates d_0, θ_0, ϕ_0, LennardJones parameters σ

and ϵ, and atomic charges Q, defines the FF. The bond-torsion term is often complemented by an out-of-plane displacement potential (improper dihedrals). In this work we use the CHARMM22 FF [94]. It is important to note that force fields cannot accurately describe electronic reorganizations, such as electronic excitations, charge-fluctuations or -transfer, and the forming and breaking of chemical bonds, and therefore we need to treat a part of the molecule with QM.

2.5.2 Embedding Scheme

The QM/MM approach divides the molecule in a region that will be treated MM and a region that will be treated QM, as is shown in Fig. 2.2, and completes the MM potential (Eq. 2.32) with a term corresponding to the QM part and a QM/MM coupling term:

$$E_{tot} = E_{MM} + E_{QM} + E_{QM/MM}. \tag{2.33}$$

There are three types of embedding schemes of increasing complexity:

In the first one, the so called **mechanical embedding**, the two regions do not polarize each other. Their interaction is treated at the FF level, what means that van der Waals parameters and partial charges for the QM atoms are needed. For the vanderwaals parameters we have to assign each QM atom the so called FF's atom classes, or define new ones. The partial charges are derived from the wave function through population analysis or fitting to the electrostatic potential, and do not change once assigned. $E_{QM/MM}$ is then calculated with the Coulomb and Lennard-Jones potentials:

$$E^{mech}_{QM/MM} = \sum_{m}^{QM\ atoms} \sum_{n}^{MM\ atoms} \left(\frac{q_m q_n}{r_{mn}} + \epsilon_{mn} \left[\left(\frac{\sigma_{mn}}{R_{mn}} \right)^{12} - \left(\frac{\sigma_{mn}}{R_{mn}} \right)^{6} \right] \right). \tag{2.34}$$

Because the wave function is not polarized by the MM region, this approach is not adequate in most cases.

In the one called **electronic embedding** allows the wave function to be polarized by the MM region's charges but not the opposite. In this scheme the MM point charges are included in the QM calculation, polarizing the wave function. The $E_{QM/MM}$ term then excludes the electrostatic interaction of QM with MM atoms:

$$E^{elec}_{QM/MM} = \sum_{m}^{QM\ atoms} \sum_{n}^{MM\ atoms} \left(\epsilon_{mn} \left[\left(\frac{\sigma_{mn}}{R_{mn}} \right)^{12} - \left(\frac{\sigma_{mn}}{R_{mn}} \right)^{6} \right] \right). \tag{2.35}$$

This approach is the most widely used and the one selected in this work. It already gives very good results. The only problem is that the MM part is not polarized by the QM region.

Finally, the **polarized embedding** will also take into account the polarization of the MM region by the QM one. For this a polarizable FF is needed, and moreover, has the disadvantage that the mutual polarization of the regions has to be solved self consistently as it is interdependent, which multiplies the computational costs. We do not use this approach in this work.

2.5.3 Cutting Covalent Bonds

If QM and MM regions are defined in a way that they are connected by covalent bonds, some extra issues must be taken into account. The covalent bonds to be cut should be as unpolar as possible, as this way they are more electronically isolating. In proteins covalent bonds of sp^3 carbon atoms can be chosen. As long as the QM region does not get too big it is also important to chose the covalent bonds to be cut as far as possible from the region of interest, to minimize the effect of possible artifacts coming from the artificial coupling of the QM and MM regions along the bonds. The $E_{QM/MM}$ term will include the MM bonded interactions along the cut bonds (bonds, angles and diehdrals).

The most common approach to terminate the cut bond in the QM region is to use so called link atoms, which I also use in this work. Other methods exist

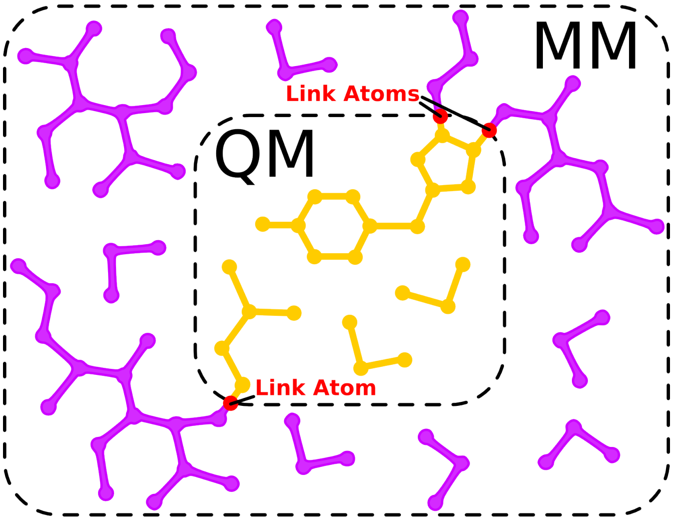

FIGURE 2.2: QM/MM concept

like the local self-consistent field (LSCF) method, 200,201 the generalized hybrid-orbital (GHO), 202,203 the frozen orbital, 204 frozen-core orbital (FCO), 205 the pseudobond, 206,207 quantumcapping potential (QCP), 206,208 effective group potential (EGP), 209,210 optimized effective Hamiltonian, 211 or the semiempirical connection-atom (CA) 212 approach (see Ref. 199 for an overview).

Link atoms replace the covalently bonded MM region at the boundary for the QM calculations. They are hydrogens because they are the simplest atoms with valence 1 (and their electronegativity is not far from the carbon one). They are placed along the cut bond (Fig. 2.2) at a typical $C - H$ distance from the corresponding QM atom (usually a carbon). One problem of the link atoms is that they are placed too close to the host MM atoms which almost always have a partial charge, so that the electrostatic interaction between them and the linc atom would be unphysically large. There are different solutions for this problem in the QM calculation. The option we use is related with how the CHARMM FF is built. The CHARMM FF has "charge groups" of atoms whose member's partial charges sum up to an integer charge. In this context we selected the divided frontier charge (DIV) scheme, where the partial charge of the MM frontier atom is deleted and evenly distributed between the rest of the atoms of its "charge group". An even simpler linking scheme is EXGROUP, which deletes all the charges of the "charge group" corresponding to the MM frontier atoms. There exist other solutions, like substituting the point charges of these frontier MM atoms by Gaussian distributions or scale down the charges. For DFTB as the QM method, different schemes have been tested by König et al. 218, and DIV was giving the best results. For the excited-state calculations we used the "*charge shift*" [95] option in ChemShell, which does the same, but it additionally adds two extra charges on the MM frontier atom, forming a dipole to compensate for the dipole created by the shift.

2.5.4 Software

For ground state calculations CHARMM package and SCC-DFTB [88] as the QM method using the QM/MM scheme built in the package with standard parameters.

For excited-state calculations (or ground state calculations with a method other than SCC-DFTB) ChemShell [96, 97] was used to couple CHARMM package (with the charmm22 force field) to either the MNDO99 program package [92, 98] for semiempirical OM2/MRCI or to TURBOMOLE [99] for CC2 [90] or other ab-initio QM methods.

2.6 Molecular Dynamics

Molecular dynamics simulations (MD) are a tool to investigate system's thermodynamics, not to calculate precise energies but explore the potential energy surface, conformations, bi-stabilities... It is a classical motion description, even for the QM atoms in the QM/MM scheme, as we are within the BO approximation and treat the nucleus classically. MDs consists in a sequence of discrete propagation steps, each of a finite time step. In each step the forces (energy gradients), and hence changes in velocities, are calculated and the positions propagated for the selected time step, usually 1fs, to be at least one order of magnitude smaller than the fastest motion in the system, the motion of hydrogens which have a vibrational period of around 10fs. Some further technical details about how we perform MDs in this work are explained in the following section.

2.7 Fluorescent Protein Setup

As starting point we use the information provided by the X-ray structures available in the protein data bank (pdb.org). Those structures are not ready to use, and before starting to explore their properties, we have to do what is called the setup. First, to obtain the X-ray structure the protein has to be crystallized, and the structure might differ from the one in solution (the interesting one for applications). We assume that the difference is small, and the MDs we perform allow for small structural readjustments.

The second reason, and the most important one, is that the X-ray structure does not provide information about the hydrogens (as they have only one electron, they are the weakest scattering atoms and they are essentially invisible in the X-ray diffraction). For most residues this does not mean much more than an optimization, or an MD if unclear hydrogen bonds are involved, but for the so called titratable residues, the ones that can assume more than one protonation state (the number of hydrogens they have), it can be difficult or even impossible to determine, as it not only depends on the pH, but also in their pK_a, which is strongly affected by the residue's environment and is not trivial to calculate accurately. If it is not possible to determine a unique protonation state, various plausible setups have to be taken into account for the same FP. The protonation state of titratable residues exposed to the solvent is less critical, if they are charged they are highly screened by the water, and in FPs the key region is buried inside the protein.

The procedure is as follows: The coordinates of the heavy atoms were obtained from the X-ray structures in the protein data bank. All water molecules present in the crystal structure were included in the calculations. The hydrogen atoms were added using the HBUILD utility in CHARMM [100]. Default protonation states were assumed except for titratable residues around the CRO, for which we applied a stepwise selection process of protonation patterns using structure optimizations and MD simulations. First the protonation states where chosen according to pK$_a$'s calculated with PROPKA [101] and, unless the case was clear, we constructed all plausible combinations obtaining different versions of the protein. Then we performed geometry optimizations on the different versions and we calculated the RMSD between the optimized and the X-ray geometries (only for heavy atoms). To filter the best candidates we required the RMSD to be lower than 0.6 Å for all the residues that are not solvent exposed. For some residues we accepted a higher RMSD as their atoms have a high B-factor. The B-factor describes the displacement of the atomic positions from an average value, hence bigger B-factors usually mean more flexible atoms. As a second filter we test the structural stability during MD simulations. The optimized structures were heated to 300 K within 20 ps, then propagated for 800 ps using the Nose-Hoover thermostat [102, 103], both with a time step of 1 fs. From the MD trajectories we calculate the mean structure, then we optimize it to check again the RMSD respect to the X-ray structure. It is important to be aware of the equivalent isomers (like a rotation interchanging the two sidechain oxygens of a glutamate for example) that can lead to distorted mean structures, which are not always properly solved by the minimization. Residues with a high RMSD must be examined in the mean structure and even directly in the MD to be sure about their real behavior. The setups fulfilling these criteria are selected for further calculations.

This process of building the setup is not automatic and needs time and intuition. It can be as time consuming as the investigation done with it afterwards.

The screening effect of bulk solvent was included by scaling the point charges of surface-exposed charged amino acids. The scaling factors were derived by solving the Poisson-Boltzmann equation, according to the charge scaling scheme proposed by Dinner et al. [104] To preserve the outer shape of the protein in the absence of bulk solvent, non-hydrogen atoms with a distance greater than 10 Å to the CRO were harmonically restrained to their position in the crystal structure. The force constants are derived from the B-factors [105]. It is important to note that harmonic constraints will be used only for the calculatios to prepare the setups, for the calculations to study the proteins atoms further than 10 Å from any QM atom will be kept frozen instead.

Chapter 3

Photoswitchable Fluorescent Proteins: OFF-state

3.1 Introduction

The successful application of reversibly-switchable fluorescent proteins in imaging techniques has raised an interest to understand the underlying photoswitching mechanism. This knowledge would allow a more rational design of RSFPs with a better performance in present imaging techniques, or suitable for new ones. In the study of RSFPs, most of the theoretical and experimental research has been focused on Dronpa, the first RSFP that was suitable for a wide range of practical applications [40].

The photoswitching mechanism in Dronpa [43], as in IrisFP and most of the known RSFPs, involves the isomerization of the chromophore (CRO) and a change in its protonation state (Figure 3.1).

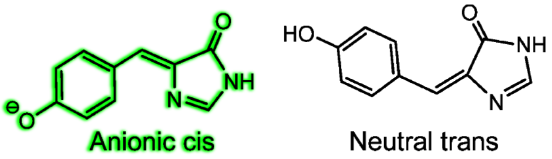

FIGURE 3.1: Chromophore in the ON- and OFF-states.

The OFF to ON switching reaction of these proteins is easier to study than the ON to OFF switching process because its quantum yield (QY) is orders of magnitude higher [2, 57], and therefore has been experimentally investigated by ultrafast spectroscopy in Dronpa [46, 51, 57], Dronpa2 [71, 72] and IrisFP [106].

Here I will study IrisFP and Dronpa, both having very similar structures for both ON- and OFF-states (see Fig. 3.2). Among the negative RSFPs mentioned in the introduction of the thesis, mTFP0.7 has a CRO environment not as similar as it is for those two (for example Thr59 is an Asn) and there is no available crystal structure for Dronpa2 in the OFF-state.

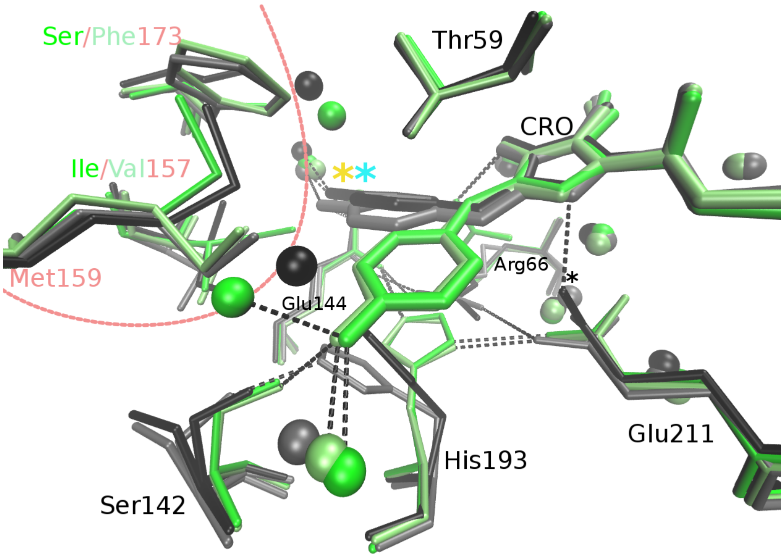

FIGURE 3.2: IrisFP in fluorescent (green) and dark (black) states; Dronpa in fluorescent (lime) and dark (gray) states. Delimited by a red dotted line a region with accumulation of structural differences between the two proteins, coming from the different conformation of Met159 and the substitutions Ser/Phe and Ile/Val of residues 173 and 157 respectively for IrisFP/Dronpa.

The first experiment of this kind proposed an excited-state proton transfer (ESPT) in Dronpa on the basis of a kinetic isotope effect (KIE) [46]. More recent studies could establish that the CRO in the excited OFF-state undergoes photoisomerization in its neutral form and that the deprotonation of the CRO follows as a consecutive step in the electronic ground state. A Fourier transform infrared study (FTIR) on Dronpa found a cis protonated ground state intermediate that decays on a ms time scale [51]. A UV-visible transient absorption spectroscopy study further supported the same conclusion through the study of the same intermediate [57]. Also a theoretical study by Xin Li et al. [49] excluded ESPT as the initial process and found photo-isomerization of the neutral CRO energetically feasible.

The most recent study by Colletier et al. comes to the same conclusion for IrisFP on the basis of ultrafast absorption spectroscopy. [106] In fact, the time-resolved data of IrisFP and Dronpa is quite similar and the main differences are the number of excited-state intermediates deduced from the decay components that are obtained by fitting the time-resolved spectra with kinetic models. While in Dronpa the photoisomerization is associated with three decay components (2.1, 10, and 64 ps), which fit a model of two excited intermediates decaying in parallel [57], in IrisFP a sequential decay via two excited intermediate states was proposed [106].

Although the structure and the first ground-state intermediate of both systems are very similar, it is still unclear how to interpret the observed differences in the excited-state dynamics. Moreover, the previous studies could not explain why the ESPT is blocked in RSFPs whereas it is observed in GFP variants that feature an intact proton wire to a glutamate acceptor. In general, ESPT is assumed as a consequence of the reduced proton affinity of the CRO in the S_1 excited state. The suppression of ESPT in RSFPs is remarkable in particular in IrisFP, where the CRO forms a direct H-bond to a glutamate and proton transfer is expected to be essentially barrierless.

In this chapter, I investigate the relation between isomerization and proton transfer (PT) in the excited state of IrisFP and Dronpa and their role for the OFF to ON switching mechanism, i.e., the excited-state dynamics of the proteins in their OFF-state. I theoretically model these processes by searching for the lowest-energy pathways on the excited-state potential energy surface using constrained geometry optimizations. These calculations are based on the quantum mechanics–molecular mechanics (QM/MM) approach, as described in chapter 2. Details are given in the following section.

3.2 Methods

Setup. The coordinates of the heavy atoms of IrisFP and Dronpa were obtained from the X-ray structures of Adam et al. [59] and Andresen et al. [43] (PDB codes 2VVI and 2POX), respectively. Protonation states and the position of H-atoms where determined following the process described in Chapter 2. Default protonation states were assumed for all residues except Glu211 and His193 (see below). The residues are named according to the numbering in Dronpa thorough the chapter.

For IrisFP it was difficult to achieve a representative structure in good agreement with the X-ray structure due to the high mobility of His193. I obtained three setups that are consistent with the X-ray data; they are called A1, A2, and B (see Appendix–A.1 for more details on the protonation state selection). In all three setups, Glu211 is protonated, which is necessary to form a H-bond to the CRO, as evident in the X-ray structure (see Figure 3.5). This is an important difference to Dronpa, as will be discussed later. Setups A1 and A2 share the same protonation states (His193 is deprotonated at N_ϵ) but are derived from different monomer chains (A and D, respectively) of the X-ray tetramer. In setup B (derived from monomer chain B), His193 is deprotonated at N_δ, instead. Our results obtained with setup B are qualitatively the same as those obtained with setup A1 and are provided in the Appendix–A.2.

For Dronpa, only one setup was found consistent with the X-ray data, with His193 (IrisFP numeration) deprotonated at N_δ and Glu211 deprotonated. One of the four monomer chains of the crystal structure, however, indicates an H-bond between Glu211 and the chromophore imidazole, which is only possible with a protonated Glu211. Although this monomer structure is likely an artifact of the crystallization, I will consider it as an intermediate case between IrisFP and Dronpa and refer to it as setup “Y”, in discrimination to the default setup “Z”, which has a deprotonated Glu211 (results for Y1 are in Appendix–A.6). Setup Y shares the same protonation states as IrisFP’s setup B.

QM region. For IrisFP I use three different QM regions, as shown in Figure 3.3. For preparatory ground-state calculations I use QM region QM1, which includes the three residues forming the CRO (Figure 3.3a, thick sticks) and two backbone atoms of the neighbor residues Phe61 (C and O) and Asn65 (N and HN) (spheres in Figure 3.3a). Excited-state (ES) calculations where performed with QM region QM2, which only includes the conjugated part of the CRO, or (when considering PT) with QM region QM3, which adds the Glu144 side chain.

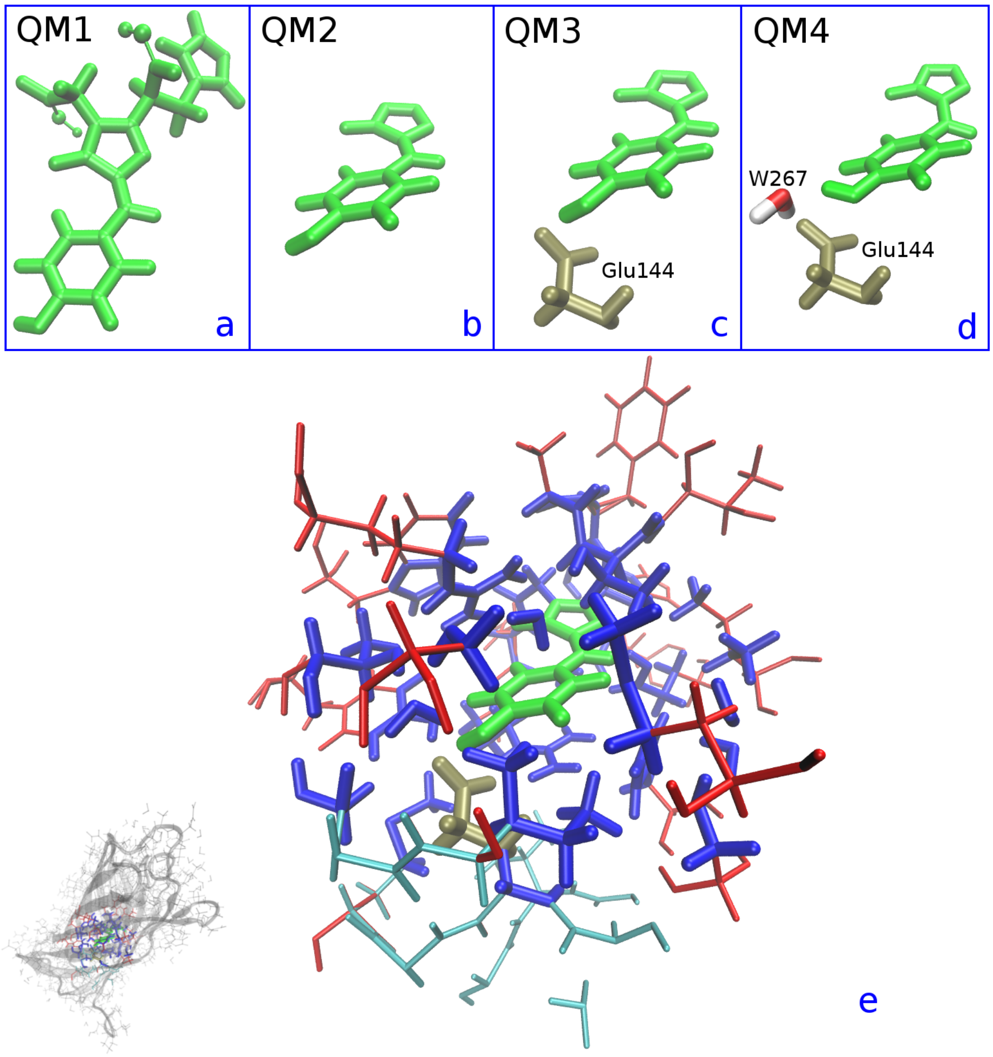

FIGURE 3.3: QM regions: a) QM1 including the whole CRO and two backbone atoms from each of the two neighboring residues. b) QM2 including just the conjugated CRO atoms c) QM3, includes Glu144 side chain d) QM4, includes H-bond mediating water267 for Dronpa e) conjugated CRO atoms (green), Glu144 side chain (tan), QM2 AR (blue), extra atoms to the AR for QM3 (cyan). In red the rest of the atoms of the residues having just some of them inside the AR (example for IrisFP).

For Dronpa the QM regions where the same ones plus a QM4 (Figure 3.3d), which is as QM3 including the water that mediates the H-bond between Glu144 and CRO described in the introduction (Figure 3.5). It is used for the calculations considering water mediated PT.

Pathway calculations. The validity of OM2/MRCI for the description of photoisomerization reactions has been demonstrated for various chromophores and GFP-like systems, [107, 108] in particular for HcRed [70]. Also the relative proton affinities of the CRO and the glutamic acid are well described by OM2 in the GS [93]. In the Appendix–A.3, I show that also in the ES, the proton affinities of OM2 compare well to CC2 results, but that OM2/MRCI underestimates the PT barrier in the GS when compared with MP2 and in the ES OM2/MRCI localizes the proton in-between the two acceptors. It was not tested though for a water-mediated H-bond. For these reasons we will use only CC2 for the calculations involving PT and both CC2 and OM2/MRCI for the others (except for the potential energy surface scans in 2D, only reasonably affordable with the faster OM2/MRCI method).

The active orbital space was optimized for each QM system individually as was explained in Chapter 2. The resulting active spaces are shown in Appendix–A.5.

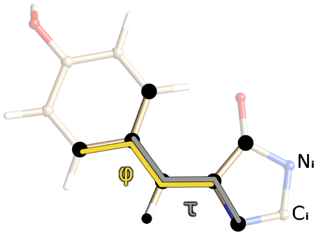

FIGURE 3.4: In black the atoms involved in the dihedrals constrained to do the τ-ϕ 2D scanning. ϕ and τ dihedrals in yellow and gray respectively.

All the energy profiles presented in this chapter are obtained by constrained pathway optimizations along a reaction coordinate, which is either the τ or ϕ dihedral for the isomerization pathways (see Figure 3.4) or λ for the PT pathway (Eq. 3.1 in section 3.3.1). During the optimization, atoms further than 10 Å from any QM atom were kept fixed at their initial positions (see above) and all other atoms were free of constraints (Figure 3.3). As the OFF to ON switching happens on a fs to ps time scale, long-range relaxation is not expected to be relevant. For the calculated potential energy surfaces, I scan both τ and ϕ dihedrals while maintaining the hydrogen atom of the bridge connecting the phenolate and the imidazole ring in plane with the three carbon atoms of the bridge. All the redundant dihedrals including the atoms marked in black in Figure 3.4 are also constrained to preserve co-planarity.

3.3 Results and Discussion

For the OFF-state I will consider two possible scenarios for the ES photodynamics: (1) The neutral CRO isomerizes and the system decays to the GS without any ESPT; (2) ESPT is allowed to happen before or during isomerization. The first scenario can be modeled using a minimal QM region that contains only the CRO (QM2 in Figure 3.3b). The second scenario will be tested using larger QM regions (QM3 for IrisFP, Figure 3.3c or QM4 for Dronpa, Figure 3.3d), which allows the proton to freely transfer between Glu144 and the CRO. Different QM regions are requited because in IrisFP the PT is direct whereas in Dronpa it is water-mediated (Figure 3.5). These two scenarios are studied in subsections 3.3.1 and 3.3.2 and tested for both proteins in comparison.

3.3.1 Isomerization of the neutral chromophore

To search for the optimal isomerization pathway of the neutral CRO I calculated a relaxed 2D scan of the ES potential energy surface in the τ and ϕ dihedrals (Figure 3.4). Figures 3.6 and 3.8 show a 2D projection of the potential energy surface for IrisFP and Dronpa, respectively. The energy (eV) is color-coded.

The global minimum of the ES surface is at (τ,ϕ)=($-160°$,$-160°$) for IrisFP and around (τ,ϕ)=($160°$,$-160°$) for Dronpa, slightly displaced from the FC point

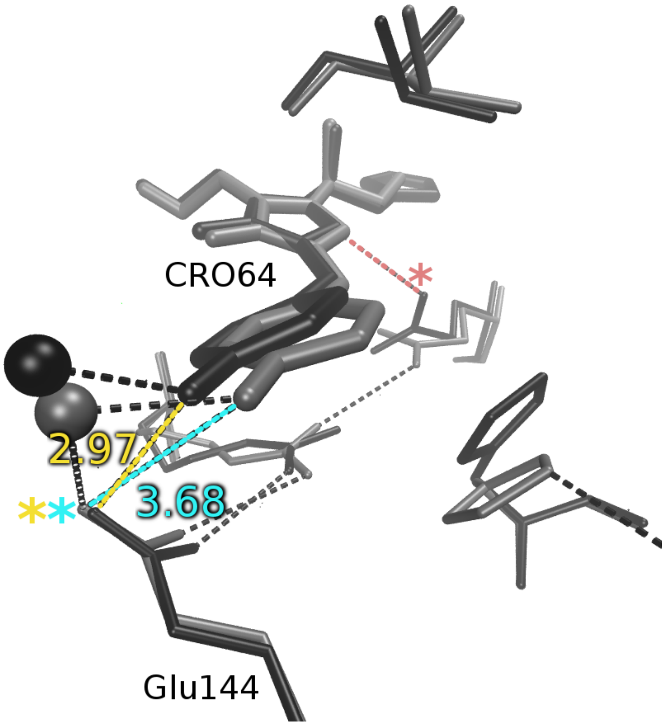

FIGURE 3.5: IrisFP CRO-Glu144 distance in IrisFP (H-bond in yellow, black structure) and in Dronpa (H-bond in blue, gray structure). In red IrisFP's Glu211-CRO H-bond.

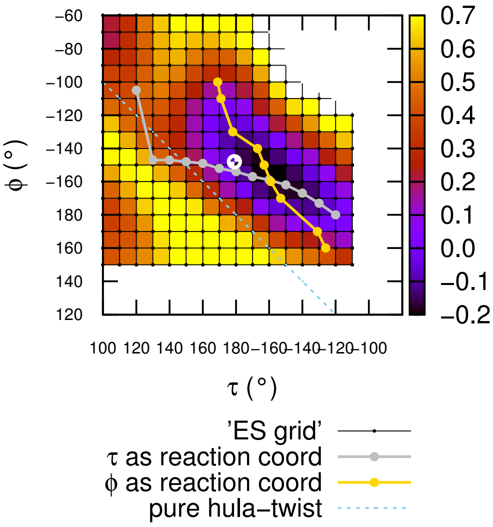

FIGURE 3.6: **IrisFP** relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The white dot is the FC point.

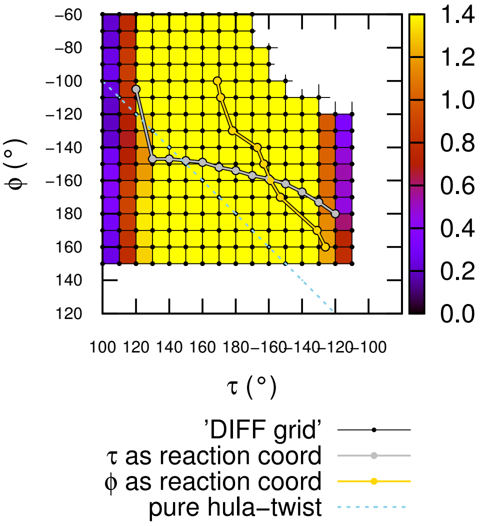

FIGURE 3.7: **IrisFP** ES-GS energy difference of the relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The lower the value the nearer to a CI seam.

(white dot). In both proteins the ES surface describes a stretched diagonal valley, roughly parallel to the ideal space-conserving hula-twist coordinate (grey dotted

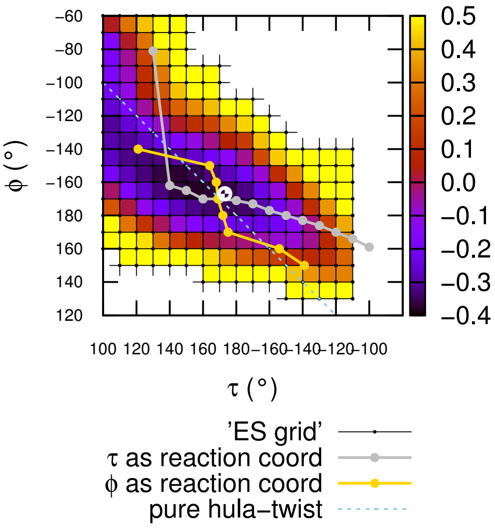

FIGURE 3.8: **Dronpa** relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The white dot is the FC point.

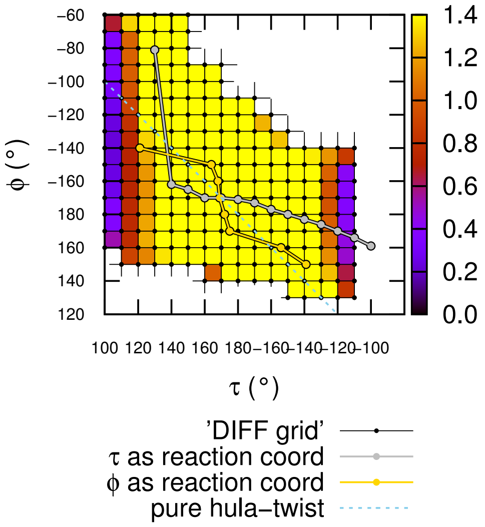

FIGURE 3.9: **Dronpa** ES-GS energy difference of the relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The lower the value the nearer to a CI seam.

line).

Figures 3.7 and 3.9 show the corresponding energy gap between the ground and excited state. Two regions are visible where the energy gap approaches zero, i.e., where the two surfaces cross in a CI seam. In these regions τ is strongly twisted, but the CI seam extends along the ϕ axis over a wide range.

The two CI regions can be reached from the ES minimum, following the bottom of the valley. The barriers for reaching the CI, however, are asymmetric and quite different in the two proteins. While IrisFP favours isomerization in the direction of negative τ, the barriers in Dronpa dictate isomerization towards positive τ. In both proteins, the minimum is in-between the FC point and the CI with the lower barrier (the FC point is at $(\tau,\phi)=(180°,-150°)$ for IrisFP and $(\tau,\phi)=(170°,-170°)$ for Dronpa). Therefore, the initial excited-state relaxation is in the direction of the favoured isomerization towards the CI.

As distortion of the methylidine bridge from co-planarity may lower the energy at which the CI is reached, I reduced the constraints to a single one (either τ or ϕ), in orther to obtain more realistic barriers. The ES pathways optimized with constrained τ (ϕ) are marked as gray (yellow) curves in the 2D-scan plots. Figure 3.10 shows the corresponding energy profiles (Figure 3.11 for Dronpa).

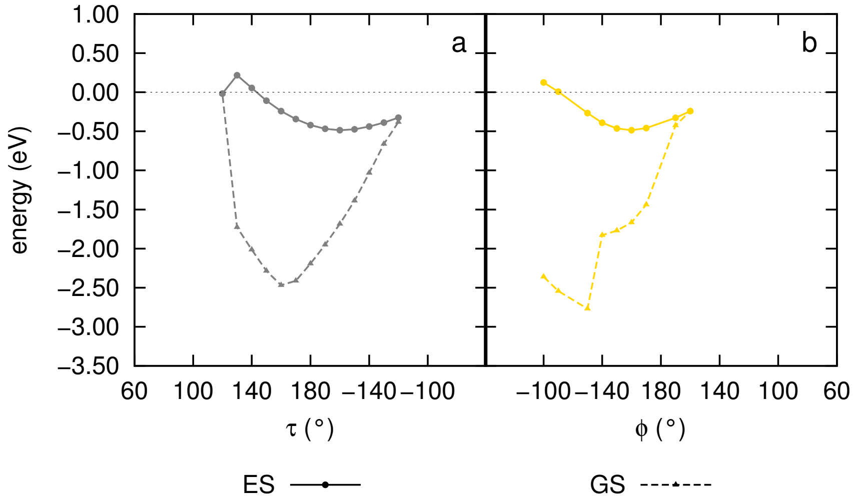

FIGURE 3.10: **IrisFP** relaxed (ES-optimized) energy profile of the τ-constrained (a) and ϕ-constrained (b) isomerization pathway for the neutral CRO. Zero energy is set to the FC point.

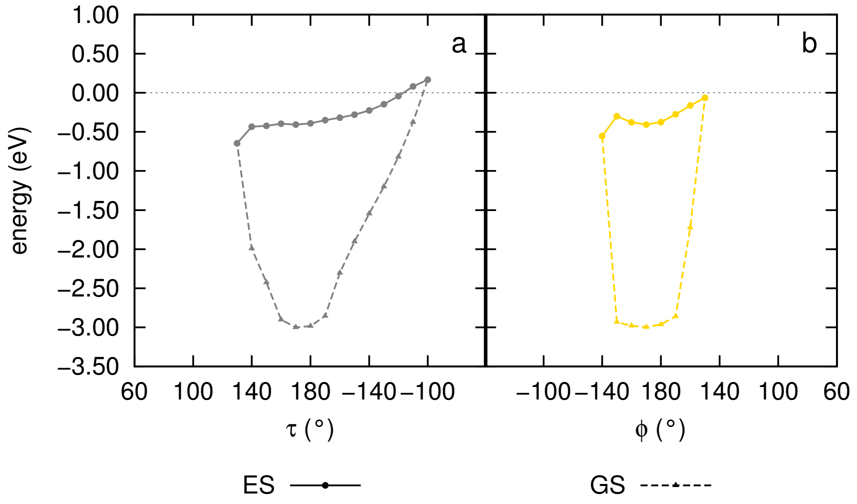

FIGURE 3.11: **Dronpa** relaxed (ES-optimized) energy profile of the τ-constrained (a) and ϕ-constrained (b) isomerization pathway for the neutral CRO. Zero energy is set to the FC point.

With both individual constraints, I obtain pathways close to the energy valley of the 2D scan. The τ-constrained pathways represent better the minimum energy pathway towards the preferred CI regions, with a small barrier of 0.20 eV above the ES minimum for IrisFP and a barrierless pathway in the case of Dronpa. In the other direction, the path rises above the FC point for IrisFP and involves a barrier of 0.34 eV for Dronpa. Using a single constraint, the pathways reach the

CIs closer to the FC point (with less distortion of the constrained dihedral) and at lower energies. The preferred CI is 0.32 eV (0.65 eV) below the FC point for IrisFP (Dronpa).

In Figures 3.12 and 3.13 I show the structures of each protein at the FC point and at the favoured CI. Due to the hula-twist character of the isomerization, the distortion is mostly on the bridge and requires little movement of the phenolate ring. The initial orientation of the two rings, however, is slightly different in the two proteins. Due to this difference in the initial values of the dihedrals, the pathway in IrisFP is a shifted parallel to the ideal hula-twist coordinate (grey dotted line in Figure 3.6) and has a smaller variation of ϕ than τ in Dronpa. This applies to both the shape of the valley in the 2D-scan and the single-constrained pathways, but the latter deviate more from the linear hula-twist coordinate. Especially the τ-constrained path bends for τ¡140, as ϕ makes a big jump in the last step before reaching the CI seam (see Figures 3.8 and 3.6), but it fits with the shape of the valley and is lower in energy than the ϕ-constrained path. These jumps in ϕ produce a kink in the energy profile (Figures 3.10 and 3.11) and simply come from the discretization of the path along the τ coordinate.

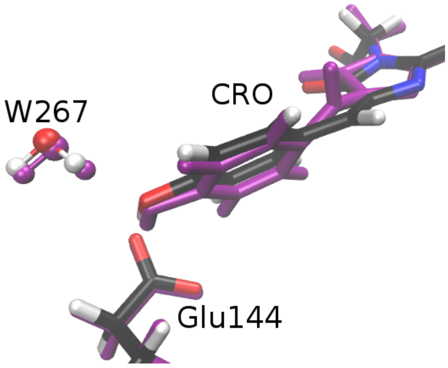

FIGURE 3.12: **IrisFP** structure at the CI point for negative τ isomerization (purple) compared with the FC structure (black).

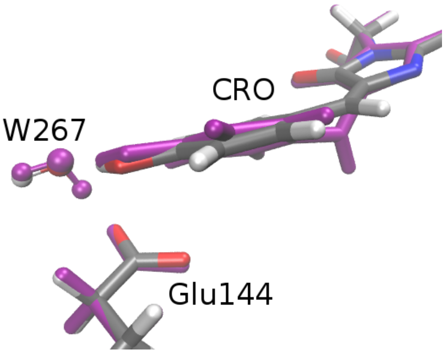

FIGURE 3.13: **Dronpa** structure at the CI point for positive τ isomerization (purple) compared with the FC structure (gray).

In summary, I found for both proteins a hula-twist-like isomerization pathway along a valley connecting two CI regions, which have a positively or negatively twisted τ dihedral. The valley is asymmetric, with different slopes towards the two CI regions, enforcing a negative τ isomerization in IrisFP and a positive one in Dronpa. In both proteins, the lower-energy CI region is below the FC point, which is consistent with the observed radiationless decay of the OFF-state and shows that trans-cis isomerization of the neutral CRO is energetically feasible without ESPT.

3.3.2 Concerted proton transfer and isomerization.

In order to model the second scenario and allow ESPT during the isomerization, I add the Glu144 side chain to the QM region (QM3, Figure 3.3c) for IrisFP's direct PT (Figure 3.23) and additionally water267 (QM4, Figure 3.3d) to describe Dronpa's water-mediated PT. (Figure 3.20). As explained in section 3.2, I use CC2 for these calculations, as it gives a more accurate description of the PT barrier. For the direct transfer, I define the reaction coordinate λ as the difference of the two distances of the proton to the oxygens of Glu144 and the CRO (Eq. 3.1).

$$\lambda = d(\mathrm{O}_{\mathrm{E144}}, \mathrm{H}^+) - d(\mathrm{H}^+, \mathrm{O}_{\mathrm{CRO}}) \tag{3.1}$$

This way, a positive λ value means that the proton is nearer to the CRO than to Glu144, as in the ground state. For the water-mediated PT, I define the reaction coordinate λ' as the difference of the Glu144–water267 and the water267–CRO O-H distance (Eq. 3.2 and Figure 3.20).

$$\lambda' = d(\mathrm{O}_{\mathrm{E144}}, \mathrm{H}^+_{\mathrm{W267}}) - d(\mathrm{H}^+_{\mathrm{CRO}}, \mathrm{O}_{\mathrm{CRO}}) \tag{3.2}$$

This way, a positive λ' value means that the phenolate's proton is nearer to the CRO than the water267's proton to Glu144 (neutral CRO and anionic Glu144), as in the ground state.

In the following, I use λ and λ' either as a constraint to calculate the PT pathway, or as an observable that may change along a τ/ϕ-constrained isomerization pathway.

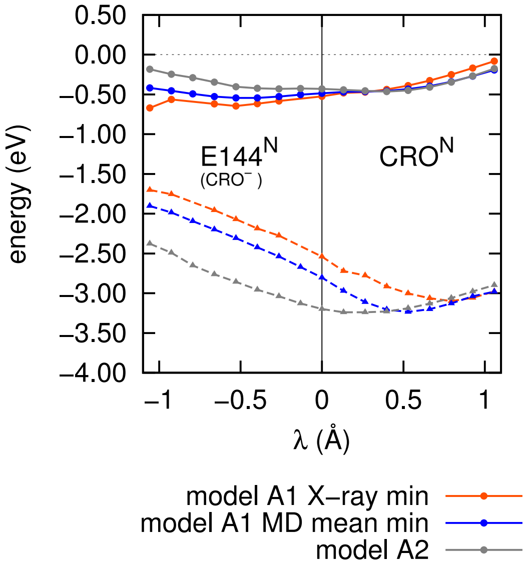

FIGURE 3.14: **IrisFP** ES energy profile along the λ-constrained PT pathway. CC2 is used as QM method. Zero energy is set to the FC point.

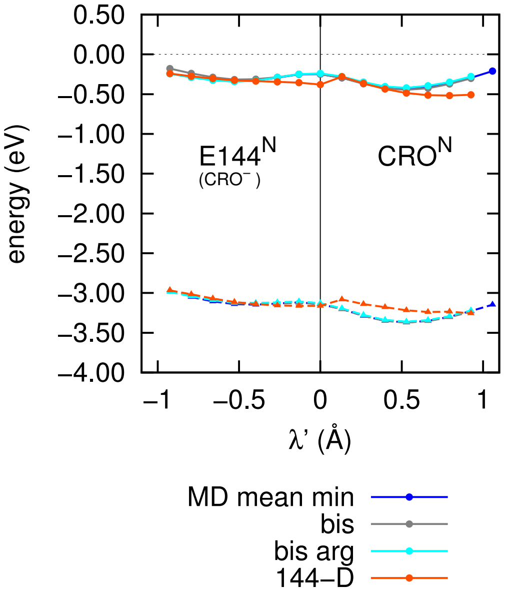

FIGURE 3.15: **Dronpa** ES energy profile along the λ'-constrained PT pathway, except for 144-D, where the constrain is λ. CC2 is used as QM method. Zero energy is set to the FC point.

Figures 3.14 and 3.15 show the ESPT pathway optimized without constraints on τ or ϕ. A1 and A2 are two different models of IrisFP in the dark state (see section 3.2), which differ only structurally but share the same protonation states. "MD mean min" refers to the minimized structure starting from the coordinates averaged over a 600 ps MD simulation, whereas "X-ray min" refers to the optimized X-ray structure. For Dronpa I will discuss four different structures produced with the same setup (model Z) to study how the structural variability in the GS affects the ES pathway. They are called "MD mean min", "bis", "bis arg" and "144-D". "MD mean min" refers to the minimized structure starting from the coordinates averaged over a 600 ps MD simulation, whereas "bis", "bis arg" and "144-D" are minimizations of individual snapshots chosen from that MD. Originally 12 snapshots (one every 50 ps) where taken, but in the minimization they converged to only 3 different structures. "bis" is similar to "MD mean min"; "bis arg" has Arg91 nearer to the CRO; "144-D" has a direct H-bond between Glu144 and the CRO, instead via water267, and the reaction coordinate is λ instead of λ'. Independent from their origin, the three (four) structures in Figure 3.14 (Figure 3.15) can be

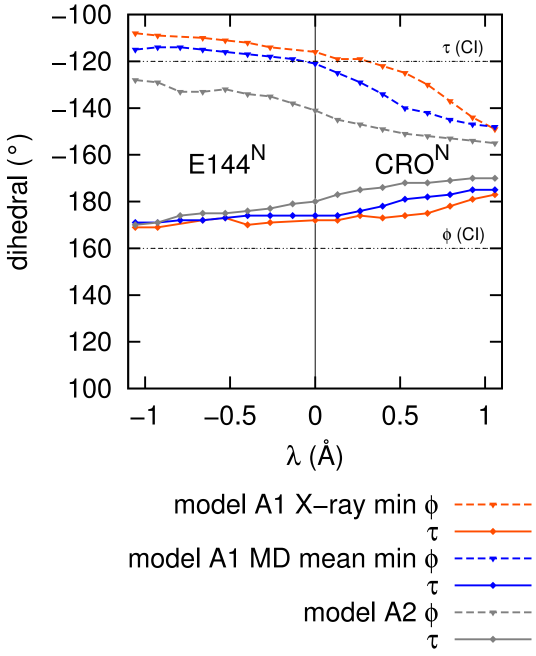

FIGURE 3.16: **IrisFP** τ and ϕ dihedrals along the λ-constrained PT pathway (CC2 is used as the QM method). The upper and lower horizontal lines mark the τ and ϕ values at the CI, respectively.

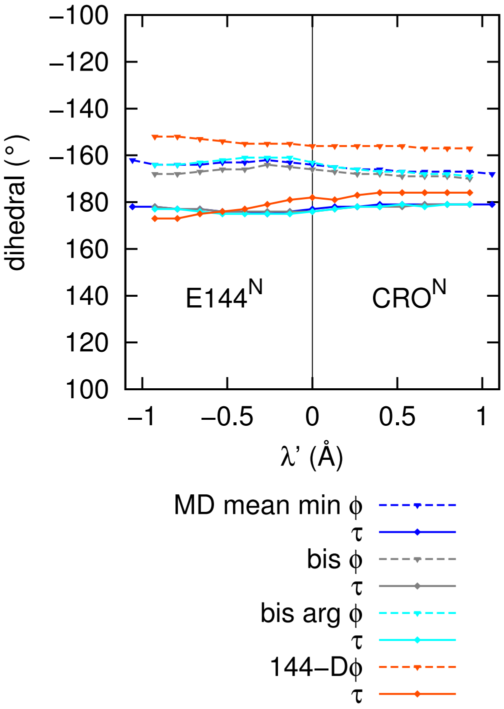

FIGURE 3.17: **Dronpa** τ and ϕ dihedrals along the λ'-constrained PT pathway (CC2 is used as the QM method).

considered as different representative points of the configuration space of each system.

In case of IrisFP, the three examples show that the lowest energy position of the proton changes due to small structural fluctuations that I observe during the room temperature MD simulations. In all three cases the excess energy from the FC point allows the proton to occupy both positions before any isomerization dynamics starts. For A2 (Figure 3.14, gray curve) two essentially isoenergetic ES minima are found separated by a very small barrier. For the other two structures (Figure 3.14, blue curves), the energy profile has a similar shape but is sloped and features only one minimum where the proton is nearer to Glu144.

In the case of Dronpa, the CRO is always more stable in its neutral state and the barrier between the two protonation states is more pronounced. Still, the excess energy at the FC point would allow the proton to occupy both positions as the transition state is 0.25 eV below the FC point. The energy profile obtained with “MD mean min”, “bis”, and “bis arg” are essentially identical and show two minima, separated by a barrier of 0.2 eV, one describing an anionic CRO and a neutral Glu and the other one a neutral CRO and an anionic Glu. The latter minimum is 0.1 eV lower in energy. For the more special case of “144-D”, with a direct H-bond between CRO and Glu144 (like in IrisFP) the positions of the two minima are shifted towards larger λ. The one with neutral CRO and anionic Glu144 is at λ=0.8 and in the other one the proton is in the middle. The latter minimum is 0.14 eV higher in energy and is produced by a discontinuity in the curve—the barrier of 0.24 eV is due to a jump of water267 forming a H-bond to either the anionic CRO or to the anionic Glu144.

Figures 3.16 and 3.17 show how the τ and ϕ dihedrals change along the ESPT pathway. The pattern is remarkably different for IrisFP or Dronpa. While for Dronpa the dihedrals are almost constant, they are clearly correlated with λ in IrisFP. At the FC point, IrisFP has (τ,ϕ)=(180°,−150°) and λ=0.5. Enforcing ESPT to Glu144 by decreasing λ, the dihedrals move along the hula-twist valley away from the preferred CI region. This indicates that ESPT and photoisomerization are competitive processes, i.e., an initial ESPT hinders the isomerization in IrisFP.

Figures 3.18a and 3.19a show the τ-constrained isomerization pathways, as calculated before, but now using the CC2 method and the larger QM3 (QM4 for Dronpa) QM region to allow ESPT (dark blue curves). I also calculate the pathway with CC2 and QM2 for comparison.

As expected from the PT pathways discussed before (Figure 3.16), IrisFP shows a correlation between isomerization and ESPT. Rotating τ in the non-preferred direction (towards the higher-energy CI), the CRO deprotonates, whereas the neutral CRO is stabilized for positive twist of τ (towards the lower-energy CI). This trend is completely missing in Dronpa. Consistent with the ESPT energy profile for Dronpa (Figure 3.17), the protonation state does not change during the isomerization (in neither direction).

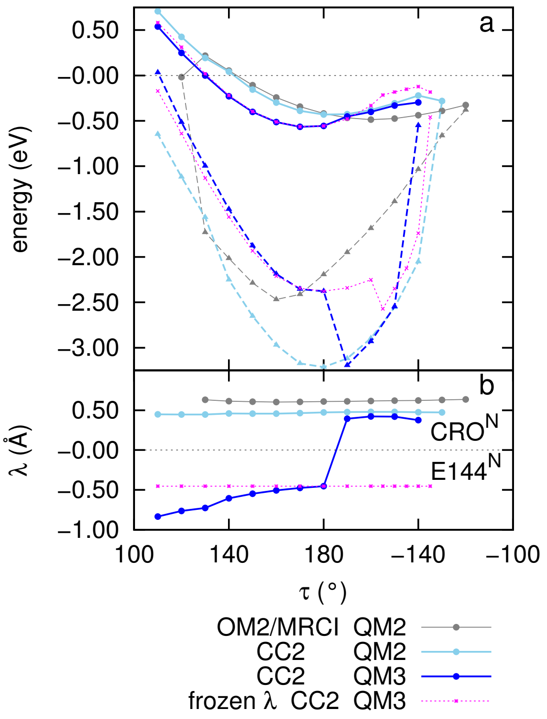

FIGURE 3.18: **IrisFP** (a) Energy profile along the τ-constrained isomerization pathway with QM2 (light blue), with QM3 and optimizing from neutral CRO (dark blue) and with QM3 and frozen λ to anionic CRO (magenta) using CC2 as the QM method. The OM2/MRCI profile for QM2 and neutral CRO is shown for comparison (gray). (b) The corresponding λ values. Zero energy is set to the FC point.

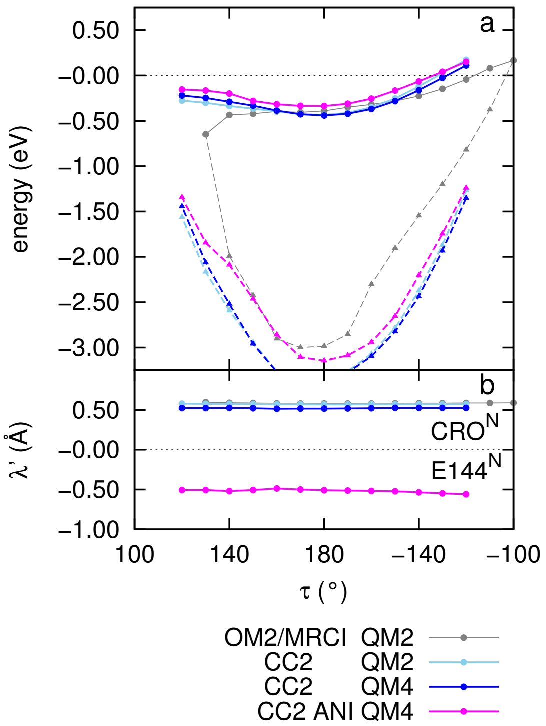

FIGURE 3.19: **Dronpa** (a) Energy profile along the τ-constrained isomerization pathway with QM2 (light blue), with QM4 and optimizing from neutral CRO (dark blue) and with QM4 and from anionic CRO (magenta) using CC2 as the QM method. The OM2/MRCI profile for QM2 and neutral CRO is shown for comparison (gray). (b) The corresponding λ' values. Zero energy is set to the FC point.

For IrisFP, the energy profiles obtained with the extended QM regions agree with those obtained with the minimal QM2 region (light blue curves) as long as the proton is close to the CRO (negative τ values). For positive τ, the ES energy drops below the QM2 curve when the CRO loses the proton, which shifts the ES minimum away from the (lower-energy) CI region and causes a redshift in the emission energy. The energy profiles are calculated with the "model A1 MD mean min" structure. This structure and the "model A1 Xray min" one perform similar curves in Figure 3.14, therefore a similar behaviour is expected. However, "model A2" has the opposite energy balance for the CRO protonation state, and ESPT would not happen in any dihedral value.In the case of Dronpa, the ES energy profile looks the same for both minimal and extended QM region as the barrier in the water-mediated PT prevents deprotonation of the CRO.

To test the same isomerization with the anionic CRO and make a direct energetic comparison, I repeat the CC2 QM3 (QM4 for Dronpa) calculations starting with the proton binding to Glu144. In the case of IrisFP, this is achieved by freezing λ to a value that confines the proton near to Glu144. For Dronpa no such constraint is needed as the barrier is high enough to prevent PT during the path optimization. The resulting path is shown as magenta curve in Figures 3.16 and 3.17. In consequence, the jump in the emission energy of IrisFP at $\tau = -170°$ disappears and the energy keeps rising towards the CI. At $\tau = -150°$, however, the emission energy drops again as ϕ relaxes to a planar configuration close to the CI (compare Figure 3.6). Due to the λ-constraint, the barrier towards the CI is 0.17 eV higher, which indicates that the neutral CRO is stabilized along the τ-isomerization pathway and will dominate when the CI is reached. This effect is much more pronounced with OM2/MRCI when I constrain λ in a QM3 calculation, as discussed in the Appendix–A.4. I can conclude then that the QM2 pathway of the neutral CRO is a good approximation for the minimum energy pathway to the CI and that the CI region is reached by isomerization of the neutral CRO. For Dronpa, the three CC2 energy profiles obtained with different QM regions and protonation states are almost equal in shape for all τ values, but the pathway with anionic CRO is constantly higher in energy (from at least 0.05 eV to up to 0.12 eV). This is because there is no proton transfer along any of the paths (Figure 3.19b) and is consistent with the results obtained for the ESPT paths. The neutral CRO is energetically more stable and there is a sufficiently high barrier for PT to impeding the PT during the optimization and there is no correlation between isomerization (τ, ϕ) and PT (λ') (compare Figure 3.17).

Although my calculations show a very different picture of the ES potential energy surface for the two compared proteins, the predicted outcome of the ES dynamics is the same. In both systems the QM2 pathway of the neutral CRO is a good approximation of the minimum energy pathway to the CI and the isomerization is not supported by ESPT. But in the case of IrisFP the probability of ESPT is high, specially if it is excited when it has a positive τ value, and thus the probability of an anionic intermediate is also high, even if the isomerization will happen in the neutral form anyway. This is not the case in Dronpa where the probability of ESPT would be low in any case.This result is consistent with the recent experiments on Dronpa and IrisFP and does not agree with the theoretical study of Li et al. [49], who suggested that ESPT in Dronpa may occur close to the CI region, as in gas-phase calculations they find an ES minimum with a twisted geometry and a strong charge separation, where phenol side of the CRO (beyond the twisted bond) accumulates a positive charge of 0.9 e. They argue, this would lower the pK_a of the chromophore, i.e., support deprotonation.

Moreover, our results are consistent with the observation of previous theoretical studies of the chromophore in vacuo [68, 109] and in asFP595 [55] that the neutral CRO shows a preference for τ-isomerization (hula twist inside the protein), which promotes efficient trans-cis isomerization, whereas the anionic CRO prefers radiationless decay via ϕ-isomerization or is fluorescent when the latter motion is suppressed by the protein. IrisFP enhances this intrinsic feature of the CRO. In the OFF-state, the direct H-bond with negatively charged Glu144 stabilizes the phenol of the neutral CRO, which accumulates positive charge when approaching the CI via hula-twist. This effect is weaker in Dronpa because it has a negative charge in the imidazole ring side of the CRO (in contrast to IrisFP it has Glu211 deprotonated). As we will see in section 3.3.3 below, this charge weakens the Glu144 H-bond to CRO's phenol making it water mediated, increasing the isomerization barrier and contributing to make Dronpa's OFF to ON switching QY lower than IrisFP's one (7% and 15% respectively).

3.3.3 The Role of Glu212 and hydrogen-bonds to the chromophore

In this section I complete my comparison of IrisFP and Dronpa by analysing how the structural differences in the two proteins affect the ES potential energy surfaces and develop a rationale for the different direction of the τ-isomerization and the different ESPT behaviour. For this purpose, I use a second setup for Dronpa, named Y, which is also described in section 3.2. This setup is equal to setup Z but has Glu212 protonated (Figure 3.21), hence assumes the same protonation states as setup B for IrisFP (which again yields the same results as the A1 and A2 setups with their different position of the proton of His194). Although the configuration shown in Figure 3.21 is stable during MD simulations for at least 800 ps, I found a more stable configuration in some of the performed MD simulations, in which Glu212 forms an H-bond to the imidazole and Glu144 forms a direct H-bond to the phenolate, just like in IrisFP. This structure is 5 kcal/mol more stable and is shown in Figure 3.22.

In the following, I will test to what extend the properties of IrisFP are reproduced by model Y2. Additional calculations on the intermediate model Y1 are provided in the Appendix–A.6.

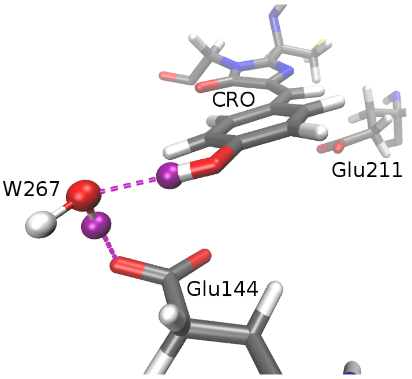

FIGURE 3.20: **Dronpa** (model Z). In purple the hydrogens and H-bonds involved in the PT. It is a water mediated H-bond. Glu211 is deprotonated.

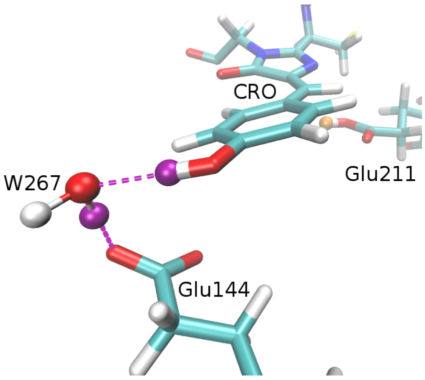

FIGURE 3.21: Dronpa model **Y1**. In purple the hydrogens and H-bonds involved in the PT. It is a water mediated H-bond. Glu211 is protonated (orange).

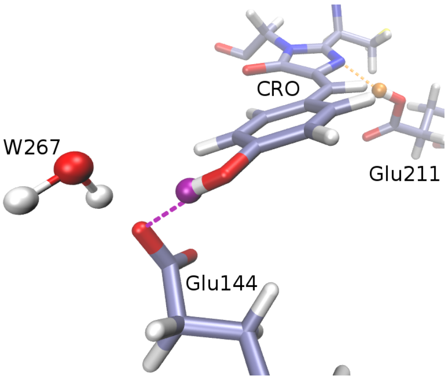

FIGURE 3.22: Dronpa model **Y2**. In purple the hydrogens and H-bonds involved in the PT. It is a direct H-bond. In orange the Glu211-CRO H-bond.

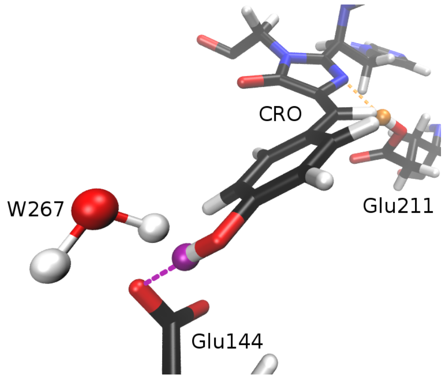

FIGURE 3.23: **IrisFP**. In purple the hydrogens and H-bonds involved in the PT. It is a direct H-bond. In orange the Glu211-CRO H-bond.

Comparison of the relaxed 2D scan. In Figure 3.24 I show the results of the 2D scan of the ES potential energy surface in the τ and ϕ dihedrals (2D projection) for IrisFP and the two Dronpa models Y2 (center) and model Z (right).

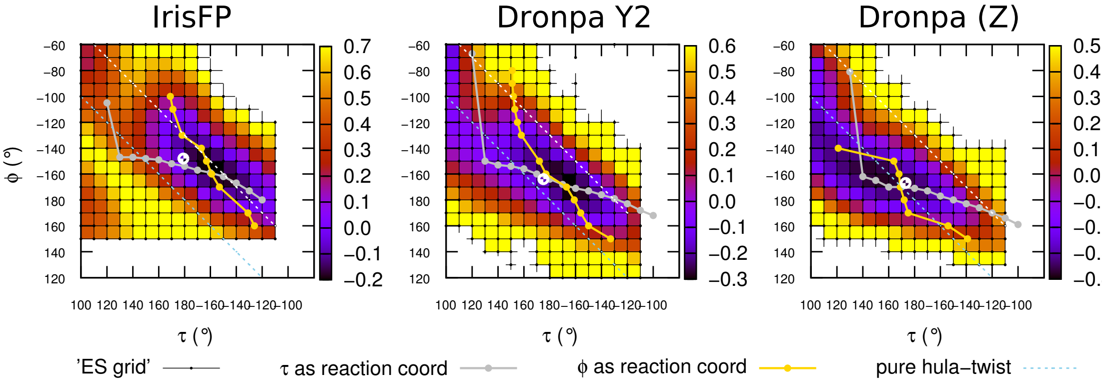

FIGURE 3.24: IrisFP, Dronpa Y1 and Dronpa (Z) relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The white dots are the FC points.

Comparing the three plots we can see how the change in protonation state and H-bonding from Dronpa model Z to model Y2 changes the potential energy surface. The minimum moves towards negative τ values and the barrier towards the CI at positive τ increases. This shows that the change in the protonation state and the resulting rearrangement of the H-bond network partially explain the differences in the potential energy surface of Dronpa and IrisFP. But even after this changes the preferred isomerization direction continues to be the opposite, indicating that there is an other important source of change.

As we can see inf Figure 3.2 there is one important structural difference between the two proteins apart from the H-bonds to the CRO we didn't discuss, the position of Met159. Looking at the structures of the τ-constrained isomerization pathways one can observe that Met159 hinders isomerization for positive τ but supports isomerization in the opposite direction by its steric interaction with the CRO phenol ring (Figures 3.26 and 3.25 respectively). In both Dronpa models this is not the case, Met159 is practically not affected by the isomerization in neither direction.

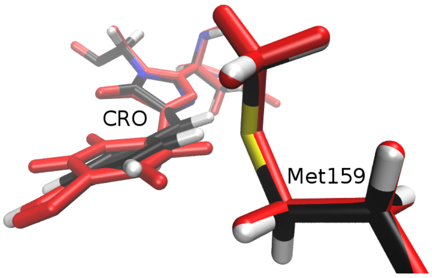

FIGURE 3.25: **IrisFP**. In red *tau*-constrained isomerization pathway's unfavoured CI point. FC structure in black for comparison.

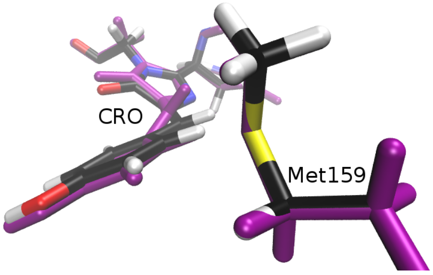

FIGURE 3.26: **IrisFP**. In purple *tau*-constrained isomerization pathway's preferred CI point. FC structure in black for comparison.

Single dihedral constrained pathways In Figure 3.27 I compare the τ- or ϕ-constrained isomerization pathways for IrisFP and the Dronpa setups Z and Y2, as calculated with OM2/MRCI.

These energy profiles clearly show that model Y2 reproduces the high isomerization barrier of IrisFP for positive τ, but there is no difference between the two Dronpa models in the other direction (for negative τ).

Correlation between isomerization and proton transfer. In Figure 3.28 I show the relaxed τ and ϕ dihedrals along the ESPT pathway of Dronpa model Y2 (center) and those of IrisFP and Dronpa model Z for comparison. The correlation between λ and the two dihedrals of IrisFP is reproduced by Dronpa model Y2. The hump in the region around λ=0, is caused by the water267 near the phenol of the CRO. In IrisFP, this water is H-bonded to the CRO even when the latter

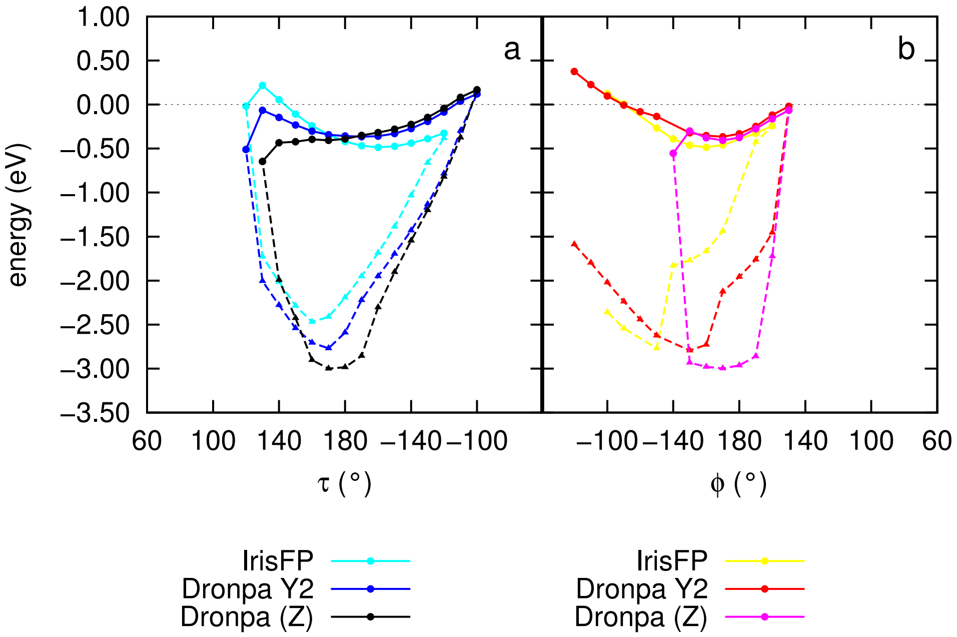

FIGURE 3.27: IrisFP and Dronpa models Z and Y2 relaxed (ES-optimized) energy profiles of the τ-constrained (a) and ϕ-constrained (b) isomerization pathway for the neutral CRO. Zero energy is set to the FC point.

is neutral (Figure 3.23), whereas in Dronpa Y2 it is initially H-bonded to Glu144 (Figure 3.22) but then forms a H-bond to the CRO when the latter becomes anionic for negative λ. The effect of water267 is also present in the ESPT pathway energy profile (Figure 3.29) making a kink similar to the one in Dronpa (Z) with direct H-bond ("144-D"). The resulting energy profile has a considerable barrier of 0,17 eV to the preferred deprotonation of the CRO. This barrier is absent in IrisFP.

In Figure 3.30 I show the τ-constrained isomerization pathways calculated with the extended QM regions, along with the relaxed λ, for Dronpa model Y2 (center). The data for IrisFP (left) and Dronpa model Z (right) are shown for comparison.

As expected from its PT pathways (Figure 3.28), when extending the QM region to QM3 for Y2 the ESPT shown by IrisFP is not reproduced, and the CRO stays neutral for all the τ range, kept by the 0.17 eV barrier. To further test the ESPT behaviour I start the calculations from the anionic CRO structure corresponding to the minimum at $\lambda \simeq 0.4$ Å, obtaining the magenta curve. This time the curve reproduces IrisFP ESPT behaviour stabilizing the neutral CRO

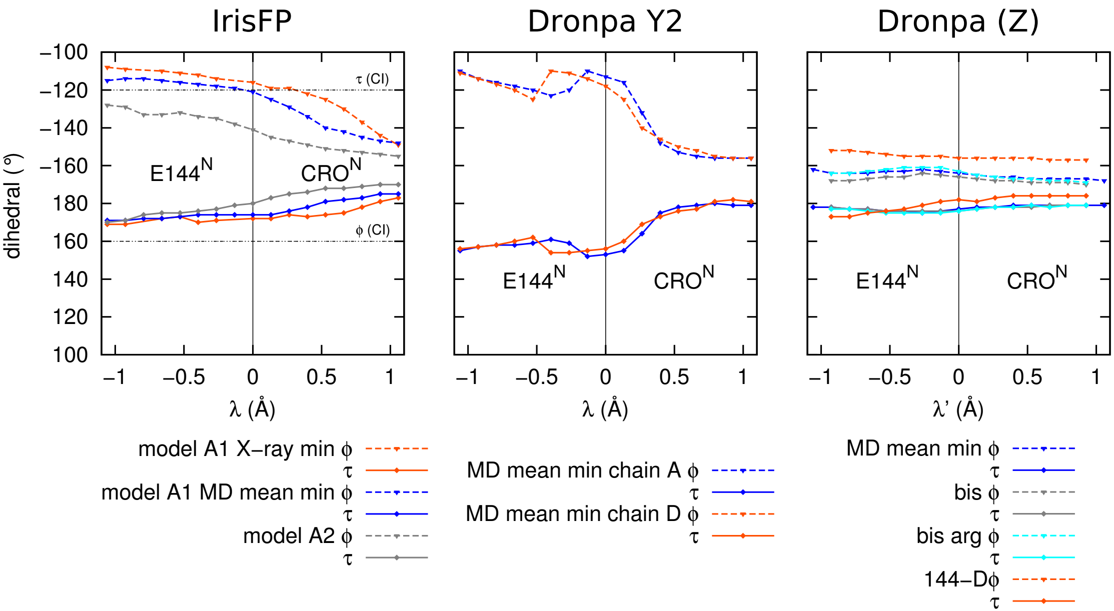

FIGURE 3.28: τ and ϕ dihedrals along the λ/λ'-constrained PT pathway for IrisFP, Dronpa (Z) and Drompa model Y2 (CC2 is used as the QM method). The upper and lower horizontal red lines mark the τ and ϕ values at the lower energy CI, respectively (For IrisFP).

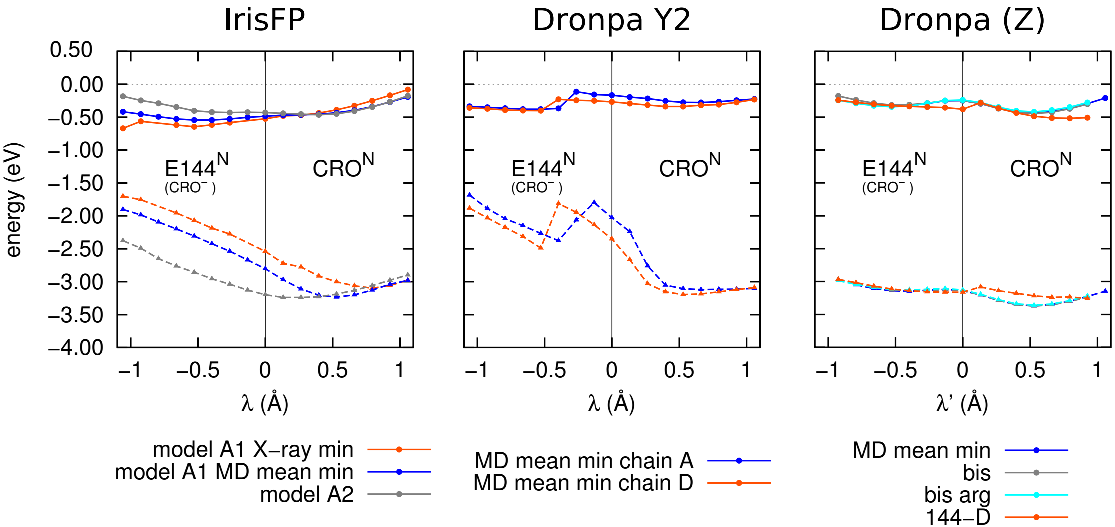

FIGURE 3.29: Energy profile along the λ/λ'-constrained PT pathway for IrisFP, Dronpa (Z) and Drompa model Y2 (CC2 is used as the QM method).

when approaching the negative τ CI. For the opposite isomerization direction, close to the CI, the H-bond between the phenolate and the protonated Glu144 breaks and is substituted by water267 (reflected in the jump from $\lambda \simeq -1$ to -2 Å). Therefore, I extended to QM region to QM4 to get a more balanced description of the two competing H-bonds to the Glu144, and to test how it influences the PT. I obtained the green curve. The water mediated H-bond highly stabilizes

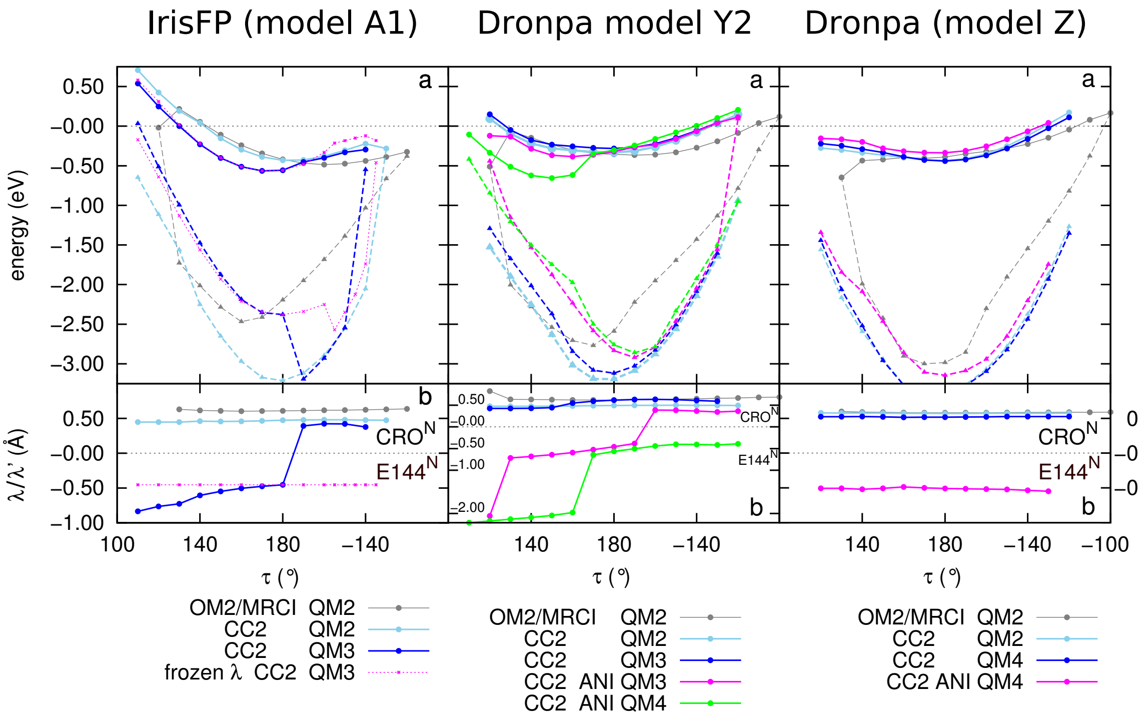

FIGURE 3.30: (a) Energy profile along the τ-constrained isomerization pathway with QM2 (light blue), with QM3 and neutral CRO (dark blue) and with QM3 or QM4 and anionic CRO (magenta or green respectively) using CC2 as the QM method. The OM2/MRCI profile for QM2 and neutral CRO is shown for comparison (gray). (b) The corresponding λ values. In Dronpa model Y2, λ values of -2 are reached because the Glu144-CRO H-bond changes from direct to water mediated. Zero energy is set to the FC point.

the anionic CRO for positive τ values, and even if the PT does not happen in the optimizations, the curve is higher than its neutral counterparts for negative τ values.

It is interesting to note that in contrast to the other two systems, Dronpa model Y2 describes a system that would isomerize at least partially to an anionic cis ground state. It is not clear though if this intermediate would be stable and relax to the fluorescent ON-state, but back protonation of the CRO is unlikely as the isomerization places the phenolate ring away from Glu211, nearby the deprotonated His193.The excited-state life time would be larger due to the higher barrier but emission might be redshifted beyond the range covered by time-resolved experiments.An anionic CRO photoproduct would be nearer to the fluorescent state with no need of ground state PT, making the whole process faster as the ground state PT is the bottleneck of the process.As in the case with OM2/MRCI calculations with a smaler QM region, the negative τ isomerization direction is almost unaffected by the changes between Dronpa models Z and Y2.

Finally, all the differences observed between Dronpa models Z and Y2 are already widely reproduced with the intermediate model Y1 (see Appendix–A.6). This shows that the main effect of changing the protonation state of Glu212 is electrostatic rather than structural. Actually, the neutral CRO potential energy surface is more similar to IrisFP for Y1 and the ESPT behabiour for Y2.

3.4 Conclusions

I have theoretically studied the OFF-state of both IrisFP and Dronpa, to search for the low-energy pathways on the first electronically excited state potential energy surface that provide channels for the characteristic radiationless decay.

The OFF-state S_1 surface of IrisFP features a sloped conical intersection, which is energetically below the FC point and is reached from the global minimum by a hula-twist-like isomerization. We find that ESPT is essentially barrierless and that the proton is shared between the CRO and Glu144, i.e., both protonation states can be populated during the short excited-state lifetime. We find, however, that the isomerization pathway that leads to the CI clearly favours a neutral CRO protonation state. Therefore, we propose a new model for the OFF to ON switching mechanism where ESPT appears as a competitive process rather than an intermediate, with the neutral cis CRO as the major GS product. The ON-state would then be established by another PT step in the electronic GS, as observed in the FTIR studies [106].

In the case of Dronpa there is a barrierless pathway to a CI, which is energetically below the FC point and is also reached from the global minimum by a hula-twist-like isomerization, but in the opposite direction of isomerization. The picture of the ESPT is completely different in this case. The neutral CRO is the most stable protonation state all along the isomerization pathway. Therefore, no kinetic isotope effect is expected for Dronpa, in contrast to the conclusions drawn by Fron et al. in their ultrafast transient absorption spectroscopy study [46]. Therefore, we theoretically support the model proposed by Warren et al. and Yadav et al. in their experiments on Dronpa, where a fast photoisomerization occurs without PT, followed by GS deprotonation [51, 57].

Considering the structural similarity of the binding site among negative RSFPs and the consistency of our results on IrisFP and Dronpa, the general OFF to ON switching mechanism for IrisFP and Dronpa, composed of a fast photoisomerization followed by a GS PT, might be transferable within this class of proteins. This applies despite the differences that I found in the microscopic process that leads from the same ES intermediate to the same GS species. More specifically, the ability of these proteins to stabilize the neutral CRO in the excited state might be essential for their high OFF to ON switching QY. This stability, however, is achieved in very different ways in the two investigated proteins. While in Dronpa the negatively charged Glu212 is the essential factor, in IrisFP it is a combination of the Glu144 interaction with the CRO and the sterical effect of Met159, which supports the isomerization in the opposite direction as in Dronpa.

One imaginable strategy to obtain a faster OFF to ON switching would be to avoid the slow GS PT by mutations that allow a concerted isomerization and ESPT.

The Glu144-CRO interaction, present in the OFF-state, is missing in the ON-state, where the charge of Glu212 is neutralized by the protonation of His194. Therefore, Glu212 does not change the pK_{a} of the CRO in the ON-state and a protonation in the excited state seems unlikely. pH-dependent measurements on the ON-state of mIrisGFP have shown that the QY for ON to OFF switching is increased by orders of magnitude when exciting the neutral (non-fluorescent) cis CRO [110]. It is reduced, however, when the population of the neutral CRO is increased at lower pH. The ON to OFF switching mechanism might therefore involve structural fluctuations in the GS. Its details remain unclear and will be addressed in the next chapter.

Chapter 4

Photoswitchable Fluorescent Proteins: ON-state

4.1 Introduction

The importance of reversibly photoswitchable fluorescent proteins (RSFPs) was described in the introduction, and Dronpa and IrisFP where already presented in Chapter 3, where I studied the OFF-state of these proteins.

An important difference between the ON- and the OFF-state for the study of the photoswitching is that the switch from the ON- to the OFF-state is a minor event competing with fluoresce, which is the dominant decay channel. Due to the scarcity of the event, there is little experimental information available about it. Lukacs et al. preformed time resolved spectroscopy experiments on Dronpa2 [72], a single-residue mutant of Dronpa. Dronpa2 has a much faster photoswitching time [48], which allowed them to obtain data for the ON-state, although the main focus of their article was the OFF-state. Kaucikas et al. also studied the time resolved spectrum of Dronpa2 for the ON-state [74], and they conclude that ON to OFF switching involves a fast excited state isomerization followed by ground state protonation of the chromophore on a ns to ms time scale. They found indications that Arg66 plays an important role.

Very recently, Morozov et al. [73] performed a theoretical study on the dynamics of the ON-state of Dronpa2, which was based on the result of Kaucikas et al. that excited-state isomerization is the first step. In their simulation of 12 excited-state MD trajectories they found that the isomerization only occurs when two of the three H-bonds to the CRO phenolate are broken before excitation. One of the broken H-bonds was to Ser142, see Figure 4.1. This means that, as they point out, photoswitching is triggered by structural heterogeneity in the ground-state dynamics of the protein. This scheme can not explain, however, why Dronpa2 is a faster switcher than Dronpa, as Dronpa has not three but only two H-bonds to the CRO phenolate (Figure 4.2). Therefore, it should be more probable for the phenolate to lose a single H-bond in Dronpa than two simultaneously in Dronpa2, and Dronpa should have the higher switching rate. Nevertheless it is important to note that Dronpa2 was obtained by a single Met159Thr mutation by Stiel et al. [48]. They speculated that this mutation reduces the steric hindrance during photo-isomerization and thereby enhances ON to OFF switching. Kaucickas et al. obtained the crystal structure of Dronpa2 and compared it to Dronpa finding an increase in the cavity volume for the chromophore due to the mutation further supporting the hypothesys. This view appears to be consistent when considering that IrisFP is a Phe173Ser mutant of the EosFP fluorescent protein [59]. This reported mutation indirectly affects the position of the equivalent Met159 side

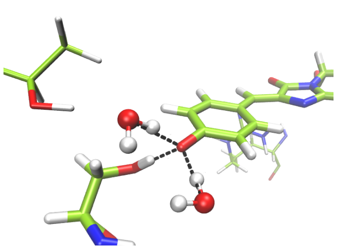

FIGURE 4.1: **Dronpa2.** In black the H-bonds to the CRO phenol. Residue 159 is a Threonine.

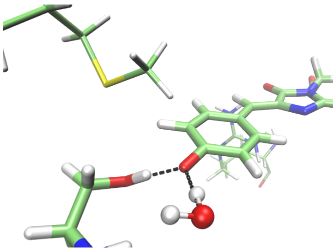

FIGURE 4.2: **Dronpa**. In black the H-bonds to the CRO phenol. Residue 159 is a Metionine, in a different position than IrisFP.

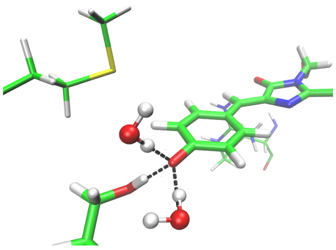

FIGURE 4.3: **IrisFP**. In black the H-bonds to the CRO phenol. Residue 159 is a Metionine, in a different position than Dronpa.

chain, which subsequently allows cis–trans photoisomerization not seen in the wild type [59].

In this chapter I will study the properties of the potential energy surface of Dronpa2, Dronpa, and IrisFP, to theoretically test the two existing hypothesis about the main factors affecting the ON to OFF switching quantum yield (QY), namely the steric hindrance of the isomerization by residue159, and the number of H-bonds to the CRO phenolate, looking also for the mechanistic details. Appart from residue 159, the three proteins have a big structural similarity, but their QY for the ON to OFF switching differ each by one order of magnitude: Dronpa2 is fastest with a $QY_{off-switch} = 4.7 \cdot 10^{-2}$ [111] followed by IrisFP with a $QY_{off-switch} = 3.2 \cdot 10^{-3}$ [59, 66] and finally Dronpa with $QY_{off-switch} = 3.0 \cdot 10^{-4}$ [40, 66].

4.2 Methods

Setup. The coordinates of the heavy atoms of IrisFP, Dronpa and Dronpa2 were obtained from the X-ray structures of Adam et al. [59], Wilmann et al. [47], and Kaucikas et al. [71] (PDB codes 2VVH, 2IE2, and 4UTS), respectively. Protonation states and the position of H-atoms where determined following the procedure described in Chapter 2. Default protonation states were assumed for all residues except His194, which was protonated in all three proteins. Thus, in the ON-state, all three proteins have the same protonation states.

QM Region. For the ON-state I use only two different QM regions, QM1 and QM2 as defined in Chapter 3 for the OFF-state (Figure 3.3 a and b). For preparatory ground-state calculations I use QM1, which includes the three residues forming the CRO (His62, Tyr63, and Gly64; Figure 3.3a, thick sticks) and two backbone atoms of the neighbor residues Phe61 (C and O) and Asn65 (N and HN) (Figure 3.3a, spheres). The excited-state (ES) calculations where performed with QM2, which only includes the conjugated part of the CRO.

Pathway calculations. The active orbital space was optimized for each QM region and, by chance, is the same for QM1 and QM2. It includes 10 occupied and 10 virtual orbitals (Appendix–A.5), which were selected as is explained in Chapter 2.

All the energy profiles presented in this chapter are obtained by constrained pathway optimizations along a reaction coordinate, which is either the τ or ϕ dihedral for the isomerization pathways (see Figure 3.4). During the optimization, atoms further than 10 Å from any QM atom were kept fixed at their initial positions and all other atoms were free of constraints (Figure 3.3). As the excited-state dynamics happens on a fs to ps time scale, long-range relaxation is not expected to be relevant. For the calculated potential energy surfaces, I scan both τ and ϕ dihedrals while maintaining the hydrogen atom of the bridge connecting the phenolate and the imidazole ring in plane with the three carbon atoms of the bridge. All the redundant dihedrals including the atoms marked in black in Figure 3.4 are also constrained to preserve co-planarity.

4.3 Methodological Test

First, in order to test the quality of our approach in predicting the energetics of the ES potential energy surface, we will present a model of the IrisFP fluorescent ON-state (Figure 4.3), which is experimentally well characterized. This gives support to our results on the excited-state dynamics of the OFF-state presented in Chapter 3 and the ON-state calculations we are going to do in this chapter.

IrisFP ON-state: Isomerization of the anionic CRO. Radiationless decay in RSFP's (and other FPs) involves τ-isomerization, ϕ-isomerization, or a combination of both (hula-twist). Therefore, we characterize the ES potential energy surface in terms of these coordinates.

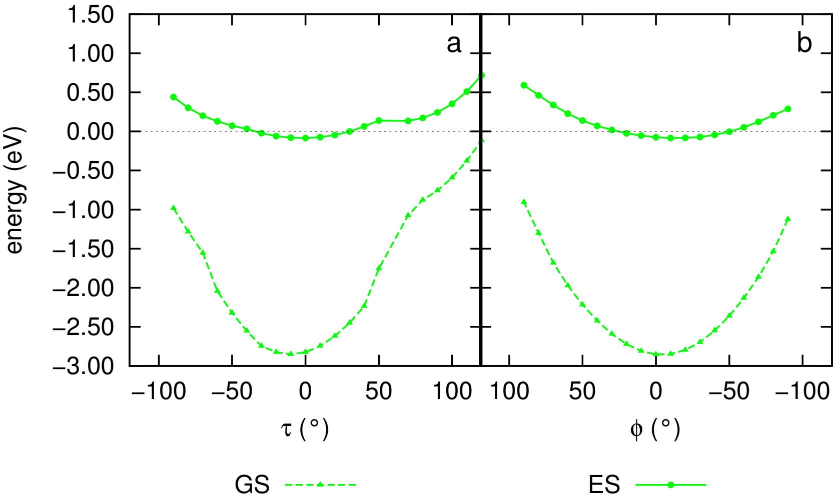

FIGURE 4.4: ON-state: Relaxed (ES-optimized) energy profile along the τ-constrained (a) and ϕ-constrained (b) isomerization pathway. The dashed line denotes the FC point.

The relaxed ES energy profile in Figure 4.4 shows a minimum $\sim$ 0.1 eV below the FC point, both for τ and ϕ as the reaction coordinate (Figure 4.4a and b, respectively). The energy rises at least 0.3 eV (Figure 4.4*b* −90°) above the FC point without reaching a conical intersection (CI). This shows that at this level of theory we recover the fluorescent character of the ON-state of the protein. The Stokes shift associated with this minimum (0.21 eV) is in good agreement with the experimental value of 0.14 eV, considering the general difficulty to theoretically predict Stokes shifts. For example, Xin Li et al. obtained a 0.29 eV Stokes shift for Dronpa with CASPT2/AMBER (experimental value: 0.07 eV). [49] For the B-form of wild type GFP, we calculated 0.11 eV with SORCI/CHARMM (experimental value: 0.18 eV) [87].

4.4 Results and Discussion

To search for a possible isomerization pathway of the photoexcited ON-state, I calculated a relaxed 2D scan of the ES potential energy surface in the τ and ϕ dihedrals. Figures 4.5, 4.7, and 4.9, for IrisFP, Dronpa2, and Dronpa, respectively, show the 2D projection of the resulting potential energy surface (color coded by the energy in eV). Figures 4.6, 4.8, and 4.10 show the energy difference between

GS and ES. Values close to zero indicate CI regions, where efficient decay to the GS is possible.

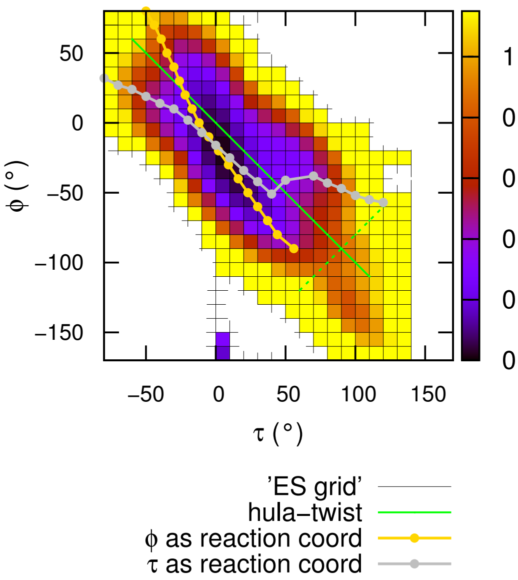

FIGURE 4.5: **IrisFP** relaxed 2D (τ, ϕ) scan of the ES potential energy surface (2D projection).

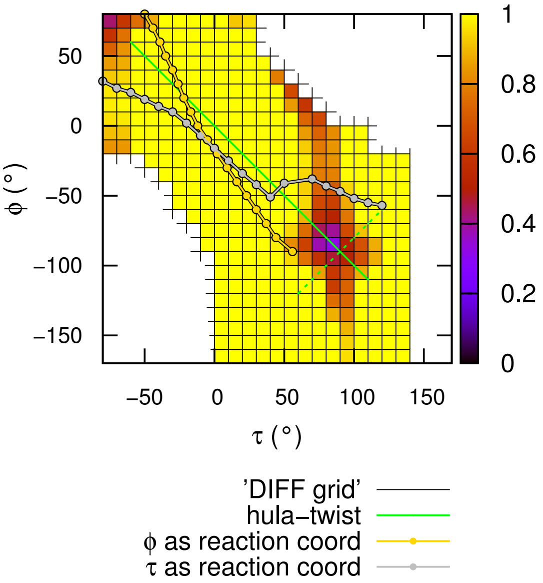

FIGURE 4.6: **IrisFP** ES-GS energy difference of the relaxed 2D (τ, ϕ) scan of the ES potential energy surface (2D projection).

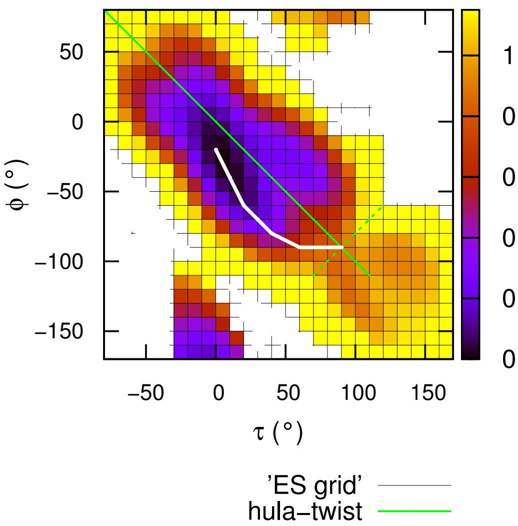

FIGURE 4.7: **Dronpa2** relaxed 2D (τ, ϕ) scan of the ES potential energy surface (2D projection). White line: manually optimized minimum energy pathway to the CI.

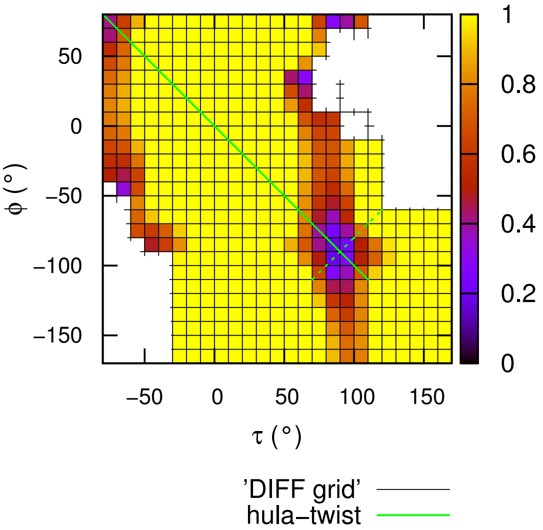

FIGURE 4.8: **Dronpa2** ES-GS energy difference of the relaxed 2D (τ, ϕ) scan of the ES potential energy surface (2D projection).

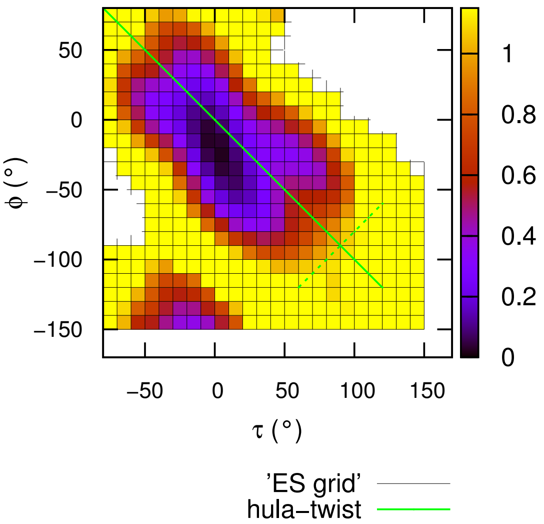

FIGURE 4.9: **Dronpa** relaxed 2D (τ, ϕ) scan of the ES potential energy surface (2D projection).

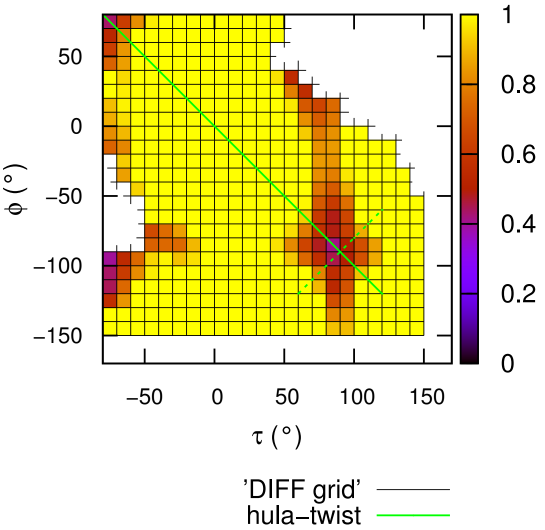

FIGURE 4.10: **Dronpa** ES-GS energy difference of the relaxed 2D (τ, ϕ) scan of the ES potential energy surface (2D projection).

For the three proteins we get a very similar qualitative picture. A butterfly shaped diagonal valley with the symmetry of the butterfly body along the hula-twist direction (green continuous line). Following the hula-twist coordinate in the direction of positive τ, we reach a CI region around the point $(\tau,\phi)=(90°,-90°)$, (marked with a crossing dashed line). The energy at this point is less than 1 eV in all three proteins. For the negative τ direction there is a similar CI region, but it is much higher in energy and I will therefore not discuss it further.

Looking at the ES-GS energy difference plots, we can see that the inverse minimal energy gap or the extend of the CI region of the three systems correlates with the order of their photoswitching QYs. Moreover, if we mark the position of the CI point in the plot of the ES energy (left side), it is evident that the energy in Dronpa is higher than in the other two proteins, consistent with the fact that it has the lowest switching QY.

The ES pathways optimized with a single constraint on τ or ϕ (see section 4.3) are shown as gray and yellow curves, respectively, in Figures 4.5 and 4.6. Both of these pathways miss the CI region and are therefore inadequate as reaction coordinate. I therefore use the ideal hula-twist coordinate ($\phi = -\tau$) to examine the energies around the CI point. A hula-twist coordinate is adequate because its trajectory passes nearby the minimum and crosses the lowest energy CI. This CI region is formed by a hill-shaped maximum of the corresponding GS surface. The hula-twist coordinate passes approximately through its cusp, where the energy gap is the smallest. A manually optimized minimum energy pathway for the potential energy surface of Dronpa2 is illustrated by a white line in Figure 4.7 for comparison.Figures 4.11 and 4.12 show the energy profile along the hula-twist coordinate, as extracted from the relaxed 2D scan.

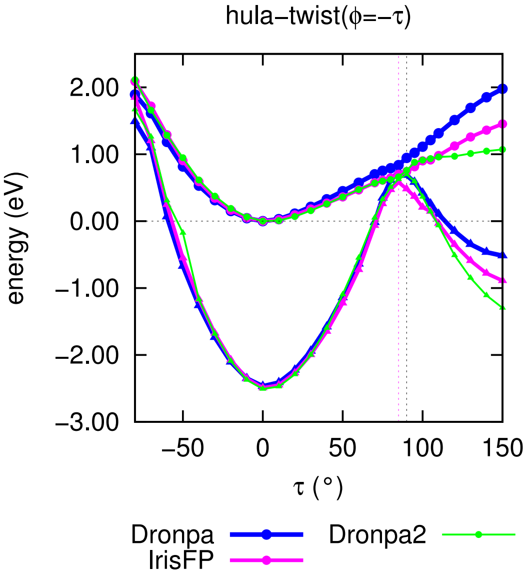

FIGURE 4.11: Energy profile along the hula-twist coordinate.

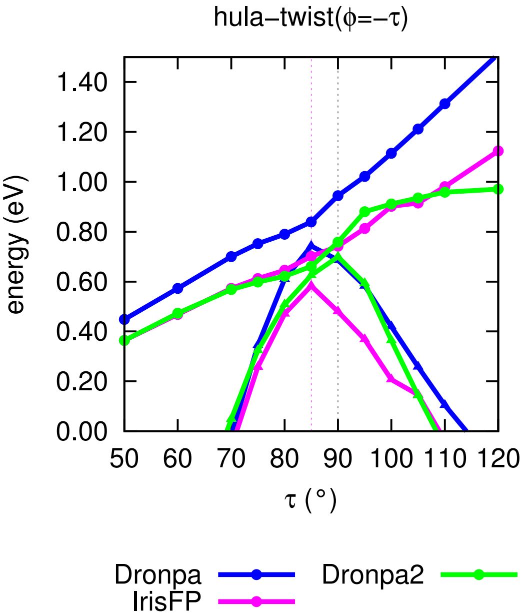

FIGURE 4.12: Close up of Figure 4.11.

The first thing we can observe in Figure 4.11 is that the gradient ($dE/d\tau$) of both the GS and ES curves after the CI region (coming from the minimum) inversely correlates with the photoswitching QY, i.e., the gradient is smallest for Dronpa2, the fastest photoswitcher, and largest for Dronpa, the slowest one.In a dynamical picture (considering, e.g., a surface hopping trajectory) this means that independent from the exact point where the decay to the GS occurs, the probability to successfully complete the isomerization (and hence the QY) is the larger the lower the gradient in this direction.

Dronpa	IrisFP	Dronpa2
0.85	0.70	0.66

TABLE 4.1: CI point Energies from the minimum (eV)

Dronpa	IrisFP	Dronpa2
0.67	0.48	0.28

TABLE 4.2: Fluorescence quantum yields

Before the CI region, the profiles of IrisFP and Dronpa2 are remarkably similar. In Dronpa, the energy rises slightly faster as we approach the CI from the minimum and at a closer look, the energies of the crossing points are slightly different. Comparing the energies at the nearest point to the CI (τ=85°, ϕ=-85°; see Figure 4.12), Dronpa has the highest energy at the CI point, followed by IrisFP, then Dronpa2 (see Table4.1). As the energy of the CI point can be considered as the effective barrier height for radiationless decay, it should inversely correlate with the fluorescence QY. Indeed, this is qualitatively the case, as shown in Table4.2.

Looking at the structures along the hula-twist coordinate (Figure 4.13), we can confirm and give details about the explanation why Dronpa has the highest and Dronpa2 the lowest energy slope towards the CI. In Dronpa2, the mutation of Met159 into the smaller threonine side chain leaves a space that is occupied by the CRO phenolate during the isomerization. In Dronpa, the Met159 side chain pushes the CRO away from this region and towards Ser142 and Water2, which maintain their H-bonds to the CRO. In IrisFP, the Met159 side chain initially gives more space for the isomerization than in Dronpa because it is rotated around the C_β–C_γ bond. Beyond the CI (for $\tau > 95°$), however, it pushes against the CRO as well (Figure 4.13 c and d), which explains why the slope in IrisFP is higher than in Dronpa2 here (Figure 4.12).

We can see that in Dronpa2 and IrisFP Water2 is hindering the isomerization, as its H-bond to the the phenolate is broken after the CI point. This event is visible as a kink in the energy profile (Figure 4.12) at $\tau = 95°$ and 105° respectively.

In their theoretical study, Morozov et al. [73] find that breaking two of the

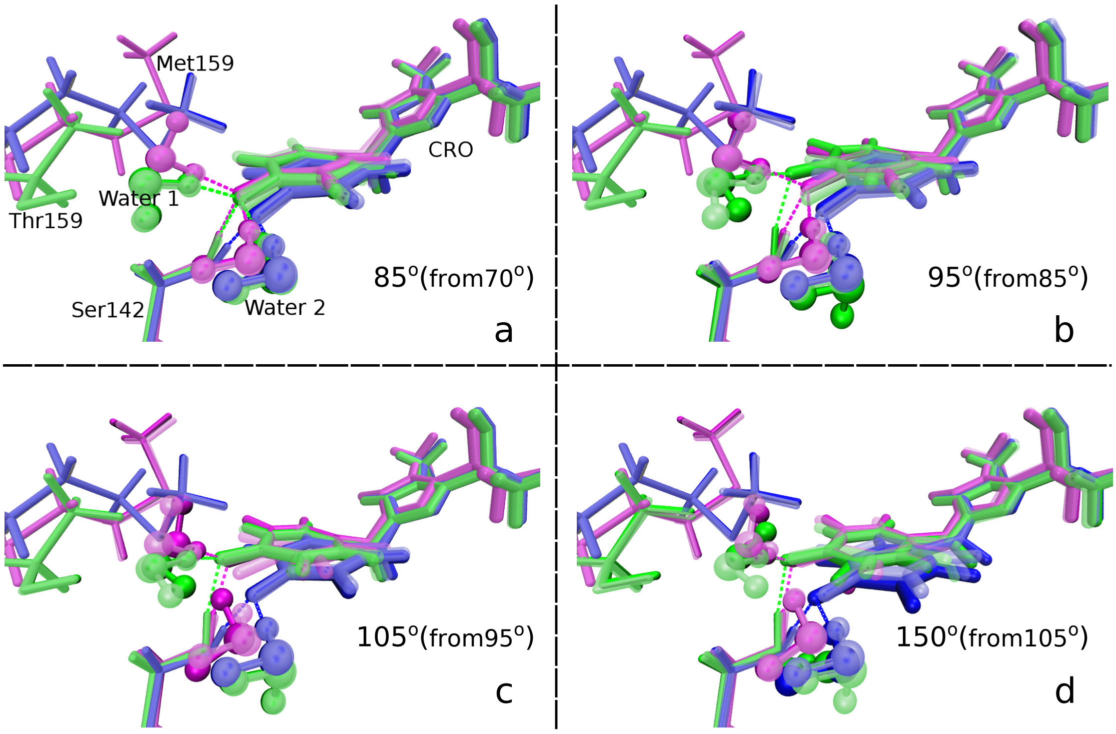

FIGURE 4.13: Dronpa (blue), IrisFP (magenta), and Dronpa2 (green). **a)** hula-twist from $\tau = 70°$ (translucent) to 85° (opaque, the CI point). No change apart from the dihedrals. **b)** hula-twist from $\tau = 85°$ (translucent) to 95° (opaque). The CRO phenolate rearranges in Dronpa2 (Water1 follows its motion) and loses the H-bond to Water2. **c)** hula-twist from $\tau = 95°$ (translucent) to 105° (opaque). In IrisFP the CRO starts to push against Met159 and loses the H-bond to Water2. **d)** hula-twist from $\tau = 105°$ (translucent) to 150° (opaque). In IrisFP the CRO pushes Met159 further. Water 2 is W278/W17/W320 in the PDB file of Dronpa/IrisFP/Dronpa2. Water 1 is W188/W319 in the PDB file of IrisFP/Dronpa2.

three H-bonds in Dronpa2 enables ultrafast isomerization. They argue that this is due to the electrostatic effect of these H-bonds and hinders the charge separation that occurs in the CRO near to the CI. If this statement is correct, the structural fluctuations that accompany the dislocation of the waters and isomerization of Ser142 should not be important and the same effect can be achieved by simply removing the waters and re-optimizing the pathway. The resulting energy profile is shown in Figures 4.14–4.16. The effect is different for each protein and I will analyse it separately for each case.

In Dronpa2 (Figure 4.14) the result is in agreement with the finding of Morozov et al. [73]. Removing one of the waters either increases the barrier (for water

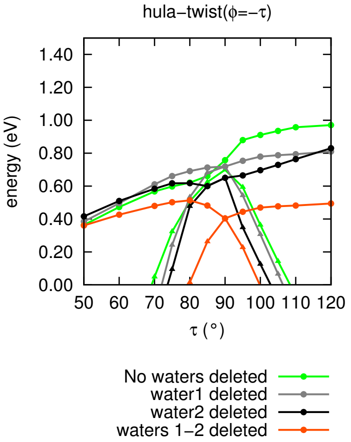

FIGURE 4.14: **Dronpa2** energy profiles along the hula-twist coordinate for each possible combination of waters 1 and 2 deleted.

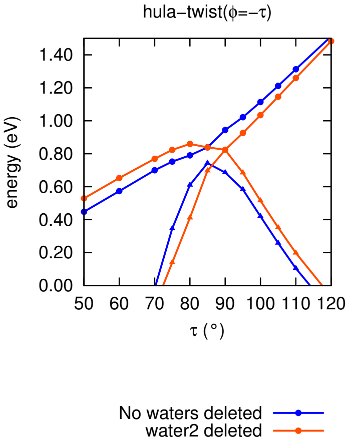

FIGURE 4.15: **Dronpa** energy profiles along the hula-twist coordinate. It has only water 2 (see Figure 4.13a). Curves with and without it. An other water H-bond to water 2 (not shown) was also deleted not the replace water 2 effect form a bit further.

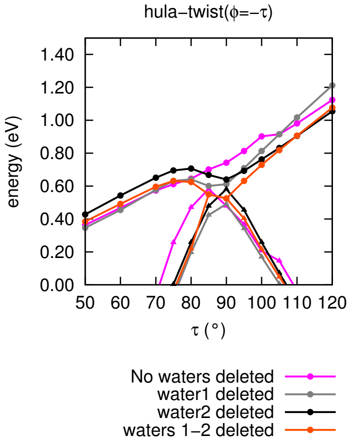

FIGURE 4.16: **IrisFP** energy profiles along the hula-twist coordinate for each possible combination of waters 1 and 2 deleted.

1) or leaves it unchanged (0.04 eV lower for water 2). Removing both lowers the barrier significantly (0.15 eV). It also changes the type of CI from a sloped one to a downhill one, separated from the barrier. Morozov et al. compute the potential energy profile extracting a generalized reaction coordinate for the hula-twist from one of the isomerization trajectories. They select one trajectory where the Ser142 H-bond is broken to extract the reaction coordinate (their figure 1b [73], Ser142 isomerizes away form the CRO; actually it is not clear from the text if this is the case for al the isomerizing trajectories). Then they place a frozen CRO with structures along the extracted reaction coordinate into the protein and optimize with the charges of the Ser142 switched off and one water deleted to ensure a single H-bond electrostatics.Our “H-bonds breaking analysis”, in contrast, always maintains the Ser142 H-bond, as it is permanent in our MDs (various 600ps trajectories orders of magnitude shorter than theirs of 100ns). Even if the shape of the curves they get are very different (figures S15 and S16 of their supplementary

material [73]), which might be due to the different Ser142 configuration, the energy barriers they get for the higher level of theory are not far from ours, 0.57 eV with the H-bonds (ours is 0.66 eV) and 0.1 eV lower, 0.47 eV, with two H-bonds broken (ours 0.15 eV lower, 0.54 eV).

For Dronpa (Firgure 4.15) we have a very different picture. There is one H-bond less as the position of Met159 takes away one of the two waters H-bond to the CRO (Figure 4.13a), and if we delete it, the energy gap merely changes. We can conclude that in this case, the collision with Met159 keeps the CRO in a position during the isomerization in which the effect of the missing H-bond is not relevant.

In IrisFP (Figure 4.16), the number of H-bonds affects the isomerization but not as much as in Dronpa2. Removing water 2 shows little effect, the barrier stays the same, but the CI changes from sloped to downhill. Taking out the other water or both, lowers the barrier by 0.06 eV, similar to the result when removing water 2 in Dronpa2 (0.04 eV). This shows that Met159 does not allow the CRO to take advantage of any freedom gained by a lesser number of H-bonds.

My results show that the effect is not purely electrostatic, but depends on the position (connectivity) of the deleted water and on the size and exact position of residue 159, which can substitute the stabilizing effect of the water H-bonds by mere steric interactions. Structural considerations can be more important than modifying the charge separation of the twisted CRO.

In Figure 4.17 I compare the energy profiles with the lowest barrier for each protein, i.e., considering the H-bond situation that is most favorable for isomerization. The differences in barrier height are 0.12 eV between Dronpa2 and IrisFP and 0.21 eV between IrisFP and Dronpa and reflect more clearly the trend in the fluorescence QYs of the three proteins than in Figure 4.12. We therefore support the conclusion by Morozov et al. that structural heterogeneity affects the OFF to ON switching dynamics, but the effect strongly depends on the structural details of the binding pocket and is in fact smallest for Dronpa, where the Met159 side chain limits the role of the H-bond network.

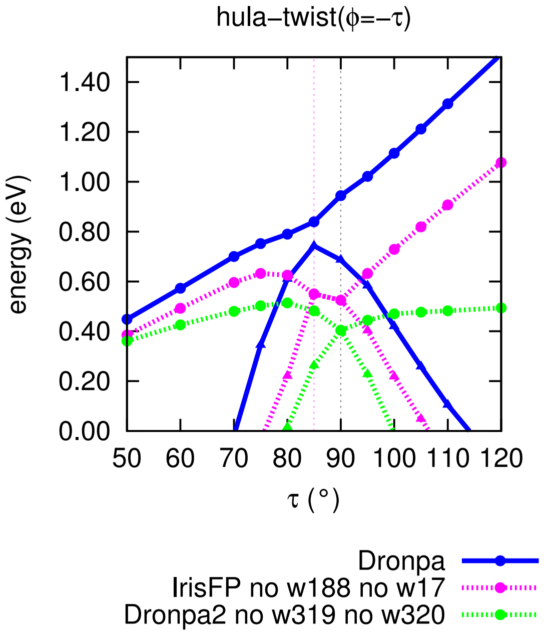

FIGURE 4.17: Hula-twist of normal Dronpa (blue), and IrisFP (magenta) and Dronpa2 (green) with waters 1-2 deleted.

4.5 Conclusions

Exploring the potential energy surface of the first excited state, I found that an ideal hula-twist is the most representative reaction coordinate for the description of the ON-state photoisomerization of Dronpa, Dronpa2, and IrisFP. Despite the coplanar geometry of the CRO in the GS, there is a clearly preferred direction of isomerization towards positive τ values in all three proteins (Figures 4.5, 4.7 and 4.9).

I confirmed the hypothesis that residue 159 is the key component controlling the ON to OFF photoswitching speed and QY. It is the main reason why Dronpa is the slowest and IrisFP is slower than Dronpa2. In Dronpa, the CRO phenolate and Met159 are in direct contact, in IrisFP they collide as the CRO moves further to complete the isomerization after the CI point is reached. In Dronpa2, the CRO is less constrained by the binding pocket along the isomerization and shows how the CRO relaxes when residue 159 does not interfere (Figure 4.13). I found that the minimum energy pathway for ON to OFF switching is very close to a space-conserving hula-twist isomerization. Therefore, it is possible that the space created by the Met159Thr mutation, which transforms Dronpa into Dronpa2, is larger than required to support the isomerization and therefore unnecessarily lowers the fluorescence QY. Other combined mutations around the CRO resulting in a better trade off between fluorescence and photoswitching yield may be possible. The residue combination in the CRO pocket in IrisFP would be an example, although Met159 in IrisFP still affects the isomerization after the CI point. Therefore there is still room for improvement in IrisFP.

I also confirmed that the H-bonds to the CRO phenolate can play a role for the photoisomerization QY, as proposed by Morozov et al. for Dronpa [73], but comparing three different proteins, my conclusions differ in several aspects. First, I find that in Dronpa2, where H-bonds have the largest effect, the position and connectivity of the of the waters, i.e., the structural function of the waters, are more important than their mere electrostatic effect. Second, I find that the effect in Dronpa is negligible (and reduced in IrisFP), as steric interactions with Met159 are dominant and prevent the displacement of the phenolate that the waters would hinder. Third, the energetic effect of the broken H-bonds has to be seen in context of the low population of these configurational states, which raises their free energy and therefore their contribution to the QY. Morozov et al., assume the H-bond between the CRO and Ser142 is broken, which might be a rather rare situation and a reliable prediction of the relevance of this configuration is not possible on the grounds of MM calculations, as the stability of this H-bonds depends critically on the MM parameters.

Outlook

The results in this thesis open a promising perspective for further improvements of reversibly photoswitchable fluorescent proteins (RSFPs). Comparison with further RSFPs with slightly different chromophore pocket like mTFP0.7, along with excited state molecular dynamics calculations, could provide further insights into the photoisomerization process, providing more clues for an optimal chromophore environment to get a better trade-off between photoswitching and fluorescence quantum yield. Further electrostatic analysis could provide information about how a concerted isomerization plus protonation state change could be achieved. This would probably increase the photoswitching speed as the protonation state change is orders of magnitude slower in the ground state and is needed to complete the ON/OFF photoswitching.

Appendix A

Photoswitchable Fluorescent Proteins: OFF-state

A.1 Protonation State Selection

Standard protonation states were assumed for all titratable amino acids except for those in the CRO vicinity: Arg66, Arg91, Glu144, His194 and Glu212. Protonation states where either chosen according to pK$_a$'s calculated with PROPKA [101] or, in unclear cases, by testing all plausible combinations. We selected the best protonation and structure setups through different filters. The first one requires the RMSD of the optimized geometry from the X-ray structure to be lower than 0.6 Å for all residues that are not solvent exposed. The second filter tests the structural stability during MD simulations. The optimized structures were heated to 300K within 20 ps, then propagated for 800 ps using the Nose-Hoover thermostat [102, 103], both with a time step of 1 fs.

For the ON-state setups, we chose chain A from the four chains in the 2VVH X-ray structure, as all 4 of them are very similar. Arg91, His194, and Glu212 have pK$_a$ values close to the pH in the X-ray experiment (8.4). For the OFF-state, we verified that Arg91 must be protonated (see below) and assumed

Table A.1: **ON-state** protonation models

model	His194	Glu212
1	prot.	deprot.
2	deprot. at ND1	deprot.
3	deprot. at NE2	deprot.
4	deprot. at ND1	prot.

Table A.2: **OFF-state** protonation models

model	His194	Glu212
1	deprot. at ND1	deprot.
2	deprot. at NE2	deprot.
3	deprot. at ND1	prot.
4	deprot. at NE2	prot.

the same for the ON-state, where we obtain a similar pK_a for this residue. For the ON-state we selected the four protonation state models shown in Table A.1. Model 1 yielded the best agreement with the X-ray structure and was stable during the MD simulation. In particular, it was the only one that maintained stable HB's to Glu144 and Glu122, consistent with the X-ray structure.

For the OFF-state it was more difficult to find a satisfying setup. His194 shows a high mobility and has different positions in each of the 4 different chains in the 2VVI X-ray structure. The X-ray structure shows a superposition of two rotamers of Ser142, which is very close to the CRO. According to pK_a calculations, Arg66 is clearly protonated and Glu144 deprotonated. The possibility of a deprotonated Arg91 was discarded in a first round working just with chain A. After that, a total of 48 different models were tested: four chains (a-c) × four protonation state combinations (Table A.2) × two Ser142 rotamer positions and some unsuccessful variants with added waters. Only three models yielded satisfactory results (A1, A2, and B). As described in the main article, A1 and A2 have the protonation state 4 but A1 is derived from chain A and A2 from chain D. Model B has protonation state 3 and originates from chain B.

A.2 Results with IrisFP model B

We obtain qualitatively the same results with model B (Figures A.1, A.2, A.3, A.4, A.5, and A.6) as with model A1 (Figures 8, 9, 10, 11, 12, and 13 in the article, respectively).

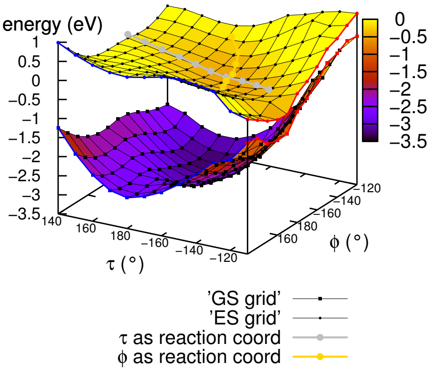

FIGURE A.1: OFF-state: Relaxed (ES-optimized) 2D (τ, ϕ) scan of the ground- and ES potential energy surfaces for the neutral CRO.

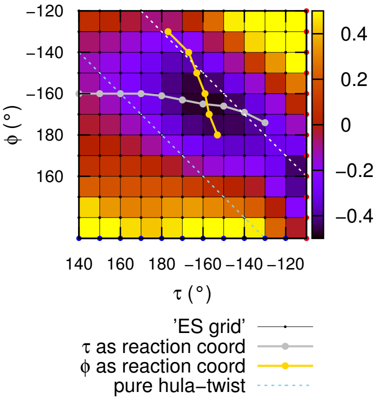

FIGURE A.2: OFF-state: Relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection).

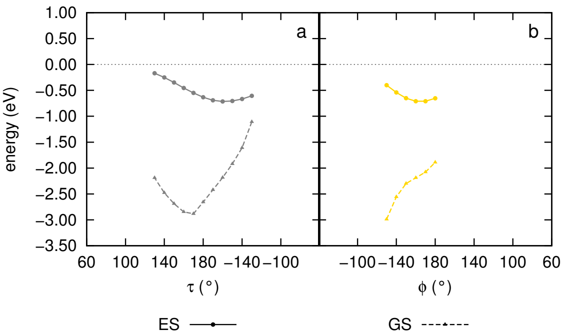

FIGURE A.3: OFF-state: Relaxed (ES-optimized) energy profile of the τ-constrained (a) and ϕ-constrained (b) isomerization pathway for the neutral CRO.

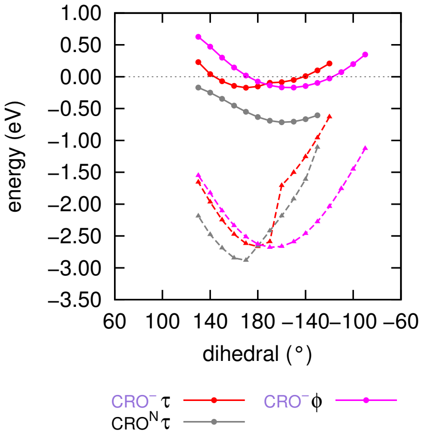

FIGURE A.4: OFF-state energy profile along τ (red), and ϕ (purple), constraint isomerization pathway for the Anionic CRO.

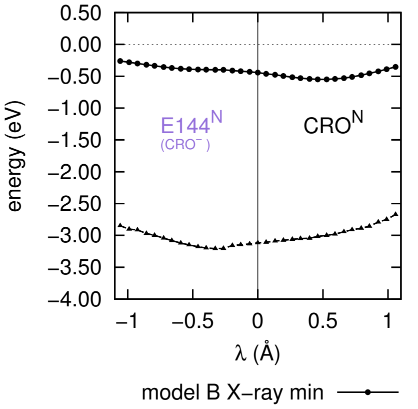

FIGURE A.5: OFF-state: Energy profile along the λ-constrained PT pathway. CC2 is used as QM method.

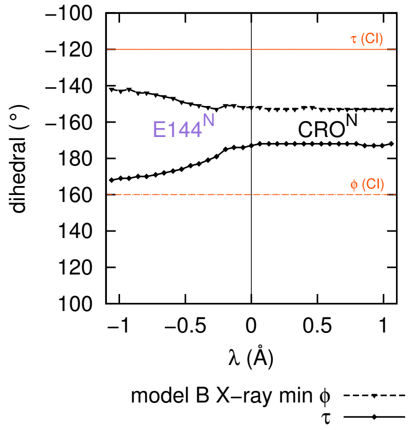

FIGURE A.6: OFF-state: τ and ϕ dihedrals along the λ-constrained PT pathway (CC2 is used as the QM method). The upper and lower horizontal lines mark the τ and ϕ values at the CI, respectively.

A.3 QM method for PT (test in IrisFP)

Figure A.7 shows the energy profile of the PT (reaction coordinate λ) in the GS and compares OM2/MRCI with MP2. OM2/MRCI underestimates the energy at the mid-point ($\lambda = 0$), which is related to the PT barrier, and places the proton too far from CRO.

In the ES, the situation is worse, as OM2/MRCI underestimates the change in the proton affinities and places the proton in the middle, a bit closer to the CRO, while in the reference calculation with CC2 the proton is closer to Glu144 than to the CRO (Fig. A.8).

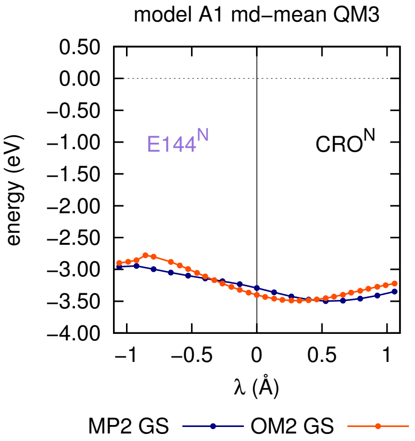

FIGURE A.7: OFF-state: OM2/MRCI and MP2 relaxed GS energy profiles along the λ-constrained PT pathway. The ϕ and τ dihedrals were fixed to the same GS values during optimization.

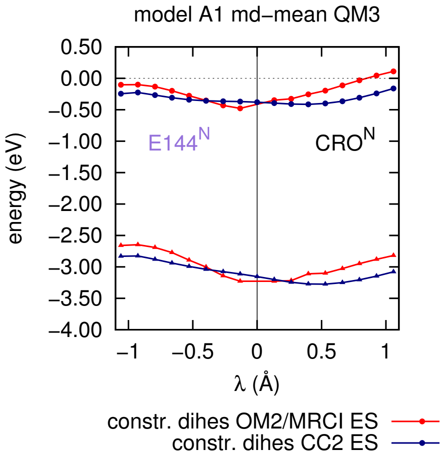

FIGURE A.8: OFF-state: OM2/MRCI and CC2 energy profiles along the ES-relaxed λ-constrained PT pathway. The ϕ and τ dihedrals were fixed to the same GS values during optimization.

A.4 OM2/MRCI τ-constrained isomerization pathway in IrisFP

As OM2/MRCI does not correctly describe the potential energy surface w.r.t. changes in λ (see Figure A.9b, orange curve), we fix λ to its value for either the neutral or the anionic CRO, as obtained from the CC2 τ-isomerization pathway. The result is shown in Fig. A.9 (black and blue curves). The barrier towards the CI still increases upon deprotonation of the CRO, in a similar manner with the QM2 region. This effect is much less pronounced when using CC2 as the QM method (Figure 14 in the main article).

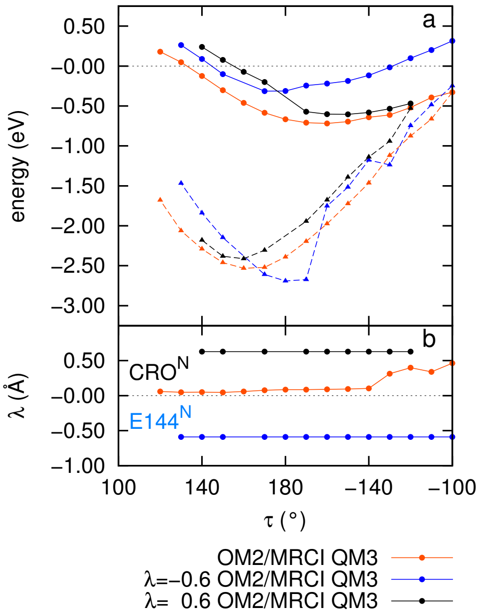

FIGURE A.9: OFF-state: (a) Energy profile along the τ-constrained isomerization pathway with QM3 and OM2/MRCI as the QM method. Optimization with free λ (orange) or with λ constrained to describe a neutral (black) or anionic (blue) CRO. (b) The corresponding λ values.

A.5 Active space (AS) selection

Table A.3) shows the active orbital spaces AS(m, n) with m occupied and n virtual orbitals for each QM region and CRO protonation state. They were chosen to be as large as possible within memory limitations and at the same time avoiding

TABLE A.3: AS(m,n)

QM region	CRO prot. state	AS
QM1	anionic	AS(10,10)
QM2	anionic	AS(10,10)
QM1	neutral	AS(10,8)
QM2	neutral	AS(11,11)
QM3	neutral	AS(8,6)

active/inactive pair mixing, which occurs when orbitals become nearly degenerate at some intermediate along the reaction coordinate. In some cases, the active space was chosen smaller to reduce SCF convergence problems.

A.6 Results with Dronpa model Y1

Results for Dronpa model Y1, which differs with Dronpa (Z) just in the protonation state of the Glu211 maintaining the same structure, already shows a behaviour much nearer to IrisFP:

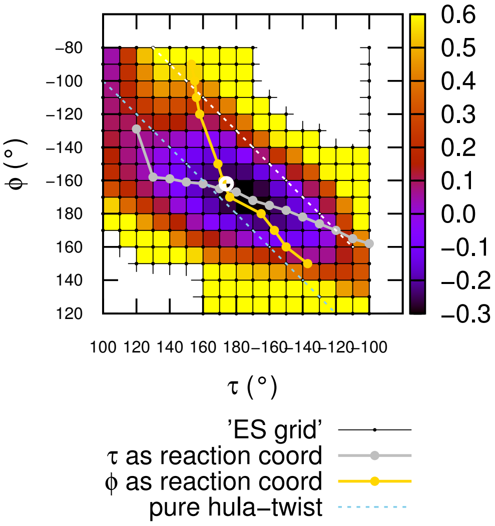

FIGURE A.10: **Dronpa model Y1** relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The white dot is the FC point.

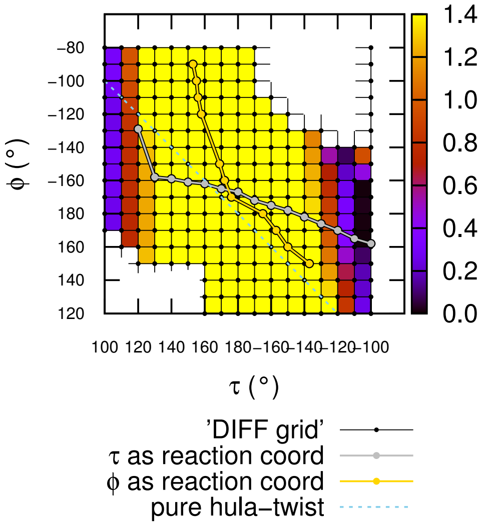

FIGURE A.11: **Dronpa model Y1** ES-GS energy difference of the relaxed 2D (τ, ϕ) scan of the ES potential energy surface for the neutral CRO (2D projection). The lower the value the nearer to a CI seam.

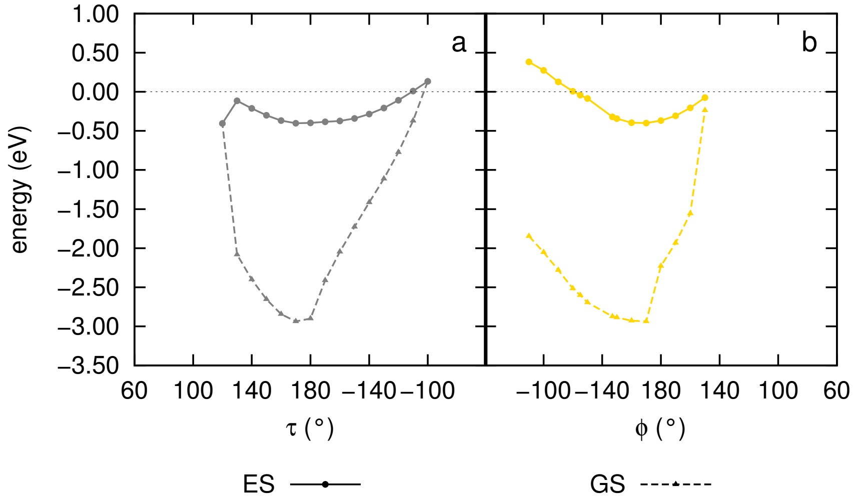

FIGURE A.12: **Dronpa model Y1** relaxed (ES-optimized) energy profile of the τ-constrained (a) and ϕ-constrained (b) isomerization pathway for the neutral CRO. Zero energy is set to the FC point.

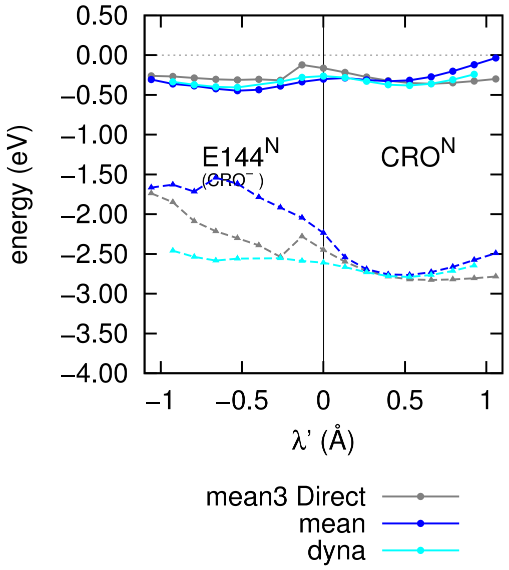

FIGURE A.13: **Dronpa model Y1** ES energy profile along the λ'-constrained PT pathway, except for "mean3 Direct", where the constrain is λ. CC2 is used as QM method. Zero energy is set to the FC point.

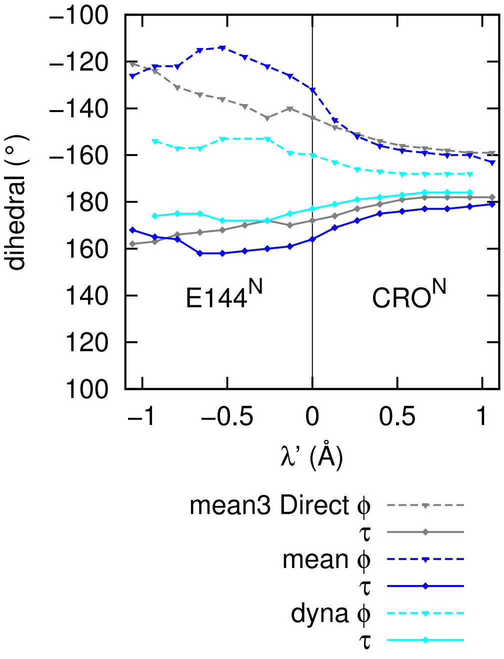

FIGURE A.14: **Dronpa model Y1** τ and ϕ dihedrals along the λ'-constrained PT pathway (CC2 is used as the QM method).

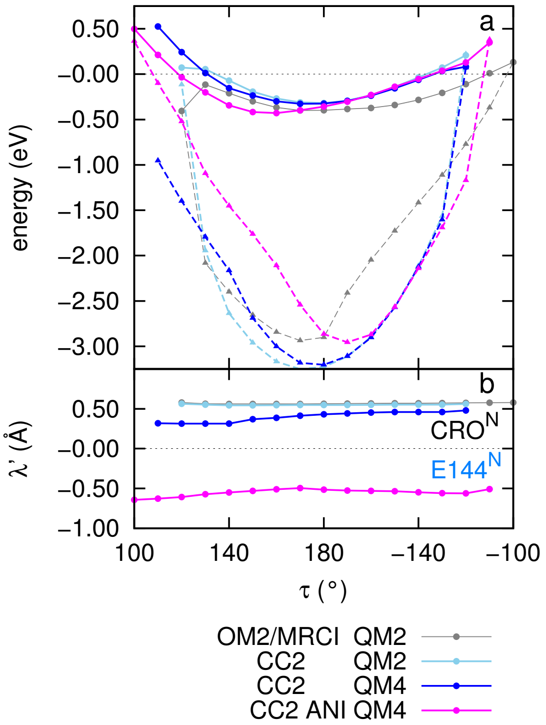

FIGURE A.15: **Dronpa model Y1** (a) Energy profile along the τ-constrained isomerization pathway with QM2 (light blue), with QM4 and optimizing from neutral CRO (dark blue) and with QM4 and from anionic CRO (magenta) using CC2 as the QM method. The OM2/MRCI profile for QM2 and neutral CRO is shown for comparison (gray). (b) The corresponding λ' values. Zero energy is set to the FC point.

Bibliography

[1] Xin X Zhou and Michael Z Lin. Photoswitchable fluorescent proteins: ten years of colorful chemistry and exciting applications. *Current Opinion in Chemical Biology*, 17(4):682–690, August 2013.

[2] Chenxi Duan, Virgile Adam, Martin Byrdin, and Dominique Bourgeois. Structural basis of photoswitching in fluorescent proteins. In Sidney Cambridge, editor, *Methods in Molecular Biology*, volume 1148, pages 177–202. Springer New York, 2014.

[3] Richard N. Day and Michael W. Davidson. The fluorescent protein palette: tools for cellular imaging. *Chem. Soc. Rev.*, 38:2887–2921, 2009.

[4] Douglas C. Prasher, Virginia K. Eckenrode, William W. Ward, Frank G. Prendergast, and Milton J. Cormier. Primary structure of the aequorea victoria green-fluorescent protein. *Gene*, 111(2):229 – 233, 1992.

[5] M Chalfie, Y Tu, G Euskirchen, WW Ward, and DC Prasher. Green fluorescent protein as a marker for gene expression. *Science*, 263(5148):802–805, 1994.

[6] Roger Heim, Andrew B. Cubitt, and Roger Y. Tsien. Improved green fluorescence. *Nature*, 373(6516):663–664, February 1995.

[7] Andrew B. Cubitt, Roger Heim, Stephen R. Adams, Aileen E. Boyd, Larry A. Gross, and Roger Y. Tsien. Understanding, improving and using green fluorescent proteins. *Trends in Biochemical Sciences*, 20(11):448 – 455, 1995.

[8] David C. Baulcombe, Sean Chapman, and Simon Santa Cruz. Jellyfish green fluorescent protein as a reporter for virus infections. *The Plant Journal*, 7(6):1045–1053, 1995.

[9] John Marshall, Raymond Molloy, Guy W.J Moss, James R Howe, and Thomas E Hughes. The jellyfish green fluorescent protein: A new tool for studying ion channel expression and function. *Neuron*, 14(2):211 – 215, 1995.

[10] Brendan P. Cormack, Raphael H. Valdivia, and Stanley Falkow. Facs-optimized mutants of the green fluorescent protein (gfp). *Gene*, 173(1):33 – 38, 1996. Flourescent Proteins and Applications.

[11] R. Heim, D. C. Prasher, and R. Y. Tsien. Wavelength mutations and post-translational autoxidation of green fluorescent protein. *PNAS*, 91:12501–12504, 1994.

[12] Roger Heim and Roger Y Tsien. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology*, 6(2):178 – 182, 1996.

[13] M. Ormo and S. J. Remington. Crystal structure of the aequorea victoria green fluorescent protein. *Science*, 273:1392–1395, 1996.

[14] F. Yang, L. G. Moss, and G. N. Phillips Jr. The molecular structure of green fluorescent protein. *Nat. Biotechnol.*, 14:1246–1251, 1996.

[15] Arkady F. Fradkov, Ying Chen, Li Ding, Ekaterina V. Barsova, Mikhail V. Matz, and Sergey A. Lukyanov. Novel fluorescent protein from discosoma coral and its mutants possesses a unique far-red fluorescence. *FEBS Letters*, 479(3):127–130, 2000.

[16] Nadya G. Gurskaya, Arkady F. Fradkov, Alexey Terskikh, Mikhail V. Matz, Yulii A. Labas, Vladimir I. Martynov, Yurii G. Yanushevich, Konstantin A. Lukyanov, and Sergey A. Lukyanov. Gfp-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Letters*, 507(1):16–20, 2001.

[17] Mikhail V. Matz, Arkady F. Fradkov, Yulii A. Labas, Aleksandr P. Savitsky, Andrey G. Zaraisky, Mikhail L. Markelov, and Sergey A. Lukyanov. Fluorescent proteins from nonbioluminescent anthozoa species. *Nat Biotech*, 17(10):969–973, October 1999.

[18] Marco A Mena, Thomas P Treynor, Stephen L Mayo, and Patrick S Daugherty. Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library. *Nat Biotech*, 24(12):1569–1571, December 2006.

[19] N. C. Shaner, R. E. Campbell, P. A. Steinbach, B. N. G. Giepmans, A. E. Palmer, and R. Y. Tsien. Improved monomeric red, orange and yellow fluorescent proteins derived from discosoma sp red fluorescent protein. *Nat. Biotechnol.*, 22(12):1567–1572, December 2004.

[20] D. Shcherbo, E. M. Merzlyak, T. V. Chepurnykh, A. F. Fradkov, G. V. Ermakova, E. A. Solovieva, K. A. Lukyanov, E. A. Bogdanova, A. G. Zaraisky, S. Lukyanov, and D. M. Chudakov. Bright far-red fluorescent protein for whole-body imaging. *Nat. Methods*, 4(9):741–746, September 2007.

[21] O. M. Subach, I. S. Gundorov, M. Yoshimura, F. V. Subach, J. H. Zhang, D. Gruenwald, E. A. Souslova, D. M. Chudakov, and V. V. Verkhusha. Conversion of red fluorescent protein into a bright blue probe. *Chemistry & Biology*, 59(10):1116–1124, October 2008.

[22] Wataru Tomosugi, Tomoki Matsuda, Tomomi Tani, Tomomi Nemoto, Ippei Kotera, Kenta Saito, Kazuki Horikawa, and Takeharu Nagai. An ultramarine fluorescent protein with increased photostability and ph insensitivity. *Nat Meth*, 6(5):351–353, May 2009.

[23] Lei Wang, W. Coyt Jackson, Paul A. Steinbach, and Roger Y. Tsien. Evolution of new nonantibody proteins via iterative somatic hypermutation. *PNAS*, 101(48):16745–16749, 2004.

[24] Daniele Arosio, Gianpiero Garau, Fernanda Ricci, Laura Marchetti, Ranieri Bizzarri, Riccardo Nifos, and Fabio Beltram. Spectroscopic and structural

study of proton and halide ion cooperative binding to gfp. *Biophysical Journal*, 93(1):232–244, March 2007.

[25] Nadya G Gurskaya, Vladislav V Verkhusha, Alexander S Shcheglov, Dmitry B Staroverov, Tatyana V Chepurnykh, Arkady F Fradkov, Sergey Lukyanov, and Konstantin A Lukyanov. Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nat Biotech*, 24(4):461–465, April 2006.

[26] George H. Patterson and Jennifer Lippincott-Schwartz. A photoactivatable gfp for selective photolabeling of proteins and cells. *Science*, 297(5588):1873–1877, 2002.

[27] Rebekka M. Wachter and S. James Remington. Sensitivity of the yellow variant of green fluorescent protein to halides and nitrate. *Current Biology*, 9(17):R628 – R629, 1999.

[28] Tsai-Wen Chen, Trevor J. Wardill, Yi Sun, Stefan R. Pulver, Sabine L. Renninger, Amy Baohan, Eric R. Schreiter, Rex A. Kerr, Michael B. Orger, Vivek Jayaraman, Loren L. Looger, Karel Svoboda, and Douglas S. Kim. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, 499(7458):295–300, July 2013.

[29] Yi Shen, Morgane Rosendale, Robert E. Campbell, and David Perrais. phuji, a ph-sensitive red fluorescent protein for imaging of exo- and endocytosis. *The Journal of Cell Biology*, 207(3):419–432, 2014.

[30] Michael J Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nat Meth*, 3(10):793–796, October 2006.

[31] Martin Andresen, Andre C Stiel, Jonas Folling, Dirk Wenzel, Andreas Schonle, Alexander Egner, Christian Eggeling, Stefan W Hell, and Stefan Jakobs. Photoswitchable fluorescent proteins enable monochromatic multilabel imaging and dual color fluorescence nanoscopy. *Nat Biotech*, 26(9):1035–1040, September 2008.

[32] Peter Dedecker, Gary C. H. Mo, Thomas Dertinger, and Jin Zhang. Widely accessible method for superresolution fluorescence imaging of living systems. *Proceedings of the National Academy of Sciences*, 109(27):10909–10914, July 2012.

[33] Stefan W Hell, Marcus Dyba, and Stefan Jakobs. Concepts for nanoscale resolution in fluorescence microscopy. *Current Opinion in Neurobiology*, 14(5):599–609, October 2004.

[34] Mats G. L. Gustafsson. Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 102(37):13081–13086, September 2005.

[35] Robert M. Dickson, Andrew B. Cubitt, Roger Y. Tsien, and W. E. Moerner. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature*, 388(6640):355–358, July 1997.

[36] Daniel Sinnecker, Philipp Voigt, Nicole Hellwig, and Michael Schaefer. Reversible photobleaching of enhanced green fluorescent proteins. *Biochemistry*, 44(18):7085–7094, May 2005.

[37] R. Nifosi, A. Ferrari, C. Arcangeli, V. Tozzini, V. Pellegrini, and F. Beltram. Photoreversible dark state in a tristable green fluorescent protein variant. *J. Phys. Chem. B*, 107(7):1679–1684, February 2003.

[38] Tim B. McAnaney, Wei Zeng, Camille F. E. Doe, Nina Bhanji, Stuart Wakelin, David S. Pearson, Paul Abbyad, Xinghua Shi, Steven G. Boxer, and Clive R. Bagshaw. Protonation, photobleaching, and photoactivation of yellow fluorescent protein (yfp 10c): a unifying mechanism. *Biochemistry*, 44(14):5510–5524, April 2005.

[39] Dmitriy M. Chudakov, Alexei V. Feofanov, Nikolay N. Mudrik, Sergey Lukyanov, and Konstantin A. Lukyanov. Chromophore environment provides clue to kindling fluorescent protein riddle. *Journal of Biological Chemistry*, 278(9):7215–7219, 2003.

[40] Ryoko Ando, Hideaki Mizuno, and Atsushi Miyawaki. Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science*, 306(5700):1370–1373, November 2004.

[41] Satoshi Habuchi, Ryoko Ando, Peter Dedecker, Wendy Verheijen, Hideaki Mizuno, Atsushi Miyawaki, and Johan Hofkens. Reversible single-molecule photoswitching in the gfp-like fluorescent protein dronpa. *PNAS*, 102(27):9511–9516, 2005.

[42] M. Andresen, M. C. Wahl, A. C. Stiel, F. Grater, L. V. Schafer, S. Trowitzsch, G. Weber, C. Eggeling, H. Grubmuller, S. W. Hell, and S. Jakobs. Structure and mechanism of the reversible photoswitch of a fluorescent protein. *PNAS*, 102(37):13070–13074, September 2005.

[43] Martin Andresen, Andre C. Stiel, Simon Trowitzsch, Gert Weber, Christian Eggeling, Markus C. Wahl, Stefan W. Hell, and Stefan Jakobs. Structural basis for reversible photoswitching in dronpa. *Proceedings of the National Academy of Sciences*, 104(32):13005–13009, August 2007.

[44] Hideaki Mizuno, Tapas Kumar Mal, Markus Wlchli, Akihiro Kikuchi, Takashi Fukano, Ryoko Ando, Jeyaraman Jeyakanthan, Junichiro Taka, Yoshitsugu Shiro, Mitsuhiko Ikura, and Atsushi Miyawaki. Light-dependent regulation of structural flexibility in a photochromic fluorescent protein. *Proceedings of the National Academy of Sciences*, 105(27):9227–9232, 2008.

[45] Peter Dedecker, Jun-ichi Hotta, Ryoko Ando, Atsushi Miyawaki, Yves Engelborghs, and Johan Hofkens. Fast and reversible photoswitching of the fluorescent protein dronpa as evidenced by fluorescence correlation spectroscopy. *Biophysical Journal*, 91(5):L45–L47, June 2006.

[46] Eduard Fron, Cristina Flors, Gerd Schweitzer, Satoshi Habuchi, Hideaki Mizuno, Ryoko Ando, Frans C. De Schryver, Atsushi Miyawaki, and Johan Hofkens. Ultrafast excited-state dynamics of the photoswitchable protein dronpa. *J. Am. Chem. Soc.*, 129(16):4870–4871, April 2007.

[47] Pascal G. Wilmann, Kristina Turcic, Jion M. Battad, Matthew C.J. Wilce, Rodney J. Devenish, Mark Prescott, and Jamie Rossjohn. The 1.7 crystal structure of dronpa: A photoswitchable green fluorescent protein. *Journal of Molecular Biology*, 364(2):213 – 224, 2006.

[48] Andre C. Stiel, Simon Trowitzsch, Gert Weber, Martin Andresen, Christian Eggeling, Stefan W. Hell, Stefan Jakobs, and Markus C. Wahl. 1.8 å bright-state structure of the reversibly switchable fluorescent protein dronpa guides the generation of fast switching variants. *Biochemical Journal*, 402(1):35–42, 2007.

[49] X. Li, L. W. Chung, H. Mizuno, A. Miyawaki, and K. Morokuma. A theoretical study on the nature of on- and off-states of reversibly photoswitching fluorescent protein dronpa: Absorption, emission, protonation, and raman. *J. Phys. Chem. B*, 114(2):1114–1126, January 2010.

[50] Xin Li, Lung Wa Chung, Hideaki Mizuno, Atsushi Miyawaki, and Keiji Morokuma. Primary events of photodynamics in reversible photoswitching fluorescent protein dronpa. *J. Phys. Chem. Lett.*, 1(23):3328–3333, December 2010.

[51] Mark M. Warren, Marius Kaucikas, Ann Fitzpatrick, Paul Champion, J. Timothy Sage, and Jasper J. van Thor. Ground-state proton transfer in the photoswitching reactions of the fluorescent protein dronpa. *Nat Commun*, 4:1461, February 2013.

[52] M. L. Quillin, D. A. Anstrom, X. K. Shu, S. O'Leary, K. Kallio, D. A. Chudakov, and S. J. Remington. Kindling fluorescent protein from anemonia sulcata: Dark-state structure at 1.38 angstrom resolution. *Biochemistry*, 44(15):5774–5787, April 2005.

[53] T. A. Schüttrigkeit, T. von Feilitzsch, C. K. Kompa, K. A. Lukyanov, A. P. Savitsky, A. A. Voityuk, and M. E. Michel-Beyerle. Femtosecond study of light-induced fluorescence increase of the dark chromoprotein asfp595. *Chem. Phys.*, 323(2-3):149–160, April 2006.

[54] L. V. Schäfer, G. Groenhof, A. R. Klingen, G. M. Ullmann, M. Boggio-Pasqua, M. A. Robb, and H. Grubmüller. Photoswitching of the fluorescent protein asfp595: Mechanism, proton pathways, and absorption spectra. *Angew. Chem. Int. Ed.*, 46(4):530–536, 2007.

[55] L. V. Schäfer, G. Groenhof, M. Boggio-Pasqua, M. A. Robb, and H. Grubmüller. Chromophore protonation state controls photoswitching of the fluoroprotein asfp595. *PLoS Comput. Biol.*, 4(3):e1000034, March 2008.

[56] Bella L. Grigorenko, Igor V. Polyakov, Alexander P. Savitsky, and Alexander V. Nemukhin. Unusual emitting states of the kindling fluorescent protein: Appearance of the cationic chromophore in the gfp family. *J. Phys. Chem. B*, 117(24):7228–7234, June 2013.

[57] Dheerendra Yadav, Fabien Lacombat, Nadia Dozova, Fabrice Rappaport, Pascal Plaza, and Agathe Espagne. Real-time monitoring of chromophore isomerization and deprotonation during the photoactivation of the fluorescent protein dronpa. *J. Phys. Chem. B*, 119(6):2404–2414, February 2015.

[58] J. Nathan Henderson, Hui-wang Ai, Robert E. Campbell, and S. James Remington. Structural basis for reversible photobleaching of a green fluorescent protein homologue. *Proceedings of the National Academy of Sciences*, 104(16):6672–6677, 2007.

[59] V. Adam, M. Lelimousin, S. Boehme, G. Desfonds, K. Nienhaus, M. J. Field, J. Wiedenmann, S. McSweeney, G. U. Nienhaus, and D. Bourgeois. Structural characterization of irisfp, an optical highlighter undergoing multiple photo-induced transformations. *PNAS*, 105(47):18343–18348, November 2008.

[60] Tanja Brakemann, Gert Weber, Martin Andresen, Gerrit Groenhof, Andre C Stiel, Simon Trowitzsch, Christian Eggeling, Helmut Grubmller, Stefan W Hell, Markus C Wahl, and Stefan Jakobs. Molecular basis of the light-driven switching of the photochromic fluorescent protein padron. *The Journal of Biological Chemistry*, 285(19):14603–14609, February 2010.

[61] Aline Regis Faro, Philippe Carpentier, Gabriella Jonasson, Guillaume Pompidor, Delphine Arcizet, Isabelle Demachy, and Dominique Bourgeois. Low-temperature chromophore isomerization reveals the photoswitching mechanism of the fluorescent protein padron. *J. Am. Chem. Soc.*, 133(41):16362–16365, October 2011.

[62] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 103(2):227–249, 1976.

[63] Hans Martin Senn and Walter Thiel. QM/MM methods for biological systems. *Top. Curr. Chem.*, 268:173–290, 2007.

[64] Dominique Bourgeois and Virgile Adam. Reversible photoswitching in fluorescent proteins: A mechanistic view. *IUBMB Life*, 64(6):482–491, 2012.

[65] S. Violot, P. Carpentier, L. Blanchoin, and D. Bourgeois. Reverse ph-dependence of chromophore protonation explains the large stokes shift of the red fluorescent protein mkeima. *J. Am. Chem. Soc.*, 131(30):10356–+, August 2009.

[66] V. Adam, B. Moeyaert, C. C. David, H. Mizuno, M. Lelimousin, P. Dedecker, R. Ando, A. Miyawaki, J. Michiels, Y. Engelborghs, and J. Hofkens. Rational design of photoconvertible and biphotochromic fluorescent proteins for advanced microscopy applications. *Chemistry & Biology*, 18(10):1241–1251, October 2011.

[67] Ranieri Bizzarri, Michela Serresi, Francesco Cardarelli, Stefania Abbruzzetti, Barbara Campanini, Cristiano Viappiani, and Fabio Beltram. Single amino acid replacement makes aequorea victoria fluorescent proteins reversibly photoswitchable. *J. Am. Chem. Soc.*, 132(1):85–95, January 2010.

[68] Seth Olsen, Kristina Lamothe, and Todd J. Martnez. Protonic gating of excited-state twisting and charge localization in gfp chromophores: A mechanistic hypothesis for reversible photoswitching. *J. Am. Chem. Soc.*, 132(4):1192–1193, January 2010.

[69] Aline Regis Faro, Virgile Adam, Philippe Carpentier, Claudine Darnault, Dominique Bourgeois, and Eve de Rosny. Low-temperature switching by photoinduced protonation in photochromic fluorescent proteins. *Photochem. Photobiol. Sci.*, 9(2):254–62, February 2010.

[70] Q. Sun, Z. Li, Z. G. Lan, C. Pfisterer, M. Doerr, S. Fischer, S. C. Smith, and W. Thiel. Isomerization mechanism of the hcred fluorescent protein chromophore. *Phys. Chem. Chem. Phys.*, 14(32):11413–11424, 2012.

[71] Marius Kaucikas, Ann Fitzpatrick, Elana Bryan, Abelone Struve, Robert Henning, Irina Kosheleva, Vukica Srajer, Gerrit Groenhof, and Jasper J. Van Thor. Room temperature crystal structure of the fast switching m159t mutant of the fluorescent protein dronpa. *Proteins*, 83(3):397–402, 2015.

[72] Andras Lukacs, Allison Haigney, Richard Brust, Kiri Addison, Michael Towrie, Gregory M. Greetham, Garth A. Jones, Atsushi Miyawaki, Peter J. Tonge, and Stephen R. Meech. Protein photochromism observed by ultrafast vibrational spectroscopy. *The Journal of Physical Chemistry B*, 117(40):11954–11959, 2013. PMID: 24033093.

[73] Dmitry Morozov and Gerrit Groenhof. Hydrogen bond fluctuations control photochromism in a reversibly photo-switchable fluorescent protein. *Angewandte Chemie International Edition*, 55(2):576–578, 2016.

[74] Marius Kaucikas, Martijn Tros, and Jasper J. van Thor. Photoisomerization and proton transfer in the forward and reverse photoswitching of the fast-switching m159t mutant of the dronpa fluorescent protein. *The Journal of Physical Chemistry B*, 119(6):2350–2362, 2015. PMID: 25369171.

[75] M. Born and R. Oppenheimer. Quantum theory of molecules. *Analen d. Physik*, 84(4):457, 1927.

[76] M. R. Silva-Junior and W. Thiel. Benchmark of electronically excited states for semiempirical methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2. *J. Chem. Theory Comput.*, 6(5):1546–1564, May 2010.

[77] Martial Boggio-Pasqua, Carl F. Burmeister, Michael A. Robb, and Gerrit Groenhof. Photochemical reactions in biological systems: probing the effect of the environment by means of hybrid quantum chemistry/molecular mechanics simulations. *Phys. Chem. Chem. Phys.*, 14:7912–7928, 2012.

[78] W. Klopper, F. R. Manby, S. Ten-No, and E. F. Valeev. R12 methods in explicitly correlated molecular electronic structure theory. *International Reviews in Physical Chemistry*, 25(3):427–468, July 2006.

[79] E . Hylleraas and B. Undheim. Numerische berecbnung der 2 s-terme von ortho- und par-helium. *Z. Phys.*, 65:759–772, 1930.

[80] J. K. L. MacDonald. Successive approximations by the Rayleigh–ritz variation method. *Phys. Rev.*, 43:830–833, 1933.

[81] R. J. Bartlett. Many-body perturbation-theory and coupled cluster theory for electron correlation in molecules. *Annu. Rev. Phys. Chem.*, 32:359–401, 1981.

[82] C. C. J. Roothaan. New developments in molecular orbital theory. *Rev. Mod. Phys.*, 23:69, 1951.

[83] GG Hall. The molecular orbital theory of chemical valency. 8 . a method of calculating ionization potentials. *Proc. of the Royal Society of London. Series A. Mathematical and physical sciences.*, 1951.

[84] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry (Introduction to advanced electronic structure theory)*. Dover, 1996.

[85] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, 1964.

[86] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, 1965.

[87] Marius Wanko, Michael Hoffmann, Paul Strodel, Axel Koslowski, Walter Thiel, Frank Neese, Thomas Frauenheim, and Marcus Elstner. Calculating

absorption shifts for retinal proteins: Computational challenges. *J. Phys. Chem. B*, 109:3606–3615, 2005.

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

[89] C. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46(7):0618–0622, October 1934.

[90] O. Christiansen, H. Koch, and P. Jorgensen. The 2nd-order approximate coupled-cluster singles and doubles model CC2. *Chem. Phys. Lett.*, 243(5-6):409–418, September 1995.

[91] A. Köhn. Can coupled-cluster theory treat conical intersections? *J. Chem. Phys.*, 127(4):044105, 2007.

[92] W. Weber and W. Thiel. Orthogonalisation corrections for semiempirical methods. *Theor. Chem. Acc.*, 103:495–506, 2000.

[93] Marius Wanko, Pablo García-Risueño, and Angel Rubio. Excited states of the green fluorescent protein chromophore: Performance of ab initio and semi-empirical methods. *Phys. Status Solidi B*, 249(2):392–400, 2012.

[94] Jr A. D. MacKerell, D. Bashford, M. Bellott, Jr. R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wirkiewicz-Kuczera, D. Yin, , and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 1998.

[95] Paul Sherwood, Alex H. de Vries, Simon J. Collins, Stephen P. Greatbanks, Neil A. Burton, Mark A. Vincent, and Ian H. Hillier. Computer simulation

of zeolite structure and reactivity using embedded cluster methods. *Faraday Discuss.*, 106(0):79–92, 1997.

[96] Chemshell, a computational chemistry shell, see www.chemshell.org.

[97] J. Kastner, J. M. Carr, T. W. Keal, W. Thiel, A. Wander, and P. Sherwood. Dl-find: An open-source geometry optimizer for atomistic simulations. *J. Phys. Chem. A*, 113(43):11856–11865, October 2009.

[98] Walter Thiel. Mndo program version 6.1, mülheim, 2007.

[99] TURBOMOLE V6.0 2009, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`.

[100] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4(2):187–217, 1983.

[101] Mats H. M. Olsson, Chresten R. Sondergaard, Michal Rostkowski, and Jan H. Jensen. Propka3: Consistent treatment of internal and surface residues in empirical pk(a) predictions. *J. Chem. Theory Comput.*, 7(2):525–537, February 2011.

[102] S. Nosé. A unified formulation of the constant-temperature molecular-dynamics methods. *J. Chem. Phys.*, 81:551, 1984.

[103] W.G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695, 1985.

[104] A. R. Dinner, X. Lopez, and M. Karplus. A charge-scaling method to treat solvent in qm/mm simulations. *Theor. Chem. Acc.*, 109:118, 2003.

[105] Michael Hoffmann, Marius Wanko, Paul Strodel, Peter H. König, Thomas Frauenheim, Klaus Schulten, Walter Thiel, Emad Tajkhorshid, and Marcus Elstner. Color tuning in rhodopsins: The mechanism for the spectral shift between bacteriorhodopsin and sensory rhodopsin II. *J. Am. Chem. Soc.*, 128:10808–10818, 2006.

[106] Jacques-Philippe Colletier, Michel Sliwa, Franois-Xavier Gallat, Michihiro Sugahara, Virginia Guillon, Giorgio Schir, Nicolas Coquelle, Joyce Woodhouse, Laure Roux, Guillaume Gotthard, Antoine Royant, Lucas Martinez Uriarte, Cyril Ruckebusch, Yasumasa Joti, Martin Byrdin, Eiichi Mizohata, Eriko Nango, Tomoyuki Tanaka, Kensuke Tono, Makina Yabashi, Virgile Adam, Marco Cammarata, Ilme Schlichting, Dominique Bourgeois, and Martin Weik. Serial femtosecond crystallography and ultrafast absorption spectroscopy of the photoswitchable fluorescent protein irisfp. *The Journal of Physical Chemistry Letters*, 7(5):882–887, 2016. PMID: 26866390.

[107] Xuhui Zhuang, Jun Wang, and Zhenggang Lan. Tracking of the molecular motion in the primary event of photoinduced reactions of a phytochromobilin model. *The Journal of Physical Chemistry B*, 117(50):15976–15986, 2013. PMID: 24261565.

[108] Ganglong Cui, Zhenggang Lan, and Walter Thiel. Intramolecular hydrogen bonding plays a crucial role in the photophysics and photochemistry of the GFp chromophore. *J. Am. Chem. Soc.*, 134(3):1662–1672, December 2011.

[109] Wolfgang Weber, Volkhard Helms, J. Andrew McCammon, and Peter W. Langhoff. Shedding light on the dark and weakly fluorescent states of green fluorescent proteins. *PNAS*, 96:6177–6182, 1999.

[110] Susan Gayda, Karin Nienhaus, and G. Ulrich Nienhaus. Mechanistic insights into reversible photoactivation in proteins of the GFp family. *Biophys. J.*, 103(12):2521–2531, December 2012.

[111] Ryoko Ando, Cristina Flors, Hideaki Mizuno, Johan Hofkens, and Atsushi Miyawaki. Highlighted generation of fluorescence signals using simultaneous two-color irradiation on dronpa mutants. *Biophysical Journal*, 92(12):L97 – L99, 2007.

Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

image kaibara87 / © Some rights reserved