PHOTOINDUCED REDOX REACTIONS IN BIOLOGICALLY RELEVANT SYSTEMS

by

Atanu Acharya

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

December 2016

Copyright 2016 Atanu Acharya
To Mr. Nalini R. Mabhai and Prof. N. Chandrakumar

for being my sources of encouragement
Acknowledgements

Graduate school at USC has been a tremendous learning experience and a place of self discovery for me. I will always remember it as one of the most defining chapter of my life. I was lucky to have Professor Anna Krylov as my advisor here. Her invaluable advice and encouragement has helped me to understand and appreciate computational chemistry. In the beginning, she always presents her students with several scientific direction that one would like to take and one can choose according to their interest. She is a great scientist and an awesome advisor who always encouraged me to try different scientific challenges. This inspired me to pursue a project on green fluorescent protein, although I did not come from a biology background. I really enjoyed the freedom she gives to try out my own scientific ideas, no matter how crazy they sound. One of her best qualities is the ways she gradually exposes her students into the scientific community. I have attended 9 conferences and 1 workshop during my PhD which allowed me to present my work in front of huge scientific community. In addition, whenever a scientist visits USC, she always schedules an one-on-one meeting for the student. I cannot overstate the importance of those conferences and meetings in my professional growth.

I feel lucky to be a part of USC Department of Chemistry because of the great scientific environment. I could approach any professor for any kind of discussion. I would like to thank Prof. Jahan Dawlaty for teaching me the tricks to get other professors notice
my email, while applying for postdoctoral position. I would like to thank my committee members Professors Arieh Warshel, Alex Benderskii, Sri Narayan, Rosa Di Felice and Moh El-Naggar. When Moh had to leave for jury duty, two days before my qualifying examination, Rosa offered to join my committee in such a short notice. I would like to thank Moh for all the scientific discussion about electron transfer in bacterial nanowires.

I would also like to thank all my collaborators, Professors Ksenia Bravaya, Anatoly Kolomeisky, Konstantin Lukyanov, Lyudmila Slipchenko and Debashree Ghosh. The scientific work presented in the thesis would not have been possible without their help. Debashree and Ksenia also mentored me during my early years in USC. They have been great role models for me even after they left Anna’s group to pursue their independent scientific career.

I made some really good friends in Anna’s group; Dr. Samer Gozem, Dr. Kaushik Nanda were always there to cheer me up. I would also like to thank Dr. Shirin Faraji, Dr. Ilya Kaliman, Dr. Matthias Schneider, Dr. Marc de Wergifosse, Xintian Feng, Natalie Orms, Arman Sadybekov for being awesome labmates. A special shout-out goes to Samer and Natalie for carefully proofreading my thesis. I also made some great friends outside the lab, Amit Samanta, Anirban Roy, Saptaparna Das, Parichita Mazumder, Chayan Dutta, Gaurav Kumar, Deepak Verma. Amit and Saptaparna deserve a special thank you for critically reviewing my dumb proposal for qualifying examination and transforming it into great one. I am grateful to Anirban and Saptaparna for hosting me in their apartment when I arrived in the United States.

I will treasure the time I spent with my awesome roommates, Purnim Dhar, Piyush Deokar, Subodh Tiwari and Amit Samanta. Most of our conversations either started or ended with “Let’s go there”. From planing for a road trip at 2am and leaving the next
morning to endless horrible movies they made me watch in the theater, they will always have a special place in my memory.

I would like to take this opportunity to appreciate the support and encouragement my parents showed over the years. I thank my wife, Deepika for all her support and love during my failure and success. She deserves a big shout-out for diligently proofreading my entire thesis and correcting millions of mistakes I made. She believed in me even when I did not. She helped me realize the importance of balance between work and life. Life with her has been nothing I ever experienced before and I want to keep exploring it forever.

This list would not be complete without thanking the great teachers I have had the privilege to learn from. I would like to thank my first chemistry teacher, Mr. Nalini Ranjan Mabhai who inspired me to study chemistry. Prof. N. Chandrakumar and Prof. S. Kumar in IITM showed me how fun the world of Quantum Mechanics can be. Prof. A. Patnaik taught me to be always honest to my work. Above all, they inspired and encouraged me to pursue my dreams.
# Table of Contents

Dedication ii
Acknowledgements iii
List of Tables viii
List of Figures xi
Abbreviations xvi
Abstract xix

Chapter 1: Introduction and overview 1
  1.1 Green Fluorescent Proteins ............................. 1
  1.2 Fluorescent protein photocycle ........................ 7
  1.3 Photoinduced transformations in fluorescent proteins 10
  1.4 Photoinduced electron transfer: A gateway step leading to multiple outcomes .................. 13
Chapter 1 References ........................................ 19

Chapter 2: Methodology 26
  2.1 Marcus theory of electron transfer ......................... 26
  2.2 Calculating $\Delta G$ and $\lambda$ for a redox process using LRA .................. 30
  2.3 Calculating $\Delta G$ and $\lambda$ for an ET process using LRA ..................... 32
  2.4 MD simulation setup ..................................... 36
    2.4.1 Protein structures and protonation states ................... 37
  2.5 The Pathways model ..................................... 40
  2.6 Calculations of electronic couplings ....................... 44
    2.6.1 Constrained DFT method and CDFT-CI ...................... 47
  2.7 Docking calculations .................................... 50
  2.8 Protocols for calculating energetics and couplings .......... 53
    2.8.1 QM/MM schemes ....................................... 53
    2.8.2 Benchmarks ........................................... 58
List of Tables

2.1 eGFP Pathways calculations: $T_{DA}$ from chromophore to potential electron acceptors. ........................................ 42
2.2 $T_{DA}$ from the chromophore to potential electron acceptors in eYFP using the Pathways model. ........................................ 43
2.3 $T_{DA}$ from Tyr203 to potential electron acceptors in eYFP using the Pathways model. ........................................ 43
2.4 Effect of different bonded network in the bridge between D and A. $H_{DA}$ calculated using CDFT-CI method. .............................. 50
2.5 Effect of different carbon-carbon bonded network in the bridge between D and A. These results illustrate the distance-dependent decay of coupling. 50
2.6 Comparison between the computed (SOS-CIS(D)/aug-cc-pVDZ) and experimental excitation energies (eV). Only the chromophore is included in the QM part and the rest of the protein and solvent were treated as point charges. Computed values were averaged over 19 snapshots. .... 58
2.7 Basis set sensitivity of the VDE of the chromophore of YFP (without halide). The first column is taken to be the default (base) values, and the other columns list deviation from the base value. ................................. 59
2.8 Basis set sensitivity on the VDE of the chromophore of YFP (with halide). The first column is taken to be the default (base) values, and second column lists deviation from the base value. ................................. 59
2.9 Effect of adding Cl$^-$ in QM region on the VDE of the chromophore of YFP (with halide). .................................................. 60
2.10 Basis set sensitivity of the VEA of the Tyr145 residue of YFP (without halide). The first column is taken to be the default (base) values, and the other three columns list deviation from the base value. .................. 61
2.11 Comparison between VEAs of Tyr145 obtained at the EOM-EA-CCSD and DFT with same basis sets for 4 frames obtained from the ground-state trajectory of eYFP without halide. Only Tyr145 was included in the QM part of the QM/MM calculation. All values are given in eV. .... 62
2.12 Effect of adding Cl$^-$ in the QM part on the VEA of Tyr203 of YFP (with halide). ................................................................. 63
2.13 Redox properties of the chromophores of eGFP, eYFP and halide-bound eYFP. VDEs of the chromophores on the reduced (ground) and oxidized surfaces were averaged over 41 snapshots using $\omega$B97X-D/aug-cc-pVDZ. Energies are in eV and the reduction potentials are in V with respect to SHE. ................................. 64

2.14 Energy differences between the ground and CT states for 41 frames along the MD trajectory calculated on the ground-state surfaces. All values are in eV. ................................. 69

2.15 Energy differences between the ground and CT states for 41 frames along the MD trajectory calculated on the CT-state surfaces. All values are in eV. ................................. 70

3.1 Vertical and adiabatic ionization/detachment energies (eV) of the model FP chromophores and phenolic species. ................................. 87

3.2 Selected geometric parameters (Å) of the model chromophores and the respective oxidized species. ................................. 92

3.3 Mulliken analysis of spin densities in the oxidized species. ................................. 94

3.4 Free energies of solvation (kcal mol$^{-1}$) in acetonitrile for the model chromophores. ................................. 94

3.5 Free energies of solvation (kcal mol$^{-1}$) in water for the model chromophores. ................................. 95

3.6 Standard reduction potentials versus SHE ($E^0$, V) in acetonitrile and water for the model protein chromophores ($HA^+/HA$ and $A^-/A$). ................................. 97

3.7 Vertical ionization energies (eV) calculated with $\omega$B97x-D and EOM-IP-CCSD ................................. 100

3.8 Entropy change for gas-phase oxidation reaction. The entropies are calculated within the rigid rotor harmonic oscillator (RRHO) approximation. The entropies are in cal/mol$\cdot$K and T$\cdot$ΔS is in kcal/mol. ................................. 101

3.9 Free energy change (eV) of gas-phase oxidation reaction. ................................. 102

4.1 Redox properties of the ground-state and electronically excited chromophores of eGFP, eGFP-Y145L, eYFP and halide-bound eYFP at T=298 K. ................................. 129

4.2 Relevant Gibbs free energy differences, reorganization energies, and couplings and Marcus rates for ET at 298 K. Energy and coupling values are given in eV and eV$^2$, respectively, and the rate constants are in s$^{-1}$. ................................. 130

4.3 Average values of relevant structural parameters for eGFP, eYFP, and eYFP+Cl$^-$. The standard deviations are shown in parenthesis. See Fig. 4.7. ................................. 133

4.4 Temperature dependence of the computed rates and yields for eGFP and eYFP+Cl$^-$ assuming T-independent couplings. ................................. 142
4.5 Direct tunneling probabilities from the chromophore to the closest docked BQ in eGFP and eGFP-Y145L. .............................. 143
4.6 Electronic couplings for direct ET from the chromophore to the closest docked BQ in eGFP and eGFP-Y145L. .............................. 144
4.7 Rates for direct ET (Chro\textsuperscript{-*} \rightarrow BQ) at T=298 K. .............................. 146
List of Figures

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

1.2 A typical structure of a fluorescent protein represented by eGFP. In all fluorescent proteins, the chromophore, which is formed autocatalytically upon protein folding, is buried inside a tight 11-stranded $\beta$-barrel comprising 220-240 amino-acids. The approximate molar weight is 25 to 30 kDa. The diameter of the barrel is $\sim$24 Å and its height is $\sim$42 Å.

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

1.4 Various light-induced phenomena in fluorescent proteins.
Different mechanisms for ET. The relevant states are the bright excited state ($S_1$) and the charge-transfer (CT) state. In photooxidation, the latter is of $D^+A^-$ character (or $D^\cdot A^-$, depending on the protonation state of the chromophore). Top left: ET between the donor and acceptor by the Marcus mechanism. Top right: Adiabatic evolution of the initially excited state leading to CT via a barrier. Bottom left: CT state accessed by radiationless relaxation from a higher excited state. Bottom right: ET via direct one- or multi-photon excitation of the excited state of CT character.

Relevant MOs and leading electronic configurations of the CT states in wt-GFP. Two CT states of the Glu222→Chro character are located around 4-6 eV above the ground state.

Relevant MOs and leading electronic configurations of the locally excited chromophore and the CT states of Chro$^-\rightarrow$O$_2$ character.

Difference between Libby’s theory and Marcus theory of treating electron transfer. R and P represent reactants and products states.

Gibbs free energy curves and definitions of the key quantities in the Marcus theory of electron transfer. $O$ and $R$ denote oxidized and reduced states, respectively.

Three different regimes in Marcus theory: (a) the normal Marcus region, (b) the regime wherein $\Delta G^0 = \lambda$ giving rise to maximum rate of ET, and (c) the inverted Marcus region.

Pictorial representation of ET process. An electron is transferred from the donor (D) to the acceptor (A). The ground-state is shown in green and the CT-state is shown in red. In our study, the donor and the acceptor are the negatively charged chromophore and the tyrosine residue, respectively.

(a) Probability function and (b) the free energy curves for ground (PB) and CT (P$^+B^-$) states. $\Delta V$ represents the energy gap between two states along the trajectories.

Possible H-bonds around the chromophore in eYFPs without and with halide.

Possible H-bonds around the chromophore in eGFP.

Matrix representation of the GMH method.

Benzoquinone docked to eGFP, eYFP, and halide-bound eYFP. In the case of halide-bound eYFP, two docked BQ conformation were obtained, one close to Tyr145 and the other close to Tyr203. For the structure shown in panel (a), Chro-BQ distance is $\sim 6$ Å.

Benzoquinone docked to eGFP, eYFP, and halide-bound eYFP close to Tyr92.
2.11 Distance analysis between Tyr145 and BQ docked on the surface of eGFP along a 10 ns trajectory. Graphs show fluctuations in the relative distance along the trajectory (left) and the resulting distribution (right). The snapshots were taken every 2.5 ps.

2.12 Convergence of $\langle \Delta E'_{\text{cd ftci}} \rangle$ and $\langle \Delta E'_{\text{cd ftci}} \rangle_{\text{CT}}$ for eYFP without halide with Tyr145 as the intermediate acceptor.

2.13 Convergence of the Chro-Ty145 coupling in the ground state of eGFP.

2.14 Relevant tyrosine residues around the chromophore with the halide.

2.15 Residues included in QM in the calculations of the couplings between the chromophore and Tyr92, based on the Pathways model predictions.

3.1 Chromophores of the selected FPs of different colors: wt-GFP, eGFP (green), TagBFP (blue), EBFP(blue), CFP (cyan), YFP (yellow), DsRed, mCherry (red), mOrange (orange). Absorption/emission wavelengths are given in parenthesis. The chromophores are shown in colors corresponding to their fluorescence.

3.2 The effect of resonance stabilization of energetics of electron ejection from the neutral (left) and anionic (right) species. Since the resonance stabilization is always greater for charged species, more extensive resonance interactions lead to ionization energy decrease in the neutral species and to electron-detachment energy increase in anions.

3.3 Different protonation states of the GFP model chromophore.

3.4 The structures of the model chromophores (deprotonated forms) and atom labeling scheme. The chromophores consist of the green (phenol), pink (bridge), blue (imidazolinone) and red (acylimine) moieties. Panel (d) gives the atom labeling scheme: “p”, “b”, “i”, and “a” denote phenol, bridge, imidazolinone, and acylimine, respectively.

3.5 Leading resonance structures of the deprotonated model FP chromophores. Other resonance structures are shown in the section 3.7.

3.6 The HOMOs of the model chromophores.

3.7 Thermodynamic cycle.

3.8 Resonance structures of the deprotonated model GFP chromophore.

3.9 Resonance structures of the deprotonated model blue (TagBFP) chromophore.

3.10 Resonance structures of the deprotonated model RFP chromophore.

4.1 Structure of eGFP/eYFP. Left: $\beta$-barrel enclosing the chromophore. Right: eGFP and eYFP have the same anionic chromophore formed by cyclization and oxidation of the protein backbone at positions 65-67 (top). In YFPs, the chromophore is $\pi$-stacked with Tyr203 (bottom).
4.2 QM/MM schemes for eYFP (a) and eGFP (b) used in the calculations of the ionization energies of the chromophore. The black dotted lines denote the boundary between the QM (blue) and MM parts. The MM part in which point charges were set to zero is denoted by green and red (note that green atoms are part of the chromophore).

4.3 QM/MM scheme for eYFP and eGFP used in the calculations of electron attachment energies of tyrosines. The black dotted lines denote the boundary between the QM (blue) and MM parts. The MM part in which point charges were set to zero is denoted by green and red (note that green atoms are part of tyrosine).

4.4 eGFP and eYFP oxidative photoactivation. (a) The effect of potassium ferricyanide concentration on the main (green/yellow) fluorescent state bleaching (green full squares/yellow full triangles) and the red fluorescence increase (red open squares/magenta open triangles) in the oxidative redding of immobilized eGFP and eYFP. After one activating irradiation cycle with GFP filter set, remaining green fluorescence (normalized according to initial value) and originating red fluorescence (normalized according to maximal value) were measured and shown in the graph. (b) The red fluorescence appearance in eYFP during irradiation. Immobilized eYFP was irradiated with arc-lamp (GFP filter set, 0.6 W/cm²) in phosphate buffer (black squares), in the presence of oxidant (blue triangles), in the presence of sodium chloride (red circles), and in the presence of both oxidant and chloride (green triangles). Redding efficiency is normalized according to initial yellow fluorescence. Each data point is an average of three independent experiments. Error bars, s.d.

4.5 Mechanism of photoinduced ET in FPs. An oxidant molecule (represented by para-benzoquinone, BQ) docked to eGFP (left) and eYFP (right) and the relevant distances. The direct tunneling and two-step hopping (via Tyr145) mechanisms for ET are shown by dashed arrows.

4.6 Kinetic model of photoinduced ET via hopping mechanism. The excited state can decay to the ground state, either radiatively or non-radiatively. This channel is characterized by \( r_f \) which is inversely proportional to the excited-state lifetime \( (r_f \sim 10^9 \text{ s}^{-1}) \). Alternatively, the excited state can decay via ET from the chromophore to either Tyr145 or another acceptor, ResX (this could be Tyr203 in eYFP). ET to Tyr145 or ResX results in anion-radical (e.g., Tyr\(^{-}\)) formation that can lead to permanent bleaching \( (r_b) \). ET to Tyr145 can also lead to ET to an outside oxidant \( (r_2) \) forming a precursor of the red form. The observed bleaching is the sum of the yields of the red form precursor and of permanently bleached states. Based on our rates calculations, \( r_3 \) and \( r_6 \) are slow; \( r_5 \) is slow for Tyr203.
4.7 Relevant structural parameters. The distance between the phenolic oxygens of the chromophore and Tyr145 \((d_1)\) affects the main ET channel \((r_1)\). The extent of \(\pi\)-stacking can be quantified by \(D \equiv \frac{d_2+d_3}{2}\) and \(\Delta \equiv |d_2 - d_3|\).

4.8 Bleaching and redding kinetics in the eGFP and eYFP mutants. (a-d) Photoconversion of immobilized proteins \textit{in vitro} in PBS in the presence of 0.5 mM potassium ferricyanide. (PBS contains potassium chloride.) Graphs show the main form bleaching (a,c) and simultaneous appearance of red fluorescence (b, d) in eGFP, eGFP-Y145L, eGFP-Y145F (a, b), and eYFP, eYFP-Y145L, eYFP-Y145F (c, d). Green/yellow and red fluorescence intensities were background subtracted and normalized to the maximum values. Standard deviation values \((n = 15-20\) measurements in a representative experiment out of five independent experiments) are shown. (e) Bleaching of eGFP, eGFP-Y145L, eYFP, and eYFP-Y145L in live HEK293 cells induced by 488 nm laser in a confocal microscope. (f) Increase of photostability \((\text{time to half-bleaching})\) of the eGFP-Y145L and eYFP-Y145L mutants compared to eGFP and eYFP, respectively, under confocal and widefield microscopy of live HEK293 cells. Standard deviation values for 50-60 cells in three independent experiments are shown.

4.9 BQ docked in the vicinity of residue 145 in the eGFP-Y145L mutant.

4.10 Rate for the direct Chro-BQ ET using the following parameters: \(\Delta G=-2.189\) eV, \(\lambda=1.20\) eV, \(|H_{DA}|^2=2.6 \times 10^{-8}\) eV^2.

4.11 Two snapshots along the eYFP equilibrium trajectory illustrating two interconverting hydrogen-bond patterns. In the dominant conformation \((\text{left})\), the chromophore forms a hydrogen bond with Tyr145, similar to the eGFP. In the second conformation, Tyr145 flips and forms a hydrogen bond with the His169.

4.12 Left: Distance between the chromophore and Tyr145 along a 12 ns trajectory. Right: Chro-Tyr145 distance distribution.

5.1 Structure of horse myoglobin, with relevant residues highlighted with spherical atom representation.

5.2 Relative orientation of decaheme unit found in MtrF.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADE</td>
<td>adiabatic detachment energy</td>
</tr>
<tr>
<td>AIE</td>
<td>adiabatic ionization energy</td>
</tr>
<tr>
<td>BQ</td>
<td>para-benzoquinone</td>
</tr>
<tr>
<td>CBS</td>
<td>complete basis set</td>
</tr>
<tr>
<td>CT</td>
<td>charge transfer</td>
</tr>
<tr>
<td>CDFT</td>
<td>constrained density functional theory</td>
</tr>
<tr>
<td>DE</td>
<td>detachment energy</td>
</tr>
<tr>
<td>EA</td>
<td>electron attachment</td>
</tr>
<tr>
<td>ET</td>
<td>electron transfer</td>
</tr>
<tr>
<td>ESPT</td>
<td>excited-state proton transfer</td>
</tr>
<tr>
<td>eGFP</td>
<td>enhanced green fluorescent protein</td>
</tr>
<tr>
<td>eYFP</td>
<td>enhanced yellow fluorescent protein</td>
</tr>
<tr>
<td>FCD</td>
<td>fragment-charge difference</td>
</tr>
<tr>
<td>FDFT</td>
<td>frozen density functional theory</td>
</tr>
<tr>
<td>FP</td>
<td>fluorescent protein</td>
</tr>
</tbody>
</table>

xvi
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>FqRET</td>
<td>fluorescence quenching resonance energy transfer</td>
</tr>
<tr>
<td>FRET</td>
<td>fluorescence resonance energy transfer</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>GMH</td>
<td>generalized Mulliken-Hush</td>
</tr>
<tr>
<td>HOMO</td>
<td>highest-occupied molecular orbital</td>
</tr>
<tr>
<td>IE</td>
<td>ionization energy</td>
</tr>
<tr>
<td>ISC</td>
<td>inter-system crossing</td>
</tr>
<tr>
<td>KMC</td>
<td>kinetic Monte Carlo</td>
</tr>
<tr>
<td>LRA</td>
<td>linear response approximation</td>
</tr>
<tr>
<td>LSS</td>
<td>large Stokes shift</td>
</tr>
<tr>
<td>LSS-FP</td>
<td>large Stokes shift fluorescent protein</td>
</tr>
<tr>
<td>LUMO</td>
<td>lowest-unoccupied molecular orbital</td>
</tr>
<tr>
<td>MD</td>
<td>molecular dynamics</td>
</tr>
<tr>
<td>MO</td>
<td>molecular orbital</td>
</tr>
<tr>
<td>mVDE</td>
<td>modified vertical detachment energy</td>
</tr>
<tr>
<td>mVEA</td>
<td>modified vertical electron affinity</td>
</tr>
<tr>
<td>NBO</td>
<td>natural bond orbital</td>
</tr>
<tr>
<td>PC-FP</td>
<td>photoconvertible fluorescent protein</td>
</tr>
<tr>
<td>PES</td>
<td>potential energy surface</td>
</tr>
<tr>
<td>PA-FP</td>
<td>photoactivatable fluorescent protein</td>
</tr>
<tr>
<td>PS-FP</td>
<td>photoswitchable fluorescent protein</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>QM/MM</td>
<td>quantum mechanics/molecular mechanics</td>
</tr>
<tr>
<td>RS-FP</td>
<td>reversibly switchable fluorescent protein</td>
</tr>
<tr>
<td>ROS</td>
<td>reactive oxygen species</td>
</tr>
<tr>
<td>SHE</td>
<td>standard hydrogen electrode</td>
</tr>
<tr>
<td>TS</td>
<td>transition state</td>
</tr>
<tr>
<td>VDE</td>
<td>vertical detachment energy</td>
</tr>
<tr>
<td>VEA</td>
<td>vertical electron attachment</td>
</tr>
<tr>
<td>YFP</td>
<td>yellow fluorescent protein</td>
</tr>
<tr>
<td>ZPE</td>
<td>zero point energy</td>
</tr>
</tbody>
</table>
Abstract

GFP-like fluorescent proteins occupy a unique niche in modern science. They are the only fluorescent probes of natural origin. Their properties are of interest from both fundamental and applied points of view. Being a simple single-protein system with clear absorption and fluorescence readouts, fluorescent proteins are a useful vehicle for studying the mechanistic details of these processes in proteins, both in vitro and in cellulo. Electron transfer is of fundamental importance in biology, the respiration process of a living organism being just one example. Photoinduced redox reactions play an important role in the photocycle of fluorescent proteins from the green fluorescent protein (GFP) family. GFP-like proteins are widely used for in vivo imaging purposes and there is abundance of oxidizing and reducing agents in those environments. Yet, the redox properties of the fluorescent proteins and the consequences of photoinduced electron transfer in GFP is not fully explored.

In chapter 2, we present the theoretical methods for modeling redox and ET processes and discuss requisite computational tools. Calculation of the redox potential of GFP chromophore utilizes a thermodynamic cycle (in implicit solvent) and linear response approximation (LRA) in protein environment. The ET process includes the transfer of an electron to an external acceptor, via hopping mechanism or direct tunneling mechanism. We used Pathways model to locate possible intermediate electron
acceptors in FPs. In addition to free energy of reaction, computation of the ET rates requires two additional parameters, reorganization energy and electronic coupling. We compute the reorganization energy and coupling using LRA and constrained DFT-CI (CDFT-CI), respectively. LRA approach requires the energy of CT state relative to the ground state, which we calculate by CDFT-CI with the ωB97X-D functional. A protocol for extrapolating CDFT-CI energies computed with finite basis set to the complete basis set limit is also presented here.

In chapter 3, the redox properties of model chromophores from the green fluorescent protein family are characterized computationally using density functional theory with a long-range corrected functional, the equation-of-motion coupled-cluster method and implicit solvation models. The analysis of electron-donating abilities of the chromophores reveals an intricate interplay between the size of the chromophore, conjugation, resonance stability, presence of heteroatoms, and solvent effects. Our best estimates of the gas-phase vertical/adiabatic detachment energies of the deprotonated (i.e., anionic) model red, green, and blue chromophores are 3.27/3.15 eV, 2.79/2.67 eV, and 2.75/2.35 eV, respectively. Vertical/adiabatic ionization energies of the respective neutral species are 7.64/7.35 eV, 7.38/7.15 eV, and 7.70/7.32 eV. The standard reduction potentials ($E_{\text{red}}^0$) of the anionic (Chr/Chr$^-$) and neutral (Chr$^+$/Chr) model chromophores in acetonitrile are: 0.38/1.44 V (red), 0.22/1.26 V (green), and -0.12/1.05 V (blue) suggesting, counter-intuitively, that the red chromophore is more difficult to oxidize than the green and blue ones (in either neutral or deprotonated forms). The respective redox potentials in water follow a similar trend, but are more positive than the acetonitrile values.
In chapter 4, we discuss photoinduced electron transfer in fluorescent proteins from the GFP-family. This process can be regarded either as an asset facilitating new applications or as a nuisance leading to the loss of optical output. Photooxidation commonly results in green-to-red photoconversion called oxidative redding. It was discovered that yellow FPs do not undergo redding, but the redding is restored upon halide binding. Our calculations of the energetics of one-electron oxidation and possible ET pathways suggested that excited-state ET proceeds through a hopping mechanism via Tyr145. In YFPs, the \( \pi \)-stacking of the chromophore with Tyr203 reduces its electron donating ability, which can be restored by the halide binding. Point mutations confirmed that Tyr145 is a key residue controlling ET. Substitution of Tyr145 by less efficient electron acceptors resulted in highly photostable mutants. This strategy — calculation and disruption of ET pathways by mutations — may represent a new approach towards enhancing photostability of FPs.
Chapter 1: Introduction and overview

1.1 Green Fluorescent Proteins

The unique properties of green fluorescent proteins (GFPs) have revolutionized many areas in the life sciences\textsuperscript{1–5} by enabling \textit{in vivo} observations of protein localization and interactions, intracellular measurements of concentrations of physiologically important ions (Ca\textsuperscript{2+}, Cl\textsuperscript{−}, H\textsuperscript{+}), mapping gene expressions, etc. The importance of fluorescent proteins and related technologies was recognized with the 2008 Nobel Prize in chemistry. The 2014 Nobel Prize, conferred “for the development of super-resolved fluorescence microscopy”, is another testament to the significance of fluorescent proteins, and particularly, of their photophysical properties.

GFP was first characterized\textsuperscript{6} at the protein level in extracts from the jellyfish \textit{Aequorea victoria} in 1962. It then took more than 30 years to clone the GFP gene and demonstrate that functional GFP can be expressed in various model organisms.\textsuperscript{7,8} This discovery opened the era of GFP applications as a fluorescent label fully encoded by a single gene. In addition to their role in biotechnology applications, fluorescent proteins are interesting for their own sake. In particular, natural diversity and functioning of fluorescent proteins represent intriguing fundamental problems. So far, GFP-like proteins have been found only in multicellular animal species (Metazoa kingdom), specifically in hydroid jellyfishes and coral polyps (phylum Cnidaria), combjellies (Ctenophora),
crustaceans (Arthropoda), and lancelets (Chordata). Together with the observation that most sequenced animal genomes contain no GFP-related sequences, this suggests that the GFP gene originated very early in animal evolution but then was lost in many species. Natural GFP-like proteins demonstrate a broad spectral diversity including cyan, green, yellow, orange, and red fluorescent proteins as well as a colorful palette of non-fluorescent chromoproteins. Phylogenetic analysis and reconstruction of ancestral genes have shown that the green fluorescent phenotype (eGFP-like excitation and emission spectra) was likely characteristic of evolutionary ancient proteins, whereas other colors appeared later in evolution, independently in different taxa.

The biological functions of GFP-like proteins have been studied only sparcely and for many species remain unclear or, at least, not experimentally proven. One well-studied example is the participation of GFPs in bioluminescent systems, where they act as secondary emitters. Yet, most bioluminescent species contain no fluorescent protein, and, conversely, most fluorescent protein-containing animals are non-bioluminescent. Thus, fluorescent proteins appear to have other functions. For example, it has been proposed that fluorescent proteins play a photoprotective role in corals. A recent elegant study demonstrated that green fluorescent spots on jellyfish tentacles efficiently attract a prey. This observation explains the predominant distribution of fluorescent proteins at the tentacles and around the mouth of jellyfishes and coral polyps. An association of fluorescent proteins with a physiological state of coral larvae have been demonstrated, but possible molecular mechanisms of this phenomenon are unclear. It is reasonable to hypothesize that, at the time of their early evolution, fluorescent proteins had some basic functions not related to their visual appearance (bioluminescence, camouflage, attraction, recognition, etc.) as no organisms had eyes at that time.
Such primary biochemical functions could have be photoprotection, production or scavenging of reactive oxygen species (ROS), or light-induced electron or proton transfer. While direct observation of evolutionary ancient fluorescent protein functions is impossible, detailed studies of photophysics and photochemistry of GFP-like proteins might provide clues to the biological functioning of this protein family.

Not surprisingly, the photophysics of fluorescent proteins has motivated numerous experimental and theoretical studies.\textsuperscript{14–24} Owing to the complexity of the system, many aspects of the fluorescent protein photocycle and chromophore formation are still largely unexplored. Yet, the molecular-level understanding of these processes provides a crucial advantage in the design of new fluorescent proteins with properties to fit particular applications. While investigation of some properties (colors, Stokes shifts, brightness) is relatively straightforward, understanding the role of others (photostability, phototoxicity) and their optimal parameter space are more subtle.

Absorption and fluorescence wavelengths are among the key parameters that can be modified. Fluorescent proteins of different colors can be used to mark different proteins (multi-color imaging) and to construct FRET (fluorescence resonance energy transfer) pairs. Variations in Stokes shifts enable single-laser dual-emission type of measurements. Red fluorescent proteins are of a particular importance as suitable markers for deep-tissue imaging.\textsuperscript{25} Non-fluorescent chromoproteins can be used as efficient FRET acceptors, e.g., in FqRET (fluorescence quenching resonance energy transfer) imaging,\textsuperscript{26, 27} and for photoacoustic imaging in tissues.\textsuperscript{28} Today, fluorescent proteins span the entire range of the visible spectrum including the far-red end of the spectrum.\textsuperscript{2, 14, 21, 25, 29–32} As illustrated in Figure 1.1, color tuning in fluorescent proteins can be achieved by several distinct mechanisms, including varying the length of the $\pi$-conjugated system, changing the protonation state of the chromophore, $\pi$-stacking,
electrostatic and other specific interactions with nearby residues. Brightness is another obviously important factor: brighter fluorescent proteins, i.e., those with larger extinction coefficients and fluorescence quantum yields, make better fluorescent labels. Other properties, such as photostability, phototoxicity, sensitivity to the presence of small molecules, ions, and reducing or oxidizing agents, are very important, but these have specific functions that are only suitable for certain applications. In other words, what is optimal for one application can be undesirable in others.

Consider, for example, photostability. In many applications, bleaching, a gradual loss of optical output upon repeated irradiation, is undesirable. Consequently, protein engineering often aims at more photostable fluorescent proteins. On the other hand, bleaching is exploited in super-resolution imaging.\textsuperscript{4, 34–38} Methods based on fluorescence loss and recovery are used to trace protein dynamics; photoconversions and photoswitching enable optical highlighting and timing of biochemical processes.\textsuperscript{23, 25, 32} In a similar vein, phototoxicity, which is undesirable for \textit{in vivo} imaging applications, can be exploited in photodynamic therapies and targeted protein/cell inactivation.\textsuperscript{39} Likewise, the sensitivity of fluorescence to other chemical species may be regarded as a nuisance interfering with imaging or as an asset enabling new types of measurements and biosensing applications. For example, sensitivity of YFPs’ fluorescence to halides limits their use as general-purpose yellow fluorescent tags, but can be exploited in ratiometric measurements of halide concentrations. The same duality is engendered by photoconversion and photoswitching, phenomena entailing changes in fluorescence properties upon irradiation. For example, photooxidative redding,\textsuperscript{40} photoconversion leading to a red-shifted absorption/emission, may be exploited in applications\textsuperscript{4, 25, 41} such as timing biochemical processes, optical highlighting, or intracellular redox measurements; yet, it interferes with standard imaging measurements in live cells, which always contain
Figure 1.1: Color tuning in fluorescent proteins: Different chemical structures of the chromophore lead to different colors. Main types of chromophore structures are shown together with corresponding excitation (upper bar) and emission (bottom bar) wavelengths designated by arrows. The size of $\pi$-conjugated system is particularly important for determining the color: more extensive conjugation leads to red-shifted absorption (compare, for example, blue, green, and red chromophores). Changes in protonation states of the chromophore also affect the energy gap between the ground and excited states. Excited-state deprotonation of the chromophore is one of the mechanisms of achieving large Stokes shifts. The absorption/emission can be red-shifted by $\pi$-stacking of the chromophore with other aromatic groups (e.g., tyrosine), as in YFP (not shown). Specific interactions with nearby residues also affect the hue (for example, additional red shift in mPlum fluorescence is attributed to a hydrogen bond formed by acylimine’s oxygen). Reproduced from Ref. 33.
copious amounts of oxidizing and reducing agents. In single-molecule visualization applications, properties such as blinking frequency and photon budget need to be considered.\textsuperscript{37,42–45}

Owing to their rich photophysics and photochemistry, fluorescent proteins feature a wide array of tunable properties. Our ability to manipulate these properties is critical for designing fluorescent proteins optimal for specific applications. Knowledge of structure-function relationship and detailed molecular-level mechanistic understanding of the fluorescent proteins’ photocycle are essential prerequisites for controlling these properties.

On a fundamental level, the same molecular-level processes that operate in fluorescent proteins are encountered in other systems of technological and biological significance. For example, natural and artificial light harvesting involves photoexcitation, energy transfer (either coherent or via FRET) between multiple chromophores, and generation and transport of photoelectrons. Photocatalysis and production of solar fuels is based on photochemical transformations. Light sensing in many biological systems is initiated by photoinduced cis-trans isomerization coupled with excited-state proton transfer (ESPT). Thus, understanding fundamental aspects of fluorescent proteins’ photophysics will aid our progress in other areas.

Various aspects of FPs have been extensively reviewed.\textsuperscript{1,2,4,14–24,31,32,37} Studies prior to 2009 have been comprehensively reviewed in a topical issue of \textit{Chemical Society Reviews}.\textsuperscript{3,15–17,20} Transient dark states, their possible structure and connection to protonation equilibria, and the implication for single-molecule studies have been discussed in Ref. 20. Mechanistic details of ESPT have received a considerable attention.\textsuperscript{15,17} Photoconvertible and photoswitchable fluorescent proteins and their applications have been
The uses of fluorescent proteins in super-resolution imaging have been reviewed in Refs. 46 and 47. We have recently reviewed various excited state processes in FP, with an emphasis on the mechanistic details of those processes.

1.2 Fluorescent protein photocycle

Figure 1.2: A typical structure of a fluorescent protein represented by eGFP. In all fluorescent proteins, the chromophore, which is formed autocatalytically upon protein folding, is buried inside a tight 11-stranded \( \beta \)-barrel comprising 220-240 amino-acids. The approximate molar weight is 25 to 30 kDa. The diameter of the barrel is \( \sim 24 \) Å and its height is \( \sim 42 \) Å. Reproduced from Ref. 33.

The photophysics and photochemistry of fluorescent proteins bear considerable resemblance to those of synthetic dyes.\(^ {35,48}\) From the chemical point of view, typical fluorescent protein chromophores (Fig. 1.1) are similar to cyanine dyes, owing to their common structural feature: a methyne bridge connecting conjugated aromatic moieties. However, the presence of the protein barrel (Fig. 4.1) leads to significant differences.
Figure 1.3: Excited-state processes in fluorescent proteins. The main relaxation channel is fluorescence. Radiationless relaxation, a process in which the chromophore relaxes to the ground state by dissipating electronic energy into heat, reduces quantum yield of fluorescence. Other competing processes, such as transition to a triplet state via inter-system crossing (not shown), excited-state chemistry and electron transfer, alter the chemical identity of the chromophore thus leading to temporary or permanent loss of fluorescence (blinking and bleaching) or changing its color (photoconversion). Reproduced from Ref. 33.

The rigid protein environment restricts the chromophore’s range of motion and limits its accessibility to the solvent and other species present in solution (ambient oxygen, salt ions, oxidating and reducing agents, etc). Indeed, photophysical properties of the model chromophores in solutions differ strikingly from those of the respective parent fluorescent proteins.\(^1\)\(^,\)\(^15\) the solvated chromophores do not fluoresce, they often have different colors, and they are more efficient photosensitzers.

Figure 1.3 outlines various excited-state processes in fluorescent proteins. The photocycle is initiated by light absorption producing an initial electronically excited state of the chromophore. The main relaxation channel restoring the ground-state chromophore
is fluorescence. The color of the emitted light may differ from the absorbed light due to a structural relaxation of the chromophore, its hydrogen-bond network, or ESPT. Alternatively, the chromophore may return to the ground state by dissipating the electronic energy into nuclear motions, via radiationless relaxation. Such thermal relaxation fully dominates in GFP-like chromoproteins, which have extremely low fluorescence quantum yield ($10^{-4}$-$10^{-5}$). Since the bonding pattern in the excited states is different, electronic excitation can initiate various chemical transformations of the chromophore, such as isomerization, making or breaking covalent bonds, photooxidation/photoreduction, or reactions with nearby residues or small molecules (e.g., ambient oxygen). Changes in bonding pattern upon excitation also affect the acidity of the chromophore, which is a driving force for ESPT. These processes alter optical properties leading to the formation of transient dark or permanently bleached states as well as changing the color of the absorption/fluorescence. Thus, the yields of bleaching and blinking, photostability, phototoxicity, photoswitching and photoconversion phenomena are determined by the competition between the main relaxation channels (fluorescence and radiationless relaxation) and various photoinduced transformations. The timescales of different channels are crucially important for understanding the branching ratios and yields. A finite excited-state lifetime limits the scope of excited-state processes. Typically, for fluorescent proteins, the excited-state lifetimes are 1-10 ns. Thus, in order to have a noticeable effect on the photocycle, an excited-state process should be initiated on a timescale comparable with that the excited-state lifetime. Below we briefly review typical lifetimes and yields of these excited-state processes.

Not surprisingly, the dominant excited-state process in fluorescent proteins is fluorescence; its quantum yield ($Y_f$) is high, e.g., 0.6 in eGFP and eYFP. Interestingly,
$Y_f$ of model fluorescent protein chromophores in solutions are 3-4 orders of magnitude lower than in the protein environment; this phenomenon has been attributed to the increased flexibility of the bare chromophore and its interactions with solvent molecules.\textsuperscript{15, 50–52}

The dominant process leading to the loss of fluorescence is radiationless relaxation. In contrast to bleaching, this is a relatively benign process since it simply restores the ground-state chromophore (although, long-lived dark states can also be formed via radiationless relaxation). The upper limit for this channel is given by $1-Y_f$.

### 1.3 Photoinduced transformations in fluorescent proteins

Figure 1.4 summarizes various types of light-induced changes in optical properties, which are exploited in applications. When fluorescent proteins are used as simple fluorescent tags, light is used to excite them and then fluorescence is recorded. The difference between the absorption and emission wavelength is called Stokes shift. Combining fluorescent proteins with large and small Stokes shifts enables multicolor applications in which only one laser is required (single-excitation/dual-emission mode). These practical considerations motivated the development of fluorescent proteins with large Stokes shifts (LSS).\textsuperscript{53–56} Large Stokes shifts are also desirable in FRET applications: in FRET acceptors, they improve the spectral gap between the donor’s and the acceptor’s emission, whereas large Stokes shifts in FRET donors reduce the direct excitation of the acceptor.
The ability to use light to modify optical properties of fluorescent proteins has greatly expanded their usage.\textsuperscript{22–24, 31, 32} Light can be used to selectively activate or deactivate fluorescent proteins. In some fluorescent proteins this can be done in a reversible fashion. Photoactivation (PA) entails the conversion of a dark, non-fluorescent form of the protein into a bright one. Using light to switch between dark and bright forms is called photoswitching (proteins that are dark in their most stable state are called positive photoswitchers, in contrast to negative photoswitchers, which are naturally bright and can be switched into a long-lived dark state). Some fluorescent proteins permit photoconversion (PC) rather than just photoactivation or photoswitching. These fluorescent proteins switch between two colors (e.g., from green to red), both of which can be visualized. Photoswitchable and photoconvertible fluorescent proteins provide a basis for many super-resolution techniques.\textsuperscript{37, 38} Some phototransformations can be reversible,
giving rise to the reversibly switchable fluorescent proteins (RS-FPs); in these, the fluo-
rescent and non-fluorescent states are inter-convertible by photo-excitation of each form
using light of a specific wavelength. RS-FPs may be used in monochromatic multi-label
imaging and dual color fluorescence nanoscopy as well as in optical memory and opti-
tical switches.

Our state of knowledge on phototransformations in fluorescent proteins is rapidly
evolving. For a long time, photoconversions were perceived as an unusual property of
a few outliers from the large fluorescent protein family. The ability to undergo pho-
toconversions was attributed to a specific amino-acid environment conductive of in-
tramolecular reactions involving the chromophore and leading to its chemical modifi-
cation. This paradigm substantially shifted in 2009, when several new photoconver-
sions were described. One of them is the so-called photooxidative redding (green-to-red
photoconversion in the presence of oxidants, which occurs in many fluorescent pro-
teins with tyrosine-based chromophores and appears to be relatively insensitive to the
chromophore’s environment. Subsequent studies provided additional examples of the
ubiquity of photoconversion phenomena. Screening of the photobehavior of 12 differ-
et orange and red fluorescent proteins led to the discovery of novel red-to-green and
orange-to-far-red conversions. In cellulo red-to-green photoconversion of Katushka,
mKate, and HcRed1 was observed both in single- and two-photon excitation regimes; it
can be induced by irradiation ranging from 3.06 to 2.21 eV (405 to 561 nm). Orange
fluorescent proteins, mOrange1 and mOrange2, photoconvert to far-red forms emitting
at 1.94 eV (640 nm) upon excitation by blue lasers; it was shown that these photocon-
versions proceed via multi-photon processes. Thus, the above examples of oxidative
redding in GFPs and orange fluorescent proteins as well as greening of red fluorescent
proteins illustrate that photoconversions are rather common among spectrally diverse fluorescent proteins.

The mechanisms and structural motifs of photoactivation, photoconversions, and photoswitching include cis-trans isomerization (often coupled with changes in protonation state), oxidation/reduction of the chromophore, and chemical changes involving the breaking of covalent bonds. In PA-GFP, photoactivation is achieved by changing the chromophore’s environment (by decarboxylation of the nearby glutamine residue), which shifts the equilibrium between the two different protonation states of the chromophore. In Kaede, Dendra, and EosFP, the change in color results from the photoinduced chemical modification of the chromophore (extension of the $\pi$-system and breaking the backbone of the protein). In Dronpa, the switching between the dark and bright states involves cis-trans isomerization coupled with changes in protonation states (a similar mechanism likely operates in Padron and KFP). In Dreiklang, the switching is based on reversible photoinduced hydration/dehydration of the imidazolone ring of the chromophore. Dreiklang is the only reversible photoswitchable protein which entails a chemical change of the chromophore (and thereby changing the spectroscopic properties) in the photoswitching process.

### 1.4 Photoinduced electron transfer: A gateway step leading to multiple outcomes

Photoinduced ET to/from the chromophore can lead to a variety of outcomes. Although this is well known in dyes, it was not explored in GFPs. Photoinduced redox properties of fluorescent proteins came into the spotlight in 2009, when it was discovered that fluorescent proteins can be efficient light-induced electron donors. Bogdanov et al.
observed that many fluorescent proteins with an anionic GFP chromophore (such as one in eGFP, see Fig. 4.1) undergo photoconversion from green to red form upon irradiation in the presence of oxidants. This oxidative redding results from a series of chemical steps initiated by photooxidation, or more specifically ET from the electronically excited chromophore to an external oxidant molecule. The structure of the red form ultimately formed by the photooxidation is still unknown. Note that photoinitiated redox processes in fluorescent proteins are not specific to photooxidation of the chromophore. Photoreduction is also possible. For example, photoinduced ET from a nearby Glu to the chromophore is believed to be a gateway step leading to decarboxylation. Recently, photoreduction of the chromophore was invoked to explain the formation of long-lived red-shifted transient species in red fluorescent proteins. Photoreduction may also play a role in anaerobic redding or in greening of red fluorescent proteins. Photoinduced ET from the anionic chromophores to O$_2$ may lead to superoxide formation, which might be responsible for phototoxicity.

In short, there is a growing body of evidence of the importance of photoinduced ET in fluorescent proteins. Different types of ET may be operational, such as ET to and from the chromophore producing reduced or oxidized species. Furthermore, the redox partners of the chromophore may be different: ET may entail a nearby residue, such as glutamine as a donor or tyrosine as an acceptor, or an oxidant molecule (e.g., O$_2$).

ET can proceed by different mechanisms summarized in Figure 1.5. One possibility is ET from the electronically excited chromophore via the Marcus mechanism, which may involve the direct ET to an oxidant molecule, or a multi-step hopping process via intermediate electron acceptors. In the strong coupling regime, ET can proceed by adiabatic evolution of the initially excited state. Alternatively, the charge-transfer (CT) states can be populated directly, by photoexcitation or via radiationless relaxation.
from higher excited states (especially at high-intensity conditions when multi-photon processes become operational).

Figure 1.5: Different mechanisms for ET. The relevant states are the bright excited state ($S_1$) and the charge-transfer (CT) state. In photooxidation, the latter is of $D^+A^-$ character (or $DA^-$, depending on the protonation state of the chromophore). Top left: ET between the donor and acceptor by the Marcus mechanism. Top right: Adiabatic evolution of the initially excited state leading to CT via a barrier. Bottom left: CT state accessed by radiationless relaxation from a higher excited state. Bottom right: ET via direct one- or multi-photon excitation of the excited state of CT character. Reproduced from Ref. 33.

As summarized in Figure 1.5, ET can proceed by different mechanisms. In particular, CT states can be accessed by direct photoexcitation of the chromophore, or by radiationless relaxation from higher excited states (this channel might be very important in multi-photon regime). The CT states of various nature have been implicated in decarboxylation$^{67, 68, 76, 77}$ and in bleaching mechanisms.$^{78}$

It was proposed$^{67}$ that decarboxylation proceeds via ET from Glu222 to the electronically excited chromophore (photoreduction) by a Kolbe-like mechanism.
Subsequent electronic structure calculations\textsuperscript{68, 77} identified such CT states for the neutral (protonated) GFP chromophore; these are located around 4-6 eV vertically (Fig. 1.6). Grigorenko \textit{et al.} have proposed that these states are accessed either directly, by UV or multi-photon excitation of the chromophore, or via radiationless relaxation from a high-lying locally excited state.\textsuperscript{68} Morokuma and co-workers have put forward\textsuperscript{77} an alternative mechanism via adiabatic evolution of the initially excited state (such as one Fig. 1.5: Top right), whereas van Thor and Sage have considered\textsuperscript{79} a Marcus-like process (Fig. 1.5: Top left).

Another important CT process, Chro$^-$ $\rightarrow$ O$_2$, has been characterized computationally in Ref. 78. Fig. 1.7 shows relevant MOs and energetics of the CT and locally excited states. A mechanism of irreversible bleaching via such states has been proposed.\textsuperscript{78} The calculations showed\textsuperscript{78} that: (i) these CT states are accessible by photoexcitation; (ii) once reaching the CT state, the system can undergo series of low-barrier transformations leading to the chromophore destruction.
This thesis is organized as follows. In chapter 2, we discuss the methodology used in the rest of the thesis, including the background and benchmarking results. This section also provides a detailed derivation of the protocol extrapolating $\Delta E$ to the complete basis set limit and the kinetic schemes used in chapter 4. Detailed analysis of the search for possible acceptors present in fluorescent proteins is also given here. We found two possible acceptors, Tyr 145 and Tyr203. In chapter 3, we present spectroscopic properties (such as ionization/detachment energy) and redox properties (reduction potential) of model FP chromophores. We focus on three different model chromophores (blue, green and red), which differ by the extent of $\pi$-conjugation. We find that a delicate balance between conjugation, resonance stabilization, and solvent effects controls the redox properties of the model chromophores. In chapter 4, we report redox properties of isolated and protein-bound chromophores of green and yellow fluorescent proteins. We discovered dominant ET pathways, active in FPs, which controls the photostability of these FPs. We identify key residues that participates in the ET pathways, such

---

**Figure 1.7:** Relevant MOs and leading electronic configurations of the locally excited chromophore and the CT states of Chro$^{-}$ $\rightarrow$ O$_2$ character. Reproduced from Ref. 78.
as Tyr145 and Tyr203. We also present mutagenesis study which demonstrates that by mutating Tyr145 leads to more photostable FP.
Chapter 1 References


[27] Pettikiriarachchi, A.; Gong, L.; Perugini, M.A.; Devenish, R.J.; Prescott, M. Ultra-
marine, a chromoprotein acceptor for Förster resonance energy transfer *PLos One* 2012, 7, e41028.

[28] Li, Y.; Forbrich, A.; Wu, J.; Shao, P.; Campbell, R.E.; Zemp, R. Engineering
dark chromoprotein reporters for photoacoustic microscopy and FRET imaging

[29] Wang, L.; Jackson, W.C.; Steinbach, P.A.; Tsien, R.Y. Evolution of new nonanti-
body proteins via iterative somatic hypermutation *Proc. Nat. Acad. Sci.* 2004, 101,
16745–16749.

[30] Subach, O.M.; Patterson, G.H.; Ting, L.-M.; Wang, Y.; Condeelis, J.S.; Verkhusha,
V.V. A photoswitchable orange-to-far-red fluorescent protein, PSmOrange *Nat.
Methods* 2011, 8, 771–777.


[32] Adam, V.; Berardozzi, R.; Byrdin, M.; Bourgeois, D. Phototransformable fluores-

[33] Acharya, A.; Bogdanov, A.M.; Bravaya, K.B.; Grigorenko, B.L.; Nemukhin, A.V.;

[34] Hofmann, M.; Eggeling, C.; Jakobs, S.; Hell, S.W. Breaking the diffraction barrier
in fluorescence microscopy at low light intensities by using reversibly photoswitch-

[35] Ha, T.; Tinnefeld, P. Photophysics of fluorescent probes for single-molecule bio-
617.

[36] Burnette, D.T.; Sengupta, P.; Dai, Y.; Lippincott-Schwartz, J.; Kachar, B. Bleach-
ing/blinking assisted localization microscopy for superresolution imaging using

[37] Nienhaus, K.; Nienhaus, G.U. Fluorescent proteins for live-cell imaging with

[38] Sengupta, P.; van Engelenburg, S.B.; Lippincott-Schwartz, J. Superresolution
imaging of biological systems using photoactivated localization microscopy *Chem.


Chapter 2: Methodology

2.1 Marcus theory of electron transfer

The yields and rates of chemical reactions are determined by the potential energy surface (PES). PES is a multidimensional surface that represents the energy of the system with respect to nuclear coordinates. Chemical reactions most often involve breaking and making of chemical bonds, and the reactants have to pass through a transition state (TS), which is higher in energy than the reactants, to form products. The rate constant, $k$, of a chemical reaction can be obtained from the TS theory.

$$k = A \exp \left( \frac{-\Delta G^\ddagger}{k_B T} \right),$$

(2.1)

where $\Delta G^\ddagger$ is activation free energy. Contrary to a chemical reaction, an ET process does not include making or breaking of chemical bonds. There may be instances where ET is followed by (or occurs after) the making or breaking of a chemical bond, but here we are concerned with the ET process itself. Prior to 1956, before Marcus theory was developed, Libby invoked the Franck-Condon principle to treat ET. He considered a vertical excitation from the ground state of the reactant to the product. In this picture, the associated ET from donor to acceptor is so fast that the nuclei are assumed to be fixed on the timescale of the ET process. But this picture violates the basic law of energy conservation and is only possible with absorption of a photon. This drawback in
Figure 2.1: Difference between Libby’s theory and Marcus theory of treating electron transfer. R and P represent reactants and products states.

Libby’s theory was the motivation behind Marcus theory, as described by Prof. Rudolph A. Marcus in his Noble lecture in 1992:

“After studying Libby’s paper and the symposium discussion, I realized that what troubled me in this picture for reactions occurring in the dark was that energy was not conserved: the ions would be formed in the wrong high-energy environment, but the only way such a non-energy-conserving event could happen would be by the absorption of light (a “vertical transition”), and not in the dark.”

Prof. Marcus introduced a reorganization energy ($\lambda$) term, which arises mostly due to solvent/environment fluctuation, such that both the Franck-Condon principle and energy conservation law are satisfied. The reaction coordinate (x-axis in Figure 2.1) in the Marcus theory is a collective solvent coordinate when only two atomic ions (which have no inner-sphere relaxation) participate in the ET process. In a more general case, when there are molecules involved in the redox/ET process, even in non-polar solvents $\lambda \neq 0$ due to vibrational relaxation of the molecule itself. The total reorganization energy can be written as a sum of the two terms:

$$\lambda = \lambda_i + \lambda_O,$$  \hspace{1cm} (2.2)
where $\lambda_O$ and $\lambda_i$ denote the reorganization energy due to solvent fluctuation and molecular vibrational relaxation, respectively. The $\lambda_i$ is typically small relative to $\lambda_O$, so here we focus on the $\lambda_O$ term. Marcus theory states that in order for the ET to happen, the solvent fluctuation must bring the initially equilibrated reactant state to an intermediate configuration, a configuration that can also be attained by solvent fluctuation of the product state. At this intermediate solvent configuration, the ET process occurs, followed by a relaxation/fluctuation of the solvent necessary to reach the equilibrium product state. The free energy for the TS is $^{1-5}$

$$\Delta G^3 = \frac{\lambda}{4} \left( 1 + \frac{\Delta G^0}{\lambda} \right)^2,$$  

(2.3)

where $\Delta G^0$ is the free energy change in the ET process. According to the Marcus theory, the ET occurs at the crossing point between the two diabatic potential energy surfaces. At the crossing point, $E_P = E_R$ (energies of the reactant and the product are equal) and this occurs at one specific nuclear configuration. This satisfies both the Franck-Condon principle and the energy conservation law. When coupling between those diabatic states is small, the rate of the transition may be expressed using Fermi’s Golden rule. The full expression for the rate of ET according to the Marcus theory of ET is $^{1-5}$

$$k_{ET} = \frac{2\pi}{\hbar} |H_{DA}|^2 \frac{1}{\sqrt{4\pi \lambda k_BT}} \exp \left\{ -\frac{\left( \Delta G^0 + \lambda \right)^2}{4\lambda k_BT} \right\},$$  

(2.4)

where $\Delta G^0$, $\lambda$, and $H_{DA}$ are the free energy change, reorganization energy, and coupling between the electronic states involved in ET. The definitions of these quantities are given in Fig. 2.2. The $\Delta G^0$ is the overall thermodynamic drive for ET, the reorganization energy, $\lambda$, is related to an effective barrier for ET, and $H_{DA}$ is an electronic coupling between the ground and charge-transfer (CT) states.
Figure 2.2: Gibbs free energy curves and definitions of the key quantities in the Marcus theory of electron transfer. O and R denote oxidized and reduced states, respectively.

\[ \Delta G_0 \lambda \Delta G_0 \lambda \]

(a) \( \lambda > -\Delta G^0 \)

(b) \( \lambda = -\Delta G^0 \)

(c) \( \lambda < -\Delta G^0 \)

Figure 2.3: Three different regimes in Marcus theory: (a) the normal Marcus region, (b) the regime wherein \( \Delta G^0 = \lambda \) giving rise to maximum rate of ET, and (c) the inverted Marcus region.

In the limit of strong coupling, the ET proceeds in the adiabatic regime such that a single passage over the activation barrier completes ET.\(^6\) In this regime, the Marcus expression assumes the following form:

\[
k_{ET} = k_0 \exp \left\{ -\frac{\left( \Delta G^0 + \lambda \right)^2}{4\lambda k_B T} \right\},
\]

(2.5)

with pre-exponential factor \( k_0 \approx 10^{12} - 10^{13} \ \text{s}^{-1} \). The pre-exponential factor in transition-state theory is \( k_0 = \frac{k_B T}{\hbar} = 6.2 \times 10^{12} \).
The beauty of Eqs. 2.3 and 2.4 is that they predict that the rate of ET does not always increase with an increase in the free energy of reaction, and may even decrease. To elaborate upon this, Marcus defined three distinct cases shown in Fig. 2.3.

### 2.2 Calculating $\Delta G$ and $\lambda$ for a redox process using LRA

The most practical aspect of Marcus theory is that, in order to compute the rate of an ET process, we need to compute three parameters: (i) the free energy change of the process ($\Delta G^0$), (ii) the reorganization energy ($\lambda$), and (iii) electronic coupling ($H_{DA}$). The first difficulty one faces when calculating these quantities lies in the representation of the reaction coordinate. In the original Marcus theory, collective solvent coordinate is used, which is difficult to represent explicitly in calculations. Later, the idea of using the energy difference between initial and final states, $\Delta E$, was introduced and successfully exploited by Warshel and co-workers.\(^7\)\(^{-10}\) Another difficulty in computing the parameters for the Marcus rate expression is that one needs to go beyond electronic energies and deal with free energies. We use linear response approximation (LRA)\(^{11,12}\) to compute free energy and reorganization energy.

The redox reaction we study is:

$$Chro^- \rightarrow Chro^{\cdot} + e^-.$$  \hspace{1cm} (2.6)

The free energy of a reaction, $\Delta G^0$, is related to the partition functions of the initial and final states as $\Delta G^0 = -k_B T \ln \frac{Q_f}{Q_i}$, where $Q_f$ and $Q_i$ are the partition functions for the final state and initial state, respectively. This expression can be rewritten as
a relation between the free energy and the ensemble average of the energy difference \((\Delta E)\) between those two states:

\[
\Delta G^0 = -k_B T \ln \langle \exp \left( -\frac{\Delta E}{k_B T} \right) \rangle_i.
\] (2.7)

We are interested in an oxidation reaction (Eq. 2.6), thus the final and initial states are oxidized and reduced states, respectively. By truncating this expression after the first term, two expressions for oxidized and reduced state are obtained: \(\Delta G^0 = \langle \Delta E \rangle_R\) and \(\Delta G^0 = \langle \Delta E \rangle_O\). Combining these two expressions, we obtain the LRA expression for free energy and reorganization energy:

\[
\Delta G^0 = \frac{1}{2} \left( \langle E_O - E_R \rangle_R + \langle E_O - E_R \rangle_O \right),
\] (2.8)

\[
\lambda = \frac{1}{2} \left( \langle E_O - E_R \rangle_R - \langle E_O - E_R \rangle_O \right).
\] (2.9)

We use the following protocol to compute these quantities. First, we run MD for the initial (Chro\(^{-}\)) and oxidized (Chro\(^{+}\)) states of the protein to generate equilibrium sampling (see Section 2.8.1). We then follow with the QM/MM calculations of vertical detachment energy (VDE) on both states, where VDE is defined as:

\[
VDE = E_O - E_R.
\] (2.10)
where $E_O$ and $E_R$ are energies of the oxidized and reduced species, respectively. The detailed protocol is described in Section 2.8.3. The same expressions can also be derived from the pictorial representation of the free energy surfaces shown in Fig. 2.2.

\[ \lambda = \langle E_O - E_R \rangle_R - \Delta G^0, \] (2.11)
\[ \lambda = \langle E_R - E_O \rangle_O + \Delta G^0. \] (2.12)

Rearranging Eqs. 2.11 and 2.12, we obtain

\[ \langle E_O - E_R \rangle_R - \Delta G^0 = \langle E_R - E_O \rangle_O + \Delta G^0, \] (2.13)
\[ \langle E_O - E_R \rangle_R - \lambda = \lambda - \langle E_R - E_O \rangle_O. \] (2.14)

Rearranging Eqs. 2.13 and 2.14, one obtains Eqs. 2.8 and 2.9.

### 2.3 Calculating $\Delta G$ and $\lambda$ for an ET process using LRA

We use the term “ET” to refer a pure electron transfer reaction (i.e., one that is uncoupled from other processes). The states that are involved in ET are denoted as the “ground-state” and “charge-transfer (CT) state.” Here we will use the same model as in the previous section, but with the oxidized state (O) and reduced state (R) now referring to the CT and the ground-state, respectively. We apply LRA to calculate energy terms by computing the energies of the CT states (relative to the ground state, $\Delta E_{CT}$) on the ground-state and CT-state free energy surfaces. In our system, the CT state corresponds
Figure 2.4: Pictorial representation of ET process. An electron is transfered from the donor (D) to the acceptor (A). The ground-state is shown in green and the CT-state is shown in red. In our study, the donor and the acceptor are the negatively charged chromophore and the tyrosine residue, respectively.

to ET from the chromophore to a tyrosine residues. The expression for $\Delta E_{CT}$ can be written as:

$$\Delta E_{CT} = VDE_{chr^-} + VEA_Y + E_{coul} \approx VDE_{chr^-} + VEA_Y.$$  \hspace{1cm} (2.15)

Since only the donor or the acceptor is charged in both states, we neglect the Coulomb interaction term between Chro and Tyr. The expressions for free energy and reorganization energy of charge transfer are then:

$$\Delta G_{CT} = \frac{1}{2} \left( \langle \Delta E_{CT} \rangle_g + \langle \Delta E_{CT} \rangle_{CT} \right),$$  \hspace{1cm} (2.16)

and:

$$\lambda_{CT} = \frac{1}{2} \left( \langle \Delta E_{CT} \rangle_g - \langle \Delta E_{CT} \rangle_{CT} \right).$$  \hspace{1cm} (2.17)
where $\Delta E_{CT} \equiv E_{CT} - E_g$. The terms in Eqns. (2.16) and (2.17) can be computed as the vertical detachment and the vertical attachment energies (VEA) of the donor (chromophore) and acceptor (tyrosine). These expressions and the sign convention are the same as in refs. 13–15.

Prior to computing these quantities, we run MD for the ground ($\text{Chro}^- \ldots \text{Tyr}$) and CT ($\text{Chro}^+ \ldots \text{Tyr}^-$) states to generate equilibrium sampling (see Section 2.8.1). We then follow with the QM/MM calculations of VDE, VEA, and the couplings. The detailed protocol is described in Section 2.8. A more appropriate way to treat the ET processes is to actually compute the free energy surfaces,\textsuperscript{10} which would give a good estimate of the reorganization energy. In constructing the free energy profiles for each electron transfer in an ET reaction, the first step is to define the reaction coordinate. The straightforward choice, following Warshel’s pioneering work,\textsuperscript{7–10} would be the vertical energy gap ($\Delta E_{CT}$) between the reactant and product states. Below, we use a simple example to illustrate how one can use this energy gap as a reaction coordinate and construct a free energy curve.

Warshel and co-workers\textsuperscript{10} constructed free energy curves (Fig.2.5a) of the reactant and product states for ET in a photosynthetic bacterial reaction center. For detailed description, please see Ref. 10. In Fig. 2.5, P and B are denoted as the electron donor and acceptor. Here, the two curves refer to the free energy curves for the ground (PB) and CT ($P^+B^-$) states. Sampling $\Delta E_{CT}$ obtained from MD trajectory in a constant time interval generates a histogram of $\Delta E_{CT}$. This provides a statistical distribution. Again, $\Delta E_{CT}$ is related to free energy, $G(\Delta E_{CT})$, by the $k_BT$-weighted probability function as:

$$G(\Delta E_{CT}) = -k_BT \ln P(\Delta E_{CT}).$$ \hspace{1cm} (2.18)
The probability function, $P(\Delta E_{CT})$ may be obtained from a Gaussian fit of the histogram. With the finite sampling limitation, the reorganization energy obtained from each of these free energy surfaces may be slightly different from each other (Fig.2.5b). In this particular example, the authors used an average of those two reorganization energies.\textsuperscript{10} Using the constructed free energy curves (Fig.2.5b) they obtained two values, $\lambda_1 = 1.45$ kcal/mol, and $\lambda_2 = 1.55$ kcal/mol. They reported an average value of $\lambda \approx 1.5$ kcal/mol, from this calculation.

![Figure 2.5: (a) Probability function and (b) the free energy curves for ground (PB) and CT (P$^+$B$^-$) states. $\Delta V$ represents the energy gap between two states along the trajectories. Adapted from Ref. 10.](image)

In principle, one can follow this recipe and use even better methods to perform the sampling. For example, one could use QM/MM.\textsuperscript{16} But even with modern computers, QM sampling calculations are still very expensive for a system such as GFP. The GFP chromophore is a large $\pi$-conjugated molecule and is negatively charged. This means that one needs to use relatively large basis sets to obtain quantitative accuracy. These difficulties motivated us to design a cheaper protocol circumventing the sampling at the QM level. We perform the sampling at the MM level and then extract snapshots from
the MM trajectory. We only perform QM/MM calculations on those snapshots and the MM part is included as point charges. The drawback of this approach is that we neglect the effect of the polarization in the MM part.

2.4 MD simulation setup

Preparation of systems for MD and QM/MM simulations

Molecular dynamics. MD simulations employed the CHARMM27 parameters for standard protein residues\textsuperscript{17} and the parameters derived by Reuter \textit{et al.} for the anionic GFP chromophore.\textsuperscript{18} Parameters for the oxidized chromophore were obtained by adjusting the structural parameters and point charges using a protocol based on the extrapolation between the reduced and oxidized structures. The details of the parameters is provided in supporting information of Ref. 19. We used X-ray structures by Watcher of eYFP-H148Q for eYFP, with and without iodide anion.\textsuperscript{20} The structures were obtained from the protein data bank with pdb id 1F0B and 1F09 for eYFP without and with the iodide anion, respectively.\textsuperscript{20} For GFP, we used the 1EMA structure.\textsuperscript{21}

The TIP3P water model was used to describe explicit solvent molecules around the protein. Since CHARMM27 only has parameters for chloride, the iodide from the X-ray structures was replaced by Cl\textsuperscript{−}. The protein was solvated in a box, producing a water buffer of about 15 Å. The surface charges were neutralized with Na\textsuperscript{+} and Cl\textsuperscript{−} ions at appropriate positions. The MD calculations were performed on these systems as follows:

1. Minimization for 2000 steps with 2 fs time step prior to adding the water box.

2. Minimization for 2000 steps with 2 fs time step of the solvated structure.
3. Equilibration of the solvent using periodic boundary condition (PBC) with 1 fs time step for 500 ps. In this step, protein structure was frozen and only the solvent was allowed to relax.

4. Equilibration run for 2 ns with 1 fs time step with PBC in which the whole system was allowed to move under constant pressure and temperature (NPT ensemble).

5. Production run for 2 ns with 1 fs time step with PBC.

6. The snapshots for the QM/MM calculations were collected from the production run.

The MD simulations were performed using NAMD in an isobaric-isothermal ensemble with Langevin dynamics. The pressure and temperature used for the simulations were 1 atm and 298 K. All simulations invoked the rigid-bond option of NAMD, which kept the OH bonds frozen.

2.4.1 Protein structures and protonation states

Determining protonation states requires a combination of techniques. Only indirect information about protonation states is provided by X-ray structures. For instance, the distances between heavy atoms may suggest the presence of a proton participating in a hydrogen bond. Experimental kinetics studies, especially isotope effects, and the pH dependence of optical properties are often used to elucidate protonation states. Protonation states can be unambiguously determined by vibrational spectroscopy. Computational methods, which include several complementary approaches, are also particularly useful for this task. The most rigorous approach is to compute Gibbs free energies of
various protonation states, in order to identify the most stable form.\textsuperscript{23, 24} Such calculations require high accuracy from an underlying electronic structure method and extensive thermodynamic averaging. This approach has been used, for example, to calculate pKa shifts due to cis-trans photoisomerization in Dronpa and Padron.\textsuperscript{25} As a shortcut, one can consider optimized structures of the protein in different protonation forms. Unfavorable protonation states might be found unstable or cause large deformation of the hydrogen-bonding network around the chromophore, allowing them to be ruled out.\textsuperscript{19, 26, 27} Finally, one can compute spectroscopic properties of different forms and compare them with the experimental absorption maxima.\textsuperscript{26} The combination of the latter two approaches has allowed the determination\textsuperscript{26} of the protonation state of the blue intermediate (a transient form in the red chromophore maturation process), for which several protonation states had been proposed. Apart from the protonation states of the active site (in case the GFP chromophore), it is also important to determine the protonation states of other amino acids present in the protein structures. Sometimes it is very straightforward; one simply performs a combined analysis of pKa and the hydrogen-bonding network involving the residue of interest. But there are instances where such analyses do not lead to a single possibility. We encountered one such instance in the system preparation stage of our study of YFPs.

Protonation states for all proteins were checked with \textit{Propka software} before the residues around the chromophore were checked manually, as described below. \textit{Propka} suggested that Glu222 should be protonated in both YFPs (with or without halide). We performed dynamics on two protonation states for both eYFP structures. The two possible protonation states for eYFPs are:

1. GLU 222: Deprotonated form of GLU 222

2. GLUP 222: Protonated form of GLU 222
We ran a 2 ns trajectory for both protonation states of the two eYFP structures and computed average hydrogen-bond distances. We then compared these distances with those from the X-ray structures. The protonated form of Glu222 yielded the best agreement. We protonated the oxygen that is closest to the chromophore (CR2 66). The other oxygen atom of Glu222, which is closer to Tyr203, was not protonated. In the protonated form, the hydrogen bond is formed between the carboxylic oxygen atom of Glu222 and the nitrogen atom of the imidazolinone ring.

![Figure 2.6](image)

Figure 2.6: Possible H-bonds around the chromophore in eYFPs without and with halide.

The $\pi$-stacked tyrosine (Tyr203) is not present in eGFP. In eGFP, the deprotonated form of Glu222 must be considered since it can form a hydrogen bond with the threonine-like side chain of the eGFP chromophore in the deprotonated form (Fig. 2.7). The phenolate oxygen of the eGFP chromophore forms a hydrogen bond with nitrogen of His148. Therefore, we protonated the N-atom of the His148 residue (HSD form) that is closest to the chromophore. Arg96 may also form a hydrogen bond with the oxygen atom of the imidazolinone group of the eGFP chromophore. These protonation states are the same as in Ref. 28, except for Glu222, which was protonated in Ref. 28 but is deprotonated in our model.
To check the effect of mutating the Tyr145 residue, we constructed eGFP-Y145L by mutating Tyr145 to Leu145 using Mutator plugin in VMD followed by the energy minimization of the mutated protein with the same force-field parameters described above.

2.5 The Pathways model

To identify possible intermediate electron acceptors, we applied the Pathways model\textsuperscript{29, 30} in which the tunneling probability ($T_{DA}$) between the specified donor and acceptor moieties is given by

$$T_{DA} = K \prod_C \epsilon_C \prod_H \epsilon_H \prod_S \epsilon_S,$$

(2.19)
where \( C, H, \) and \( S \) refer to the pathways through covalent bonds, H-bonds, and space, respectively. \( \epsilon_C, \epsilon_H \) and \( \epsilon_S \) are empirical factors given by:

\[
\begin{align*}
\epsilon_C &= 0.6, \\
\epsilon_H &= \epsilon_C^2 e^{-1.7(R-2.8)}, \\
\epsilon_S &= \epsilon_C e^{-1.7(R-1.4)},
\end{align*}
\]

which take into account that tunneling through the covalent bonds is more efficient than through hydrogen bonds, etc. Pathways model treats tunneling through covalent bonds as the most efficient mechanism of tunneling, whereas it treats through space tunneling as the least efficient mechanism.

We computed \( T_{DA} \) for all possible electron acceptors in eGFP. The results are summarized in Table 2.1. The computed values roughly correlate with the DA distances. Based on the data in Table 2.1, the most likely electron acceptors in eGFP are: Tyr145, Tyr92, Phe64, Phe165, and His148. Based on the corresponding EAs, we can neglect His148 and consider Tyr145 and Tyr92 as the most likely acceptors. Results of docking and calculations of coupling presented below indicate that Tyr145 is the most important intermediate electron acceptor.

As the next step, we compare \( T_{DA} \) in eGFP with those in eYFP (with and without chloride). The results are given in Table 2.2. The most important observations are: (i) \( T_{DA} \) to Tyr203 in eYFP is comparable to that for Tyr145; (ii) the chloride has a very small effect on all tunneling probabilities, except for that to Tyr92 (chloride binding increases \( T_{DA} \) by a factor of 15). We then consider \( T_{DA} \) from Tyr203 to other possible acceptors (see Table 2.3). We observe that these rates are not affected by \( \text{Cl}^- \), except for Tyr92. We also note that although \( T_{DA} \) to Tyr145 is relatively large, the ET is
Table 2.1: eGFP Pathways calculations: $T_{DA}$ from chromophore to potential electron acceptors.

<table>
<thead>
<tr>
<th>Acceptor</th>
<th>$T_{DA}$</th>
<th>Mediated by</th>
<th>Distance, Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr145</td>
<td>$1.7 \times 10^{-2}$</td>
<td>through space</td>
<td>4.6</td>
</tr>
<tr>
<td>Tyr92</td>
<td>$1.4 \times 10^{-3}$</td>
<td>Val68, Gln69</td>
<td></td>
</tr>
<tr>
<td>Tyr143</td>
<td>$6.2 \times 10^{-5}$</td>
<td>Tyr145</td>
<td></td>
</tr>
<tr>
<td>Tyr151</td>
<td>$7.3 \times 10^{-4}$</td>
<td>Val150</td>
<td></td>
</tr>
<tr>
<td>Tyr200</td>
<td>$1.4 \times 10^{-4}$</td>
<td>Val150, Leu201</td>
<td></td>
</tr>
<tr>
<td>Tyr182</td>
<td>$5.2 \times 10^{-4}$</td>
<td>Arg96, Gln183</td>
<td></td>
</tr>
<tr>
<td>Phe46</td>
<td>$1.2 \times 10^{-3}$</td>
<td>Phe64</td>
<td></td>
</tr>
<tr>
<td>Phe64</td>
<td>$7.8 \times 10^{-2}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe71</td>
<td>$7.8 \times 10^{-4}$</td>
<td>Val68, Gln69, Cys70</td>
<td></td>
</tr>
<tr>
<td>Phe84</td>
<td>$2.5 \times 10^{-4}$</td>
<td>Gln69</td>
<td></td>
</tr>
<tr>
<td>Phe165</td>
<td>$1.1 \times 10^{-2}$</td>
<td>through space</td>
<td>3.77</td>
</tr>
<tr>
<td>Phe223</td>
<td>$1.3 \times 10^{-3}$</td>
<td>Glu222</td>
<td></td>
</tr>
<tr>
<td>His148</td>
<td>$3.6 \times 10^{-1}$</td>
<td>through space</td>
<td></td>
</tr>
<tr>
<td>Trp57</td>
<td>$1.7 \times 10^{-4}$</td>
<td>Phe64</td>
<td></td>
</tr>
</tbody>
</table>

mediated by the chromophore. These calculations suggest no pathways that would allow the electron to hop efficiently from Tyr203 to any acceptor, other than the chromophore.
Table 2.2: $T_{DA}$ from the chromophore to potential electron acceptors in eYFP using the Pathways model

<table>
<thead>
<tr>
<th>Acceptor</th>
<th>$T_{DA}$ (eYFP)</th>
<th>$T_{DA}$ (eYFP+Cl$^-$)</th>
<th>Ratio (Cl$^-$/no Cl$^-$)</th>
<th>Ratio (eYFP/eGFP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr203</td>
<td>$2.3 \times 10^{-2}$</td>
<td>$2.4 \times 10^{-2}$</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Tyr145</td>
<td>$1.9 \times 10^{-2}$</td>
<td>$2.7 \times 10^{-2}$</td>
<td>1.4</td>
<td>1.1</td>
</tr>
<tr>
<td>Tyr92</td>
<td>$4.8 \times 10^{-4}$</td>
<td>$7.1 \times 10^{-3}$</td>
<td>14.8</td>
<td>0.3</td>
</tr>
<tr>
<td>Tyr143</td>
<td>$6.8 \times 10^{-5}$</td>
<td>$8.5 \times 10^{-5}$</td>
<td>1.3</td>
<td>1.1</td>
</tr>
<tr>
<td>Tyr151</td>
<td>$6.0 \times 10^{-4}$</td>
<td>$3.3 \times 10^{-4}$</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>Tyr200</td>
<td>$1.3 \times 10^{-4}$</td>
<td>$7.6 \times 10^{-5}$</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Tyr182</td>
<td>$1.6 \times 10^{-4}$</td>
<td>$1.6 \times 10^{-4}$</td>
<td>1.0</td>
<td>0.3</td>
</tr>
<tr>
<td>Phe46</td>
<td>$1.4 \times 10^{-3}$</td>
<td>$1.1 \times 10^{-3}$</td>
<td>0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>Phe64</td>
<td>$7.8 \times 10^{-2}$</td>
<td>$7.8 \times 10^{-2}$</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Phe71</td>
<td>$7.8 \times 10^{-4}$</td>
<td>$7.8 \times 10^{-4}$</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Phe84</td>
<td>$2.3 \times 10^{-4}$</td>
<td>$2.6 \times 10^{-4}$</td>
<td>1.1</td>
<td>0.9</td>
</tr>
<tr>
<td>Phe165</td>
<td>$1.5 \times 10^{-2}$</td>
<td>$1.4 \times 10^{-2}$</td>
<td>0.9</td>
<td>1.4</td>
</tr>
<tr>
<td>Phe223</td>
<td>$4.4 \times 10^{-3}$</td>
<td>$2.9 \times 10^{-4}$</td>
<td>0.1</td>
<td>3.4</td>
</tr>
<tr>
<td>Trp57</td>
<td>$1.3 \times 10^{-4}$</td>
<td>$2.1 \times 10^{-4}$</td>
<td>1.6</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Table 2.3: $T_{DA}$ from Tyr203 to potential electron acceptors in eYFP using the Pathways model.

<table>
<thead>
<tr>
<th>Acceptor</th>
<th>$T_{DA}$ (eYFP)</th>
<th>$T_{DA}$ (eYFP+Cl$^-$)</th>
<th>Ratio (Cl$^-$/no Cl$^-$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr145</td>
<td>$3.1 \times 10^{-4}$</td>
<td>$2.4 \times 10^{-4}$</td>
<td>0.8</td>
</tr>
<tr>
<td>Tyr92</td>
<td>$5.7 \times 10^{-5}$</td>
<td>$3.8 \times 10^{-3}$</td>
<td>66.7</td>
</tr>
<tr>
<td>Tyr143</td>
<td>$1.2 \times 10^{-5}$</td>
<td>$1.1 \times 10^{-5}$</td>
<td>0.9</td>
</tr>
<tr>
<td>Tyr151</td>
<td>$7.7 \times 10^{-4}$</td>
<td>$8.3 \times 10^{-4}$</td>
<td>1.1</td>
</tr>
<tr>
<td>Tyr200</td>
<td>$3.6 \times 10^{-3}$</td>
<td>$3.6 \times 10^{-3}$</td>
<td>1.0</td>
</tr>
<tr>
<td>Tyr182</td>
<td>$1.1 \times 10^{-5}$</td>
<td>$2.2 \times 10^{-6}$</td>
<td>0.2</td>
</tr>
<tr>
<td>Trp57</td>
<td>$6.3 \times 10^{-7}$</td>
<td>$1.1 \times 10^{-6}$</td>
<td>1.7</td>
</tr>
</tbody>
</table>


2.6 Calculations of electronic couplings

The calculation of the coupling matrix element associated with ET can be mapped into a $2 \times 2$ problem. All one needs to do is diagonalize the following matrix:

\[
\begin{pmatrix}
E_D & V_{DA} \\
V_{AD} & E_A
\end{pmatrix},
\]

where $E_D$ and $E_A$ are the energies of the diabatic ground and CT states, respectively, and $V_{DA}$ and $V_{AD}$ are the off-diagonal matrix elements representing the coupling between ground and CT states. The corresponding eigenvalues of the above matrix can be written as:

\[
E_{1,2} = \frac{E_D + E_A}{2} \pm \sqrt{\left(\frac{E_D - E_A}{2}\right)^2 + V_{DA}^2}.
\]  (2.23)

In a special case of $E_D = E_A$, which is only satisfied at the point of avoided crossing, one can solve for coupling arriving at: $V_{DA} = (E_2 - E_1)/2$. In a more general case, the generalized Mulliken-Hush (GMH)\textsuperscript{31, 32} or the fragment-charge difference (FCD)\textsuperscript{33} methods can be used. The GMH method exploits the fact that in the basis of charge-localized states, the Hamiltonian looks like the one shown in Fig. 2.8, but the dipole operator is a diagonal matrix.

\[
\begin{pmatrix}
E_R & V_{RP} \\
V_{RP} & E_P
\end{pmatrix}
\]

\[
\begin{pmatrix}
\mu_R & 0 \\
0 & \mu_P
\end{pmatrix}
\]

\[
\begin{pmatrix}
E_1 & 0 \\
0 & E_2
\end{pmatrix}
\]

\[
\begin{pmatrix}
\mu_1 & \mu_{12} \\
\mu_{12} & \mu_2
\end{pmatrix}
\]

**Figure 2.8: Matrix representation of the GMH method**
The electronic coupling, according to the GMH method, is given in Eq. 2.24. In a nutshell, GMH relies on the change in dipole moment that occurs upon ET:

$$V_{DA} = \frac{\mu_{12}(E_1 - E_2)}{\sqrt{(\mu_1 - \mu_2)^2 + 4\mu_{12}^2}}. \tag{2.24}$$

One of the drawbacks of this method is that it requires an excited state calculation, either CIS or TDDFT, from which the energy of the CT state and the transition dipole moment ($\mu_{12}$) can be obtained. CIS and TDDFT often grossly overestimate the energies of the CT states leading to large errors in the couplings.

Therefore, to compute coupling, one must construct accurate diabatic states by localizing the electron density in the relevant regions of the system. Describing ET process by using density functional theory (DFT) is not simple. Often, self-interaction errors present in DFT lead to over-delocalization of the excess charge. One way to circumvent this problem is by adding an additional potential to the Hamiltonian as a constraint. Since in the ET process the net charge of the donor and acceptor (both before and after ET) is characteristic of each diabatic state, the constraining potential should be derived such that the charges of each part of the system are fixed. The initial idea was introduced by Wesolowski and Warshel\textsuperscript{34} and was called frozen-DFT (FDFT). In FDFT, the system is divided into the two subsystems, and the entire system is treated quantum-mechanically. One subsystem contains the solute molecule and the other contains solvent molecules. The density on the solvent is frozen by the use of the constraint, and the energy of the system is minimized with respect to the density of the solute in the presence of the frozen density (as an effective potential) of solvent. Although it was initially developed and benchmarked for a single closed-shell solvated molecule, later it was used to describe more complicated biological reactions including metal-catalyzed
reactions.\textsuperscript{35} This method was also successfully used in free energy calculations of metal-containing redox proteins.\textsuperscript{12} More recent studies extended this idea and paired it with empirical valence bond (EVB) model to extract the off-diagonal EVB terms (coupling) for $S_N2$ reactions in condensed phase.\textsuperscript{36, 37}

Van Voorhis and co-workers\textsuperscript{38–41} generalized the FDFD idea and introduced constrained DFT (CDFT). They applied CDFT to compute ET parameters for several systems, ranging from small sized molecules to fairly large systems including metal complexes. CDFT is now implemented in Q-Chem and NWChem. In addition to charge constraint, CDFT can be invoked with a spin constraint, which is also important since either the initial or the final state of an ET process corresponds to open-shell species.

In the context of ET processes, two different (ground and CT) states can be constructed with a constraint such as localization of electron density on D in the reactant state and on A in the product state. Those states can be considered as diabatic states. We can use the energies of the diabatic states to compute $\Delta E_{CT}$, which is the reaction coordinate we used in LRA.

In principle, one could extract electronic coupling from the CDFT calculations, but the following issues present in the theory of CDFT complicate this task:

1. The constraint is different for the two diabatic states, those states are essentially derived as a ground state of two different Hamiltonians, so they are not orthogonal to each other.

2. The Hamiltonian matrix is written in a constrained basis using Kohn-Sham wave functions ($\Phi$) instead of using the true wave functions ($\Psi$). The diagonal elements are CDFT energies, and $H_{DA} = H_{AD}$ is not necessarily true here, since there would be a small difference between them that originates from the fact that the
basis does not consist of exact wave functions. This may be solved by taking an average of $H_{DA}$ and $H_{AD}$.

In the next section, we describe the CDFT procedure as implemented by Wu and Van Voorhis in Q-Chem.

### 2.6.1 Constrained DFT method and CDFT-CI

In the CDFT formalism a unique constraining potential is added to the Hamiltonian. The ground state wavefunction in the presence of the constraining potential satisfies

$$\left( H + V_c w_c \right) \Psi_c = F \Psi_c,$$  \hspace{1cm} (2.25)

where $w_c$ defines the property of interest, e.g., the electronic population. If $N_c$ is the target value, then $\int w_c(r) \rho_c(r) dr = N_c$. In Eq.2.25, $V_c w_c$ is the unique constraining potential for a particular state. Using Hohenberg-Kohn theorem, the wavefunctions in these constrained states can be written as a functional of the densities in those states. Therefore, the coupling element, $H_{DA}$, can be written as:

$$H_{DA}[\rho_D, \rho_A] \equiv \langle \Psi_D(\rho_D)|H|\Psi_A(\rho_A) \rangle.$$  \hspace{1cm} (2.26)

Using Eq.2.25, Eq. 2.26 can also be written as:

$$H_{DA} = \langle \Psi_D|H + V_c^A w_c - V_c^A w_c|\Psi_A \rangle$$

$$= F_A \langle \Psi_D|\Psi_A \rangle - V_c^A \langle \Psi_D|w_c|\Psi_A \rangle.$$  \hspace{1cm} (2.27)
$F_A$ may be expressed as\(^\text{39}\)

$$F_A = \langle \Psi_A | H + V_c^A w_c | \Psi_A \rangle = E[\rho_A] + V_c^A \int w_c \rho_A = E_A + V_c^A N_c. \quad (2.28)$$

In the constrained basis the matrix form of the Hamiltonian is\(^\text{39}\)

$$H = \begin{pmatrix} H_{DD} & H_{DA} \\ H_{AD} & H_{AA} \end{pmatrix},$$

where

$$H_{DD} = \langle \Phi_D | H | \Phi_D \rangle = E[\rho_D] = E_D, \quad (2.29)$$

$$H_{AA} = E_A, \quad (2.30)$$

$$H_{DA} = F_A \langle \Phi_D | \Phi_A \rangle - V_c^A \langle \Phi_D | w_c | \Phi_A \rangle, \quad (2.31)$$

$$H_{AD} = F_D \langle \Phi_A | \Phi_D \rangle - V_c^D \langle \Phi_A | w_c | \Phi_D \rangle. \quad (2.32)$$

Because of the nonorthogonality of the diabatic states, $H_{DA}$ is not equal to the coupling element that enters the Marcus rate expression. To obtain the desired coupling via CDFT, one needs to compute an overlap matrix, $S$ and a projection matrix, $w_c$, and solve for $w_c$ from $w_c C = SCn$, where $n$ is a diagonal matrix. The overlap matrix can be written as\(^\text{39}\)

$$S = \begin{pmatrix} 1 & S_{DA} \\ S_{AD} & 1 \end{pmatrix},$$

The off-diagonal elements are: $S_{DA} = S_{AD} = \langle \Phi_D | \Phi_A \rangle$. The projection matrix is defined as\(^\text{39}\)
\[ w_c = \begin{pmatrix} w_c^{DD} & w_c^{DA} \\ w_c^{AD} & w_c^{AA} \end{pmatrix}, \]

where

\[ w_c^{DD} = \langle \Phi_D | w_c | \Phi_D \rangle = \int w_c \rho_D = N_c^D, \]
\[ w_c^{AA} = N_c^A, \]
\[ w_c^{DA} = w_c^{AD} = \langle \Phi_D | w_c | \Phi_A \rangle. \quad (2.33) \]

The basis is transformed further to a new basis as \( C^\dagger H C \). The off-diagonal elements in this transformed orthogonal basis are the ones we require for the calculations of the coupling. The authors called this method CDFT-CI.

In the present work, we used Q-Chem to compute the coupling with the CDFT-CI/MM method, where important residues were included in the QM part of the calculation. These calculations were performed along the trajectory and the average of the quantity along the trajectory was used to compute \( k_{ET} \) via Eq. 2.4.

Ding et al tested CDFT-CI on two bridge-separated ferrocene (Fc) moieties\(^{41}\) and studied bridges of different length and chemical character. Their results were in good agreement with the experiment. The results for the \( Fc - bridge - Fc^+ \) system shown in Table 2.4 clearly demonstrate that different bonding patterns affect the coupling, the effect ignored in the tunneling pathway model.

It was also demonstrated\(^{41}\) that coupling is most effective through a double bond, compared to a single or triple bond, using the same system shown in Table 2.4 and varying the C–C bonds in the bridge between the \( Fc \) units (see Table 2.5).
Table 2.4: Effect of different bonded network in the bridge between D and A. $H_{DA}$ calculated using CDFT-CI method. These values are from Ref. 41.

<table>
<thead>
<tr>
<th>Network</th>
<th>$H_{DA}$(kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Fc - CH = CH - benzene - CH = CH - Fc^+$</td>
<td>2.11</td>
</tr>
<tr>
<td>$Fc - CH = CH - triazole - CH = CH - Fc^+$</td>
<td>1.43</td>
</tr>
<tr>
<td>$Fc - C ≡ C - triazole - C ≡ C - Fc^+$</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Table 2.5: Effect of different carbon-carbon bonded network in the bridge between D and A. These results illustrate the distance-dependent decay of coupling. These values are from Ref. 41.

<table>
<thead>
<tr>
<th>Network</th>
<th>$H_{DA}$(kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Fc - (CH = CH)_3 - Fc^+$</td>
<td>3.42</td>
</tr>
<tr>
<td>$Fc - (CH = CH)_6 - Fc^+$</td>
<td>1.02</td>
</tr>
<tr>
<td>$Fc - (CH_2 - CH_2)_3 - Fc^+$</td>
<td>0.15</td>
</tr>
<tr>
<td>$Fc - (C ≡ C)_3 - Fc^+$</td>
<td>2.00</td>
</tr>
</tbody>
</table>

2.7 Docking calculations

To evaluate the distances between an external oxidant and selected residues, we performed docking calculations using AutoDock.\textsuperscript{42} In our system (GFP), the chromophore is protected by the $\beta$-barrel, so the oxidant cannot diffuse into the protein barrel. Rather, the oxidant is more likely to stay on the surface. The starting point in the docking calculation was the X-ray structure. We added hydrogen atoms according to the protonation states of each molecule and optimized the resulting structure. We used para-benzoquinone (BQ) as a model oxidant (structure optimized at $\omega$B97X-D/cc-pVTZ) and analyzed its docking to eGFP, eYFP, and halide-bound eYFP. In these calculations, we used a $22.5 \times 22.5 \times 22.5$ Å grid, centered around Tyr145, to perform the docking calculation. This grid covered the volume around the chromophore, Tyr145, and Tyr203. The
AutoDock software\textsuperscript{42} then determined 20-100 lowest-energy docking sites within this box.

We note that the differences in binding energies between different docking sites (as calculated by AutoDock) were less than 1 kcal/mol for the set of 100 lowest structures. An energy difference of 1 kcal/mol at 298 K leads to a Boltzmann population of about 20\%. Several clusters of docked structures were identified. Few of the lowest energy structures corresponded to BQ inserted into the barrel (close to the chromophore). For example, in eGFP, the fraction of such structures is 6/20 (\sim 30\%), and this cluster of conformations is ranked as 2nd lowest in energy.

Among surface-docked structures, some correspond to the relatively large distances between the chromophore and BQ. For example, a commonly occurring motif is one in which BQ is docked at the bottom of the barrel. For these structures the distance from the chromophore is about 8 Å. We focused on the structures with the shortest distance to the chromophore, Tyr145, Tyr203, and Tyr92. The structures that have the shortest Tyr145-BQ distance also have the shortest chromophore-BQ distance. The representative structures are shown in Figs. 2.9 and 2.10. As one can see from Fig. 2.9, docked BQ is partially inserted into the surface of eGFP and eYFP. The fraction of such structures among the manifold of the 20 lowest-energy surface-docked structures is 4/20 (\sim 20 \%) for eGFP, and those structures are lowest in energy.

As illustrated in Fig. 2.9, BQ can approach Tyr145 as close as 3.5-4.5 Å (which is similar to the Chro-Tyr145 distance), whereas Tyr203 is considerably less accessible (the shortest computed distance was about 7 Å). Thus, Tyr203 is unlikely to serve as an efficient ET to an outside oxidant; because it is buried inside, it is likely to be a dead end for ET.
Figure 2.9: Benzoquinone docked to eGFP, eYFP, and halide-bound eYFP. In the case of halide-bound eYFP, two docked BQ conformation were obtained, one close to Tyr145 and the other close to Tyr203. For the structure shown in panel (a), Chro-BQ distance is $\sim 6$ Å.

We performed similar docking analysis centered around Tyr92; the results are shown in Fig. 2.10. As one can see, Tyr92 is also not accessible to the external oxidants (the closest BQ-Tyr92 distance is 8 Å); thus, ET to this residue is unlikely to lead to the redding.

Figure 2.10: Benzoquinone docked to eGFP, eYFP, and halide-bound eYFP close to Tyr92.

To verify the results of the docking simulations, we performed MD simulations for the docked structure of eGFP with the shortest chromophore-BQ distance. The force-field parameters used for the BQ MD simulations were obtained as follows. The $\omega$B97X-D/aug-cc-pVTZ NBO point charges for the neutral form have been used.
Equilibrium bond lengths and valence angles were taken from the gas-phase equilibrium ωB97X-D/aug-cc-pVTZ geometry. Force constants and van der Waals parameters were taken from the CHARMM General Force Field parameters for phenol. We then performed equilibrium simulations in the ground state (deprotonated chromophore, neutral Tyr145, neutral BQ) for 10 ns. We then analyzed the distance between BQ and Tyr145 by computing the distances between one selected aromatic carbon of Tyr145 and BQ (see Fig. 2.11). We observed that the distance between Tyr145 and BQ stays mostly within 3.9-5.4 Å, for a 10 ns long MD trajectory. The average distance and the standard deviation are 4.63 Å and 0.46 Å, respectively. The averaging is performed using 4,000 snapshots along the trajectory.

To evaluate the feasibility of the direct tunneling (from the chromophore to an outside oxidant) and whether ET to a particular residue can lead to an efficient ET to an outside oxidant molecule, we performed docking calculations (see Section 2.7). The distance between docked species (BQ) and different residues characterizes the accessibility of these residues to an outside oxidant. The resulting structures can be used to calculate tunneling probabilities using the Pathways model and to compute electronic couplings using CDFT-CI.

2.8 Protocols for calculating energetics and couplings

2.8.1 QM/MM schemes

The VDE and the modified VDE (mVDE) of the chromophore and modified VEA (mVEA) of the tyrosine residues were computed using the following QM/MM protocol. The definition and relevance of the modified DE and EA is described in section 2.8.4. In a nutshell, the VDE term is only relevant in pure redox reactions. However, because of
Figure 2.11: Distance analysis between Tyr145 and BQ docked on the surface of eGFP along a 10 ns trajectory. Graphs show fluctuations in the relative distance along the trajectory (left) and the resulting distribution (right). The snapshots were taken every 2.5 ps.

In the way we developed our protocol, a modified definition of vertical detachment energy (mVDE) and vertical electron attachment (mVEA) is used when computing the Marcus parameters for the ET rate.

In VDE/mVDE calculations (as well as in the calculations of excitation energies), only the chromophore was included in the QM region, as shown in Fig 4.2. In mVEA calculations, the QM part consisted of the tyrosine residue (Fig. 4.3). For the CDFT-CI
calculations, both the chromophore and the tyrosine were included in the QM part. The rest of the system (protein+solvent) was treated as point charges.

The MD snapshots were converted to the QM/MM Q-Chem input file by using a Python script. VMD Tkconsole was used to extract the frames from the MD trajectories. We observed that convergence with respect to the number of snapshots is achieved relatively fast. In our production calculations, 41 snapshots were used for averaging.

The convergence of the redox parameters with respect to the number of snapshots has been studied in detail in Ref. 44. They reported the values of thermodynamic parameters obtained from 50 snapshots. However, their data showed that the convergence was achieved at 20 snapshots. We also observed fast convergence for the redox parameters (∼20 snapshots); however, the quantities relevant to the ET process required more extensive sampling.

Fig. 2.12 illustrates the convergence of thermodynamic averaging in our calculations. As follows from Eqns. (2.43) and (2.44), the convergence of $\Delta G$ and $\lambda$ is driven by $\langle \Delta E^\prime_{cdefcis}\rangle_g$ and $\langle \Delta E^\prime_{cdefcis}\rangle_{CT}$. These quantities for eYFP are shown in Fig. 2.12. We chose eYFP (without halide) as a representative system because it shows the slowest convergence, i.e., when we increased the number of snapshots from 19 to 41, the highest change in the ET rates was observed for this system, whereas the rates for eGFP were essentially converged at 19 snapshots.

We also checked the convergence of the electronic coupling values in eGFP along the ground-state trajectory. We observed that after about 25 snapshots the coupling values are essentially converged (see Fig. 2.13).

mVDEs, mVEAs and couplings were calculated with $\omega$B97X-D/aug-cc-pVDZ, $\omega$B97X-D/aug-cc-pVTZ and $\omega$B97X-D/cc-pVDZ. Since the side chains of eGFP were always in the QM region, the charge distributions at the QM-MM boundary in the
Figure 2.12: Convergence of $\langle \Delta E'_{\text{cdftci}} \rangle_g$ and $\langle \Delta E'_{\text{cdftci}} \rangle_C \text{T145}$ for eYFP without halide with Tyr145 as the intermediate acceptor.

mVDE and mVEA calculations were identical.

Three QM/MM schemes were used in the eGFP calculations:

1. QM - only chromophore for VDE/mVDE
Figure 2.13: Convergence of the Chro-Ty145 coupling in the ground state of eGFP.

2. QM - only Tyr145 for mVEA

3. CDFT-CI: QM - chromophore and Tyr145

Five QM/MM schemes were used in the eYFP calculations:

1. QM - only chromophore for VDE/mVDE

2. QM - only Tyr145 for mVEA

3. QM - only Tyr203 for mVEA

4. CDFT-CI: QM - chromophore and Tyr145

5. CDFT-CI: QM - chromophore + Tyr203

To validate our QM/MM schemes, we computed excitation energies using SOS-CIS(D)/aug-cc-pVDZ and compared them with the experimental absorption energies.
We observe excellent agreement between the computed and experimental values of excitation energies, the largest error being 0.18 eV for the halide-bound eYFP.

Table 2.6: Comparison between the computed (SOS-CIS(D)/aug-cc-pVDZ) and experimental excitation energies (eV). Only the chromophore is included in the QM part and the rest of the protein and solvent were treated as point charges. Computed values were averaged over 19 snapshots.

<table>
<thead>
<tr>
<th>System</th>
<th>Computed</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFP</td>
<td>2.55</td>
<td>2.54</td>
</tr>
<tr>
<td>eYFP</td>
<td>2.55</td>
<td>2.41</td>
</tr>
<tr>
<td>eYFP + Cl(^-)</td>
<td>2.59</td>
<td>2.41</td>
</tr>
</tbody>
</table>

2.8.2 Benchmarks

VDE of chromophore

Here, we check the basis set sensitivity of the conventional VDE and VEA of the chromophore and tyrosine residues. We expect to observe the same trend in these parameter for modified VDE and VEA. These terms are defined in section 2.8.4. The effect of different basis sets on the VDE of the chromophore and the VEA of the tyrosine was studied in YFP without a halide. The basis sets used in this study were cc-pVDZ, cc-pVTZ, aug-cc-pVDZ and aug-cc-pVTZ. We also examined the effect of adding a chloride ion to the QM region. The default basis sets for VDEs are aug-cc-pVDZ and for VEAs are aug-cc-pVTZ. For all DFT calculations, the \( \omega \)B97X-D functional was used and we calculate the effect of basis set in terms of deviation of VDE or VEA from the default level, defined as \( \Delta VDE_i = VDE_{\text{aug-cc-pVDZ}} - VDE_i \) and \( \Delta VEA_j = VEA_{\text{aug-cc-pVTZ}} - VEA_j \). We performed these benchmark calculations for 4 random snapshots from the ground-state MD trajectory.
Table 2.7: Basis set sensitivity of the VDE of the chromophore of YFP (without halide). The first column is taken to be the default (base) values, and the other columns list deviation from the base value.

<table>
<thead>
<tr>
<th>snapshots</th>
<th>$VDE$ (eV)</th>
<th>$\Delta VDE_i$ (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>aug-cc-pVDZ</td>
<td>cc-pVDZ</td>
</tr>
<tr>
<td>1</td>
<td>6.40</td>
<td>+0.35</td>
</tr>
<tr>
<td>2</td>
<td>6.80</td>
<td>+0.33</td>
</tr>
<tr>
<td>3</td>
<td>6.21</td>
<td>+0.34</td>
</tr>
<tr>
<td>4</td>
<td>5.93</td>
<td>+0.37</td>
</tr>
</tbody>
</table>

Table 2.7 shows that the aug-cc-pVDZ basis set gives nearly converged results since; if we increase the size of the basis set further to aug-cc-pVTZ, the value of $VDE$ changes by only 0.01 eV at most. We investigated whether a similar trend is observed for halide-bound YFP with the same QM/MM scheme (only the chromophore in QM part). We also examined the effect of adding the chloride ion in the QM region on the VDE of the chromophore.

Table 2.8: Basis set sensitivity on the VDE of the chromophore of YFP (with halide). The first column is taken to be the default (base) values, and second column lists deviation from the base value.

<table>
<thead>
<tr>
<th>snapshots</th>
<th>$VDE$ (eV), aug-cc-pVDZ</th>
<th>$\Delta VDE$ (eV), aug-cc-pVTZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.13</td>
<td>+0.01</td>
</tr>
<tr>
<td>2</td>
<td>5.73</td>
<td>+0.01</td>
</tr>
<tr>
<td>3</td>
<td>5.63</td>
<td>+0.00</td>
</tr>
<tr>
<td>4</td>
<td>6.00</td>
<td>+0.00</td>
</tr>
</tbody>
</table>

For production calculations, we computed the VDEs of the chromophore at the $\omega$B97X-D/aug-cc-pVDZ level of theory. The deviation from the aug-cc-pVTZ basis set is very small ($\leq 0.01$ eV). Therefore, extrapolation of the results obtained at the $\omega$B97X-D/aug-cc-pVDZ level of theory to complete basis set (CBS) limit is unnecessary. From Table 2.9, we noticed that the effect of including Cl$^-$ in QM part while
Table 2.9: Effect of adding Cl$^-$ in QM region on the VDE of the chromophore of YFP (with halide).

<table>
<thead>
<tr>
<th>snapshots</th>
<th>VDE (eV)</th>
<th>Chromophore</th>
<th>Chromophore + Cl$^-$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.13</td>
<td>6.09</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5.73</td>
<td>5.70</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5.63</td>
<td>5.61</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6.00</td>
<td>6.01</td>
<td></td>
</tr>
</tbody>
</table>

computing the VDE of chromophore was very small ($\leq$0.04 eV). We concluded that the effect of the chloride ion near the chromophore and relevant residues is purely electrostatic, an effect which is already captured in MD.

VEA of tyrosine

We also benchmarked the VEA of tyrosine. The effect of different basis sets was checked for YFP without the halide as a test case. These calculations were performed for four random snapshots taken from the ground-state surface of YFP without halide. The following basis sets were used for this purpose: cc-pVDZ, aug-cc-pVDZ, aug-cc-pVTZ, and aug/cc-pVQZ. In all calculations, the $\omega$B97X-D functional was used and only the relevant tyrosine residue was included in QM part. The remainder of the system was modeled by point charges.

Table 2.10 shows that although the $\omega$B97X-D/aug-cc-pVDZ level of theory provides a good estimate of the VDE of the chromophore, the same cannot be said about the VEA of tyrosine residues. The difference between the VEA of Tyr145 calculated using the aug-cc-pVDZ basis and the VEA calculated using the aug-cc-pVTZ basis is about 0.2–0.3 eV. Even with the aug-cc-pVQZ basis set, the VEA is not converged. We
Table 2.10: Basis set sensitivity of the VEA of the Tyr145 residue of YFP (without halide). The first column is taken to be the default (base) values, and the other three columns list deviation from the base value.

<table>
<thead>
<tr>
<th>snapshots</th>
<th>VEA (eV)</th>
<th>ΔVEA_i (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>aug-cc-pVTZ</td>
<td>aug-cc-pVQZ</td>
</tr>
<tr>
<td>1</td>
<td>-0.45</td>
<td>+0.17</td>
</tr>
<tr>
<td>2</td>
<td>-0.35</td>
<td>+0.14</td>
</tr>
<tr>
<td>3</td>
<td>-0.06</td>
<td>+0.10</td>
</tr>
<tr>
<td>4</td>
<td>0.61</td>
<td>+0.12</td>
</tr>
</tbody>
</table>

also found about 0.2 eV difference between the VEA of the Tyr145 using the aug-cc-pVDZ basis and the aug-cc-pVTZ basis in the case of the halide-bound YFP. To confirm that this basis set effect is due to the finite size of the basis set and not a manifestation of some other problem (such as spin contamination), we repeated the test with EOM-EA-CCSD instead of DFT. VEAs obtained from the EOM-EA-CCSD method are free from spin-contaminations, unlike DFT where spin contamination might result in artifacts in the VEA values. However, as shown in Table 2.11, EOM-EA-CCSD gave almost identical trends in basis set dependence as DFT calculations which means that the error in the VEA of tyrosine residues arises only from finite size of the basis set.

We also investigated the effect of adding Cl\(^-\) in the QM part on the VEA of tyrosine. Since the chloride ion is far away from the Tyr145, it should not affect the VEA of Tyr145. However, in our study, the chloride ion is located much closer to the other relevant tyrosine (Tyr203). It was then necessary to examine the effect of Cl\(^-\) on the VEA of the Tyr203 in halide-bound eYFP. We inspected the effect of adding Cl\(^-\) in the QM part at the ωB97X-D/aug-cc-pVTZ level of theory.

The results (Table 2.12) illustrate once again that the effect of Cl\(^-\) is purely electrostatic, since there was no difference between the VEA values obtained with a tyrosine only QM region and a tyrosine + Cl\(^-\) QM region.
Table 2.11: Comparison between VEAs of Tyr145 obtained at the EOM-EA-CCSD and DFT with same basis sets for 4 frames obtained from the ground-state trajectory of eYFP without halide. Only Tyr145 was included in the QM part of the QM/MM calculation. All values are given in eV.

<table>
<thead>
<tr>
<th>Method</th>
<th>frame</th>
<th>aug-cc-pVTZ</th>
<th>aug-cc-pVDZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$B97X-D</td>
<td>1</td>
<td>-0.45</td>
<td>-0.19</td>
</tr>
<tr>
<td>EOM-EA-CCSD</td>
<td></td>
<td>-0.43</td>
<td>-0.18</td>
</tr>
<tr>
<td>$\omega$B97X-D</td>
<td>2</td>
<td>-0.35</td>
<td>-0.12</td>
</tr>
<tr>
<td>EOM-EA-CCSD</td>
<td></td>
<td>-0.35</td>
<td>-0.12</td>
</tr>
<tr>
<td>$\omega$B97X-D</td>
<td>3</td>
<td>-0.06</td>
<td>-0.13</td>
</tr>
<tr>
<td>EOM-EA-CCSD</td>
<td></td>
<td>-0.06</td>
<td>-0.12</td>
</tr>
<tr>
<td>$\omega$B97X-D</td>
<td>4</td>
<td>0.61</td>
<td>0.84</td>
</tr>
<tr>
<td>EOM-EA-CCSD</td>
<td></td>
<td>0.61</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Figure 2.14: Relevant tyrosine residues around the chromophore with the halide.
Table 2.12: Effect of adding Cl\(^-\) in the QM part on the VEA of Tyr203 of YFP (with halide).

<table>
<thead>
<tr>
<th>frame</th>
<th>VEA (eV)</th>
<th>Tyrosine 203</th>
<th>Tyrosine 203 + Cl(^-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.05</td>
<td>-0.05</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.33</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.08</td>
<td>0.07</td>
<td></td>
</tr>
</tbody>
</table>

2.8.3 Relevant energies for chromophore oxidation and redox potentials

The Gibbs free energy of the ground-state chromophore oxidation can be computed using LRA as specified by Eqns. 2.8 and 2.9, where \(E_O - E_R \equiv VDE\). \(\Delta G_{ox}\) in the excited state was calculated as:

\[
\Delta G_{ox}^{ex} = \Delta G_{ox}^{gs} - E_{em}.
\]  

(2.34)

We computed VDE of the chromophore on the ground-state and oxidized chromophore surfaces for all three proteins. From Eqns. 2.8 and 2.34, we obtained the free energies of oxidation of the chromophore in the \(S_0\) and \(S_1\) states. The oxidation potential was calculated from the free energy of oxidation of the ground state as \(\Delta G_{ox}^{gs} = -nFE_{ox}\) (\(n=1\) for one-electron oxidation). To compute the standard oxidation potential with respect to the standard hydrogen electrode (SHE), we used \(\Delta G(SHE)=4.28\) eV (see Ref. 47).
Table 2.13: Redox properties of the chromophores of eGFP, eYFP and halide-bound eYFP. VDEs of the chromophores on the reduced (ground) and oxidized surfaces were averaged over 41 snapshots using ωB97X-D/aug-cc-pVDZ. Energies are in eV and the reduction potentials are in V with respect to SHE.

<table>
<thead>
<tr>
<th>System</th>
<th>$&lt; VDE &gt;_{red}$</th>
<th>$&lt; VDE &gt;_{ox}$</th>
<th>$\Delta G_{gs}^{ox}$</th>
<th>$\lambda_{ox}$</th>
<th>$E_{em}^{exp}$</th>
<th>$\Delta G_{ox}^{ex}$</th>
<th>$E_{red}^{0}$ vs. SHE (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFP</td>
<td>6.149</td>
<td>2.952</td>
<td>4.551</td>
<td>1.599</td>
<td>2.44</td>
<td>2.111</td>
<td>0.27</td>
</tr>
<tr>
<td>eYFP</td>
<td>6.097</td>
<td>3.297</td>
<td>4.697</td>
<td>1.400</td>
<td>2.35</td>
<td>2.347</td>
<td>0.42</td>
</tr>
<tr>
<td>eYFP + Cl$^-$</td>
<td>5.960</td>
<td>2.588</td>
<td>4.274</td>
<td>1.686</td>
<td>2.35</td>
<td>1.924</td>
<td>-0.01</td>
</tr>
<tr>
<td>eGFP-Y145L</td>
<td>6.076</td>
<td>3.020</td>
<td>4.548</td>
<td>1.528</td>
<td>2.44</td>
<td>2.108</td>
<td>0.27</td>
</tr>
</tbody>
</table>
2.8.4 Relevant energies for ET processes

In the calculations of the free energies for the CT process, Eqns. 2.8 and 2.9 need to be modified to account for the fact that the electron does not leave the protein, but is transferred to another residue (acceptor). Ideally, we would prefer to compute the energy of the CT state by a CDFT-CI calculation in which the two diabatic states, the ground and the CT state, are prepared explicitly and $\Delta E_{CT}$ is computed from the energy difference between these two states. However, due to the convergence issues of CDFT-CI with augmented bases, we employed an extrapolation scheme in which CDFT-CI energies obtained in a small basis set were corrected by the VDE and the VEA of the donor and acceptor computed using large bases.

We begin by computing $\Delta E_{CT}$ as

$$\langle \Delta E_{CT} \rangle_g = \langle E_{chr}^{A} - E_{chr}^{A^{-}} + E_{chr}^{chr^{-}} - E_{chr}^{chr^{-}} \rangle_g$$ \hspace{1cm} (2.35)$$

and

$$\langle \Delta E_{CT} \rangle_{CT} = \langle E_{chr}^{chr^{-}} - E_{chr}^{chr^{-}} + E_{chr}^{chr^{-}} - E_{chr}^{chr^{-}} \rangle_{CT},$$ \hspace{1cm} (2.36)$$

where A is an acceptor, which is neutral in the ground state and negatively charged in the CT state. Subscripts show the residue included in the QM region and superscripts represent the residue included in the MM region with remainder of the system. Eqns. 2.35 and 2.36 treat the problem as an independent redox problem for the donor and the acceptor because the change in the charge distribution of the acceptor is not accounted for.
As a next step, we compare $\langle \Delta E_{CT} \rangle$ with the CDFT-CI results. For that, we modify Eqns. 2.35 and 2.36 as,

$$\langle \Delta E_{CT} \rangle_g = \langle E_{chr}^- - E_{chr}^A + E_{chr}^{chr} - E_A^- \rangle_g$$  (2.37)

and

$$\langle \Delta E_{CT} \rangle_{CT} = \langle E_{chr}^- - E_{chr}^A + E_{chr}^{chr} - E_A^- \rangle_{CT}.$$  (2.38)

Note the subtle difference between Eqns. 2.35 and 2.37, Eqns. 2.36 and 2.38 – in the modified definition, the sum of charges on the chromophore and the acceptor is always $-1$ while computing these terms. The difference, $\delta E_{corr,i} = \langle \Delta E_{CC-pVDZ}^{ct} \rangle_i - \langle \Delta E_{cc-pVDZ}^{CT} \rangle_i$, arises from the orbital overlap and the Coulomb contribution as well as a small contribution from the adjustment of charges around the QM region in the CDFT-CI calculations.

We performed CDFT-CI calculations using $\omega$B97X-D/cc-pVDZ. We also computed $\Delta E_{CT}$ at the same level of theory using Eqns. 2.37 and 2.38 and at a higher level of theory (using aug-cc-pVTZ and aug-cc-pVQZ bases and the CBS extrapolation). In the mVDE calculations of the chromophore, the $\omega$B97X-D/aug-cc-pVDZ values are essentially converged (no difference relative to the $\omega$B97X-D/aug-cc-pVTZ values); thus, these values were used without extrapolation to CBS limit and in extrapolation of $\Delta E_{CT}$ to the CBS limit, only the mVEA of the acceptor was extrapolated.

We used a two-point extrapolation scheme to obtain mVEA at the $\omega$B97X-D/CBS limit using mVEA computed by $\omega$B97X-D/aug-cc-pVTZ and $\omega$B97X-D/aug-cc-pVQZ. We subtracted the correction term (obtained at the $\omega$B97X-D/cc-pVDZ level) from $\Delta E_{CT}$ computed at the CBS limit. The VEA or mVEA of tyrosine was defined as:

$$VEA_{tyr} = E_{tyr}^- - E_{tyr}^-.\text{ Both energies were extrapolated to the CBS limit using}$$
\[ E(X) = E_{CBS} + AX^{-3} \text{ and } E(Y) = E_{CBS} + AY^{-3}, \] where \( E(X) \) and \( E(Y) \) are the energies obtained with the aug-cc-pVTZ \((X = 3)\) and aug-cc-pVQZ \((Y = 4)\) basis sets, respectively. This was repeated for all frames.

With these modifications of Eqns. 2.16 and 2.17, the final expressions for the free energy and reorganization energy of charge transfer become:

\[
\Delta G_{CT} = \frac{1}{2} \left( \langle E_{e_{chr}^{-}}^A - E_{e_{chr}^{-}}^A + E_{e_{chr}^{-}}^{chr} - E_{A}^{chr} \rangle_g - \delta E_{corr,g} \right) \\
+ \langle E_{e_{chr}^{-}}^A - E_{e_{chr}^{-}}^A + E_{e_{chr}^{-}}^{chr} - E_{A}^{chr} \rangle_{CT} - \delta E_{corr,CT} \right) \tag{2.39}
\]

and

\[
\lambda_{CT} = \frac{1}{2} \left( \langle E_{e_{chr}^{-}}^A - E_{e_{chr}^{-}}^A + E_{e_{chr}^{-}}^{chr} - E_{A}^{chr} \rangle_g - \delta E_{corr,g} \right) \\
- \langle E_{e_{chr}^{-}}^A - E_{e_{chr}^{-}}^A + E_{e_{chr}^{-}}^{chr} - E_{A}^{chr} \rangle_{CT} - \delta E_{corr,CT} \right) \tag{2.40}
\]

Defining \( mVDE_{chr} \equiv E_{e_{chr}^{-}}^A - E_{e_{chr}^{-}}^A \) and \( mVEA_A \equiv E_{e_{chr}^{-}}^{chr} - E_{A}^{chr} \), Eqns. 2.39 and 2.40 become:

\[
\Delta G_{CT} = \frac{1}{2} \left( \langle mVDE_{chr} - mVDE_{chr} \rangle_g - \delta E_{corr,g} \right) \\
+ \langle mVDE_{chr} - mVDE_{chr} \rangle_{CT} - \delta E_{corr,CT} \right) \tag{2.41}
\]

and

\[
\lambda_{CT} = \frac{1}{2} \left( \langle mVDE_{chr} - mVDE_{chr} \rangle_g - \delta E_{corr,g} \right) \\
- \langle mVDE_{chr} - mVDE_{chr} \rangle_{CT} - \delta E_{corr,CT} \right). \tag{2.42}
\]
To further simplify these equations, we define the term, \( \langle \Delta E_{\text{cdftci}}' \rangle_i \equiv \langle mVDE_{\text{chr}} + mV EA_A \rangle_i - \delta E_{\text{corr},i} \), such that the above equations assume the following form:

\[
\Delta G_{CT} = \frac{1}{2} \left( \langle \Delta E_{\text{cdftci}}' \rangle_g + \langle \Delta E_{\text{cdftci}}' \rangle_{CT} \right), \tag{2.43}
\]
\[
\lambda_{CT} = \frac{1}{2} \left( \langle \Delta E_{\text{cdftci}}' \rangle_g - \langle \Delta E_{\text{cdftci}}' \rangle_{CT} \right). \tag{2.44}
\]

The subscript outside the ensemble average, \( \langle \ldots \rangle_i \), represents the surface on which the averaging was performed. Eqns. 2.43 and 2.44 compute the free energy change and reorganization energy involved in the ground-state CT process. But since we are interested in photoinduced ET, where the initial state is the excited state, we need to subtract \( E_{em} \) from \( \Delta G_{CT} \) to obtain \( \Delta G_{CT}^{ex} \) and \( \lambda_{CT}^{ex} \):

\[
\Delta G_{CT}^{ex} = \frac{1}{2} \left( \langle \Delta E_{\text{cdftci}}' \rangle_g + \langle \Delta E_{\text{cdftci}}' \rangle_{CT} \right) - E_{em}. \tag{2.45}
\]

Note that \( \lambda_{CT}^{ex} = \lambda_{CT} \), since subtraction of \( \langle \Delta E_{\text{cdftci}}' \rangle_g - E_{em} \) and \( \langle \Delta E_{\text{cdftci}}' \rangle_{CT} - E_{em} \) when computing reorganization energy cancels \( E_{em} \). We define:

\[
\langle \Delta E_{CT}' \rangle_i = \langle mVDE_{\text{chr}}^{CBS} + mV EA_Y^{CBS} \rangle_i \tag{2.46}
\]
\[
\langle \Delta E_{CT}^{cc-pVDZ} \rangle_i = \langle mVDE_{\text{chr}}^{cc-pVDZ} + mV EA_Y^{cc-pVDZ} \rangle_i, \tag{2.47}
\]

where \( i \) represents the surface on which these terms were computed. According to the definition, we also have, \( \delta E_{\text{corr},i} = \langle \Delta E_{CT}^{cc-pVDZ} \rangle_i - \langle \Delta E_{\text{cdftci}}' \rangle_i \) and we estimate the extrapolated energy difference between the ground and CT states as \( \langle \Delta E_{\text{cdftci}}' \rangle_i = \langle \Delta E_{CT}' \rangle_i - \delta E_{\text{corr},i} \).

Electronic couplings, \( H_{DA} \), were calculated using CDFT-CI.\(^{15,39-41} \) The relevant states are: (i) Chro\(^-\)+Tyr and (ii) Chro\(^+\)+Tyr\(^-\). Thus, both residues were included in
Table 2.14: Energy differences between the ground and CT states for 41 frames along the MD trajectory calculated on the ground-state surfaces. All values are in eV.

<table>
<thead>
<tr>
<th>System</th>
<th>Acceptor</th>
<th>(\langle \Delta E_{CT}\rangle_g)</th>
<th>(\langle \Delta E_{cdftci}\rangle_g)</th>
<th>(\delta E_{\text{corr},g})</th>
<th>(\langle \Delta E'_{CT}\rangle_g)</th>
<th>(\langle \Delta E'_{cdftci}\rangle_g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFP</td>
<td>145</td>
<td>9.4065</td>
<td>5.6426</td>
<td>3.764</td>
<td>7.501</td>
<td>3.737</td>
</tr>
<tr>
<td>eYFP</td>
<td>145</td>
<td>9.2662</td>
<td>5.4965</td>
<td>3.770</td>
<td>7.607</td>
<td>3.837</td>
</tr>
<tr>
<td></td>
<td>203</td>
<td>8.5109</td>
<td>5.0668</td>
<td>3.444</td>
<td>6.644</td>
<td>3.200</td>
</tr>
<tr>
<td>eYFP + Cl(^-)</td>
<td>145</td>
<td>8.9374</td>
<td>5.4130</td>
<td>3.524</td>
<td>7.221</td>
<td>3.697</td>
</tr>
<tr>
<td></td>
<td>203</td>
<td>9.4811</td>
<td>5.4196</td>
<td>4.062</td>
<td>7.504</td>
<td>3.442</td>
</tr>
</tbody>
</table>
Table 2.15: Energy differences between the ground and CT states for 41 frames along the MD trajectory calculated on the CT-state surfaces. All values are in eV.

<table>
<thead>
<tr>
<th>System</th>
<th>Acceptor</th>
<th>$\omega$B97X-D/cc-pVDZ</th>
<th>$\langle \Delta E_{CT} \rangle_{CT}$</th>
<th>$\langle \Delta E_{cdftci}\rangle_{CT}$</th>
<th>$\delta E_{corr,CT}$</th>
<th>$\langle \Delta E'<em>{CT} \rangle</em>{CT}$</th>
<th>$\langle \Delta E'<em>{cdftci}\rangle</em>{CT}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFP</td>
<td>145</td>
<td></td>
<td>2.0915</td>
<td>2.1531</td>
<td>-0.062</td>
<td>1.984</td>
<td>2.046</td>
</tr>
<tr>
<td>eYFP</td>
<td>145</td>
<td></td>
<td>3.0921</td>
<td>2.5891</td>
<td>0.503</td>
<td>2.932</td>
<td>2.429</td>
</tr>
<tr>
<td></td>
<td>203</td>
<td></td>
<td>3.5220</td>
<td>2.8993</td>
<td>0.623</td>
<td>3.250</td>
<td>2.627</td>
</tr>
<tr>
<td>eYFP-Y145L</td>
<td>203</td>
<td></td>
<td>3.3647</td>
<td>2.8201</td>
<td>0.545</td>
<td>3.137</td>
<td>2.592</td>
</tr>
<tr>
<td>eYFP + Cl$^-$</td>
<td>145</td>
<td></td>
<td>2.1535</td>
<td>2.1785</td>
<td>-0.025</td>
<td>2.099</td>
<td>2.124</td>
</tr>
<tr>
<td></td>
<td>203</td>
<td></td>
<td>4.7700</td>
<td>3.4647</td>
<td>1.305</td>
<td>4.277</td>
<td>2.972</td>
</tr>
</tbody>
</table>
the QM part, and the rest of the system was described by point charges. The calculations were performed for several snapshots along the equilibrium trajectory at 298 K, unless specified otherwise. We also computed the coupling on the CT surface since they are needed for computing the reverse rate constants. In calculations of the Chrom-Tyr92 couplings, additional mediating residues were included in the QM part, as shown in Fig. 2.15. In these calculations, the mediating residues obtained from the Pathways model calculations were chopped at appropriate positions (no peptide bonds) to keep the system size reasonable.

![Figure 2.15: Residues included in QM in the calculations of the couplings between the chromophore and Tyr92, based on the Pathways model predictions.](image)
Chapter 2 References


Chapter 3: Towards understanding the redox properties of model chromophores from the green fluorescent protein family: An interplay between conjugation, resonance stabilization, and solvent effects

3.1 Introduction

Fluorescent proteins (FPs) from the green fluorescent protein (GFP) family are extensively used in bioimaging as genetically encoded fluorescent labels. Motivated by a variety of exciting applications, a large number of FPs with different properties (color, Stokes shifts, brightness, photostability, phototoxicity, sensitivity to small ions, maturation rates, etc) have been developed covering the entire spectral range. Atomic-level
understanding of their properties is important for engineering new designer FPs better suited for a particular application. This has been motivating extensive experimental and theoretical studies of their optical properties (excitation/emission energies and brightness), mechanistic details of chromophores’ maturation and the photocycle.\textsuperscript{5,10–17}

![Diagram of chromophores](image)

**Figure 3.1:** Chromophores of the selected FPs of different colors: wt-GFP, eGFP (green), TagBFP (blue), EBFP(blue), CFP (cyan), YFP (yellow), DsRed, mCherry (red), mOrange (orange). Absorption/emission wavelengths are given in parenthesis. The chromophores are shown in colors corresponding to their fluorescence.

Fig. 3.1 shows selected chromophores of FPs of different color (the colors refer to fluorescence). The structural motifs of color tuning are rather diverse\textsuperscript{17} and include chemical modifications of the chromophore such as extension of the \( \pi \)-system in the red
FPs, $\pi$-stacking and electrostatic interactions with the neighboring residues, as well as protonation-deprotonation equilibria. In stark contrast to their optical properties, relatively little is known about the redox properties and ionized/electron-detached states of these biomolecules.\textsuperscript{18–20} The interest in these properties stems from the recent discovery that FPs can act as light-induced electron donors.\textsuperscript{21} Redox-sensitive FPs can be used for \textit{in vivo} measurements of the mitochondrial redox potential.\textsuperscript{22, 23} The focus of this work is on understanding the effect of structure of the chromophores on their redox properties. Better understanding of structure-function relationship can be used to develop novel fluorescent probes suited to new types of applications of genetically encoded FPs.

The properties of FPs are determined by the chemical structure of their chromophores and by the interactions of the chromophores with the surrounding protein. Following a bottom up approach, we begin by investigating the redox properties of the model chromophores in gas phase and in simple solvents. These calculations allow us to quantify the intrinsic electron-donating ability of the chromophores and to make inroads into understanding how the redox properties of the chromophores can be modulated by the environment (such as solvent and the nearby protein residues). Properties of model chromophores in gas-phase and simple solvents provide an important benchmark and can be measured experimentally.\textsuperscript{24–26} Theoretical prediction of the low electron-detachment energy of the anionic form of the model GFP chromophore,\textsuperscript{18, 27} which suggested a metastable character of the bright excited state, has stimulated several experimental studies aiming at determining DE of this system.\textsuperscript{28–30} With the exception of cyan FP,\textsuperscript{29} the DEs of other isolated anionic chromophores have not yet been characterized. The first experimental measurement of the redox potential of neutral model GFP chromophores in solution has been recently reported.\textsuperscript{20} This study demonstrated
that the electron-donating ability of the chromophores can be modulated by varying resonance stabilization via structural modifications. The computational studies have helped to quantify solvent effects.\textsuperscript{20,31} The redox potential of the protein-bound chromophore (eGFP) has only been characterized computationally.\textsuperscript{19}

The electron donating ability of the chromophores depends on several delicately balanced factors, such as the size of the $\pi$-system, resonance stabilization of the charge distribution, electronegativity of the atoms comprising the chromophore, and the presence of electron donating/withdrawing substituents, as well as solvent effects.

To illustrate these competing factors, consider homologically similar compounds of increasing size such as conjugated dyes or aromatic clusters. Using particle-in-the-box reasoning, one may anticipate that the energy levels (i.e., molecular orbitals) will be lowered in larger systems resulting in red-shifted absorption and decrease in ionization energy (IE). In the same-size systems, energy can be lowered by resonance stabilization. Since the energetic consequences of delocalization are larger for charged systems, size increase and resonance stabilization have the opposite effect on electron ejection energies from neutral and anionic species. For example, the IEs of the neutral naphthalene clusters decrease with system size (8.14, 7.58, 7.56, 7.49 eVs for (Nph)$_n$, $n=1-4$),\textsuperscript{32} whereas the detachment energies (DEs) of the anionic naphthalene clusters increase with the system size (-0.18, 0.11, 0.28, 0.48, 0.62 eVs for (Nph)$_n^-$, $n=1-5$).\textsuperscript{33}

This trend is illustrated in Fig. 3.2 for the electron-ejection processes from the neutral and anionic species:

$$\text{HA} \rightarrow \text{HA}^+ (\text{radical} \ - \ \text{cation}) + e^-, \ \text{IE} \quad (3.1)$$

$$\text{A}^- (\text{deprotonated}) \rightarrow \text{A}^\cdot (\text{neutral} \ - \ \text{radical}) + e^-, \ \text{DE} \quad (3.2)$$
Figure 3.2: The effect of resonance stabilization of energetics of electron ejection from the neutral (left) and anionic (right) species. Since the resonance stabilization is always greater for charged species, more extensive resonance interactions lead to ionization energy decrease in the neutral species and to electron-detachment energy increase in anions.

Here $H^+$ denotes neutral (protonated) species, such as neutral forms of the chromophores, whereas $A^-$ denotes closed-shell anionic chromophores derived by the deprotonation of the respective neutral species. In the case of HA ionization, the $H^+$ is strongly stabilized by resonance leading to the IE decrease with increasing resonance stabilization, whereas in the second reaction, the $A^-$ is more stabilized by resonance than $A$ leading to the DE increase.

Of course, the above considerations are valid only in homologically similar compounds. The IEs/DEs of iso-electronic species will be strongly modulated by the relative electronegativity of the constituent atoms and the presence of electron donating/withdrawing groups. For example, the effect of electronegativity of the heteroatoms can be illustrated by phenyl halides for which the IEs decrease on going from fluorine to iodine. Finally, solvent will also affect energetics of the redox reactions, Eq. (3.1) and Eq. (3.2). Solvent is expected to stabilize the charged species; more extensive resonance interactions leading to more delocalized charge are expected to reduce solvent stabilization. Thus, the effect of resonance stabilization on IEs/DEs will be offset by including solvent effects. In the previous study of the redox properties of model FP
chromophores, the trends in redox potentials were dominated by IEs of isolated chromophores, however, in the present study we observe that solvent can actually reverse the trends based on IEs.

In this work, we investigate the effect of the chromophore structure on the redox properties of model chromophores representing green (eGFP, wt-GFP), red (DsRed), and blue (mTagBFP) FPs (see Fig. 3.1). Our aim is to quantify the competing factors described above laying out the foundation for developing qualitative models that can be used to rationalize and predict the trends in the redox properties based on the sizes of chromophores, resonance stabilization, and presence of heteroatoms. This is a prerequisite for future studies of the redox properties of the protein-bound chromophores.

This study focuses on the ground-state redox properties of FPs. The redox potentials of electronically excited chromophores, which are of interest in the context of light-induced electron-donating FPs, can be estimated by using the following relationship between ground- and excited-state IEs:

$$IE^{\text{ex}} \approx IE^{\text{gs}} - E_{\text{ex}}$$

(3.3)

where $IE^{\text{ex}}$ is the IE of the electronically excited chromophore, $IE^{\text{gs}}$ is the IE of the ground state, and $E_{\text{ex}}$ is the excitation energy. For example, the computed redox potential of eGFP is 0.55 V. Using computed vertical excitation energy of 2.70 eV, we arrive at $E^0 \approx -2.15$ V for electronically excited eGFP. This is a lower-bound estimate, as it does not include relaxation of the chromophore and its protein environment in the electronically excited state.

Different protonation forms of the chromophores may exist in the protein and, especially, in solvents. For the model GFP chromophore, 4 different forms shown in Fig. 3.3 have been considered, i.e., neutral, anionic (deprotonated phenolic moiety), cationic
(protonated imidazolinone), and zwitterionic. Since the neutral and anionic states appear to be most relevant to the FP photocycle, we focus on these two forms of all model chromophores. We denote the deprotonated forms by '-D'. Experiments carried out at different pH in water can give rise to chromophores in different protonation states.42

![Figure 3.3: Different protonation states of the GFP model chromophore](image)

The model molecules representing the green, red, and blue chromophores are: (i) 4-hydroxybenzylidene-1,2-dimethylimidazolinone (HBDI), (ii) 4-hydroxybenzylidene-1-methyl-2 penta-1,4-dien-1-yl-imidazolin-5-one (HBMPDI), and (iii) N-[(5-hydroxy-1H-imidazole-2yl)methyl-methylidene]acetamide (HIMA) and N-[(5-hydroxy-1H-imidazole-2yl) methylimidene]acetamide (HHIMA), respectively. The structures of their deprotonated forms are shown in Fig. 3.4

The structure of the paper is as follows. Section 3.2 gives computational details. Section 3.3 presents our results and discussion of the gas-phase energetics (section 3.3.1), solvent effects (section 3.3.2), and the redox potentials (section 3.3.3) of the model compounds. Section 3.4 presents our concluding remarks.
Figure 3.4: The structures of the model chromophores (deprotonated forms) and atom labeling scheme. The chromophores consist of the green (phenol), pink (bridge), blue (imidazolinone) and red (acylimine) moieties. Panel (d) gives the atom labeling scheme: “p”, “b”, “i”, and “a” denote phenol, bridge, imidazolinone, and acylimine, respectively.
3.2 Computational details

The structures of the model chromophores (see Fig. 3.4) were optimized using RI-MP2/cc-pVTZ. Since MP2 is not reliable for open-shell species, the ionized species were optimized by density functional theory (DFT) with the ωB97X-D functional\(^{43}\) and the cc-pVTZ basis set. The Cartesian geometries and relevant energies are given in supporting information of Ref. 44. The optimized structures were used for calculation of IEs and DEs with ωB97x-D. Two basis sets were employed: 6-311++G(2df,2pd) and 6-31+G(d). Zero point energy (ZPE) corrections to adiabatic values as well as other thermodynamic corrections were computed by ωB97x-D/cc-pVTZ at the respective optimized geometries. In addition, IEs/DEs were calculated using equation-of-motion coupled-cluster method with single and double substitutions for ionization potentials (EOM-IP-CCSD)\(^{45–49}\) for comparison, in particular, to check for potential artifacts in the computed trends due to remaining self-interaction error. The EOM-IP-CCSD calculations were performed with the 6-31+G(d) basis set. Based on our recent calculations of phenol and phenolate,\(^{50}\) we anticipate 0.1-0.3 eV differences between ωB97X-D and EOM-IP-CCSD. The estimated error bars for the IE/DE values computed with ωB97X-D are ≈0.1 eV.\(^{51}\)

Natural bond orbital (NBO)\(^{52}\) analysis of charges and spin densities was carried out to understand the structure function correlations. The solvation free energies were computed using a continuum solvation model, SM8,\(^{53}\) and the 6-31+G(d,p) basis set. The free energies of the redox reactions were calculated by constructing thermodynamic cycles as explained in Section 3.3.3.

The main cause of error in the computed redox potentials is due to the calculation of solvation free energies with implicit solvation methods. A conservative estimate for
the error bars of the solvation free energy is $\approx 0.4$ V based on the recent benchmark studies$^{54,55}$ Explicit solvation methods with polarization effects can be used to calculate the free energy changes with higher accuracy. For example, hybrid quantum mechanical/effective fragment potential (EFP) approach has shown errors of $\approx 0.05$-0.1 eV with respect to high-level ab initio methods such as EOM-IP-CCSD$^{50,56}$. However, these methods require extensive sampling which is computationally demanding for large chromophores.

All calculations were carried out using Q-Chem$^{57}$.

3.3 Results and discussion

3.3.1 Ionization and electron detachment energies of the isolated chromophores

Table 3.1 shows the vertical and adiabatic IEs/DEs (VIE/VDE and AIE/ADE, respectively) of the model blue (HIMA and HHIMA), green (HBDI), and red (HBMPDI) chromophores. For comparison, we also present energies for phenol and phenolate$^{50}$. We have also tabulated the energies calculated using EOM-IP-CCSD. The IEs/DEs calculated by EOM and DFT methods follow similar trends, which allows us to validate that the DFT results for the chromophores of different sizes are not affected by remaining self-interaction error. We notice that the difference is relatively small for the neutral species ($\sim 0.1$ eV); and is about 0.3 eV for the anionic ones. Our previous study of phenol/phenolate$^{50}$ suggests that EOM-IP-CCSD underestimates the DEs of anionic species, e.g., the errors for VDE of phenolate were 0.3 eV (using cc-pVTZ and aug-cc-pVTZ). We note that the $\omega$B97X-D/6-311(+(+,+)G(2pd,2df) value for phenolate (see Table
Table 3.1: Vertical and adiabatic ionization/detachment energies (eV) of the model FP chromophores and phenolic species\(^a\).

<table>
<thead>
<tr>
<th>Species</th>
<th>Koopmans(^b)</th>
<th>VIE/VDE</th>
<th>VIE/VDE</th>
<th>AIE/ADE</th>
<th>AIE/ADE w/ZPE</th>
<th>(\Delta G_g)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(EOM-IP)(^c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIMA</td>
<td>8.00</td>
<td>7.70</td>
<td>7.64</td>
<td>7.36</td>
<td>7.32</td>
<td>7.29</td>
</tr>
<tr>
<td>HHIMA</td>
<td>8.08</td>
<td>7.83</td>
<td>7.77</td>
<td>7.51</td>
<td>7.50</td>
<td>7.48</td>
</tr>
<tr>
<td>HBDI</td>
<td>7.59</td>
<td>7.38</td>
<td>7.33</td>
<td>7.15</td>
<td>7.15</td>
<td>7.13</td>
</tr>
<tr>
<td>HBMPDI</td>
<td>7.94</td>
<td>7.64</td>
<td>7.59</td>
<td>7.35</td>
<td>7.35</td>
<td>7.31</td>
</tr>
<tr>
<td>Phenol</td>
<td>8.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIMA-D</td>
<td>2.97</td>
<td>2.75</td>
<td>2.43</td>
<td>2.33</td>
<td>2.35</td>
<td>2.35</td>
</tr>
<tr>
<td>HHIMA-D</td>
<td>3.07</td>
<td>2.90</td>
<td>2.58</td>
<td>2.61</td>
<td>2.63</td>
<td>2.59</td>
</tr>
<tr>
<td>HBDI-D</td>
<td>2.94</td>
<td>2.79</td>
<td>2.48</td>
<td>2.67</td>
<td>2.67</td>
<td>2.66</td>
</tr>
<tr>
<td>HBMPDI-D</td>
<td>3.45</td>
<td>3.27</td>
<td>3.01</td>
<td>3.15</td>
<td>3.15</td>
<td>3.11</td>
</tr>
<tr>
<td>Phenolate</td>
<td>2.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) \(\omega^\text{B97x-D/6-311++G(2df,2pd)}\)

\(^b\) Hartree-Fock HOMO energy, 6-311++G(2df,2pd)

\(^c\) EOM-IP-CCSD/6-31+G(d)

\(^d\) From Ref. 50, EOM-IP-CCSD/cc-pVTZ
3.1) is in much better agreement with the experimental VDE of 2.36 eV.\textsuperscript{58,59} Recent experiments\textsuperscript{29,30} have reported 2.8 eV VDE for gas-phase HBDI-D, which is close to the computed $\omega$B97X-D/6-311(+(+))G(2pd,2df) value (Table 3.1), but is about 0.3 eV higher than the EOM-IP value from Table 3.1 and, consequently, previously reported theoretical estimate\textsuperscript{18,27} derived using the EOM-IP based energy additivity scheme. Thus, in this work we rely on $\omega$B97X-D/6-311(+(+))G(2pd,2df) for DEs/IEs calculations. The EOM-IP values are used to validate that the differences between the chromophores of different sizes are not affected by remaining self-interaction error.

Our best estimates (in eVs) for VIE/VDEs are 7.38/2.79 (HBDI), 7.64/3.27 (HBMPDI), 7.83/2.90 (HHIMA) and 7.70/2.75 (HIMA) for the neutral/deprotonated forms. The best estimates of the respective adiabatic values (AIE/ADEs) are 7.15/2.67 (HBDI), 7.35/3.15 (HBMPDI), 7.50/2.63 (HHIMA) and 7.32/2.35 (HIMA) for the protonated and deprotonated forms.

As discussed above, we expect that resonance stabilization will have opposite effect in the neutral and anionic (deprotonated) species. The leading resonance structures of the anionic chromophores are shown in Fig. 3.5. As one can see, the red chromophore has most extensive resonance stabilization. The comparison between HIMA and HBDI is more complicated due to only partial overlap of their structural frameworks.

Let us first consider trends in anionic species. Among the deprotonated species, phenolate has the lowest DE (1.99 eV). The VDE of HBMPDI-D, which is the largest system, is the highest (3.27 eV) due to more extensive resonance stabilization than both HIMA-D and HBDI-D anions. The VDEs of HIMA-D and HBDI-D are 2.75 eV and 2.79 eV, respectively. HIMA-D (methylated species) has lower DE than HHIMA-D (non-methylated) due to the electron-donating methyl group. Interestingly, despite sizable differences in VDEs of phenolate, HBDI-D, and HBMPDI-D, the analysis of spin
densities (see section 3.3.1) reveals that electron detachment occurs predominantly from the phenolate moiety (see Table 3.3). In order to further support our theory, we have calculated the VIEs of the ortho, meta, and para isomers of deprotonated HBDI. We see the opposite trend from its protonated counterparts\textsuperscript{20} as expected.

Among the neutral species, phenol has the highest IE (1 eV higher than that of HBDI), as expected. The difference between HIMA and HHIMA is again due to electron-donating methyl group. However, we note that HBDI has lower IE than HBMPDI, contrary to the trend illustrated in Fig. 3.2. This can be rationalized by close
inspection of the structural parameters summarized in Table 3.2 revealing that the difference in resonance stabilization in HBDI and HBMPDI is larger in the anionic forms (relative to HBDI$^+$ and HBMPDI$^+$). In the case of ionized (cationic) HBDI and HBMPDI, the degree of resonance stabilization involving the phenol, bridge, and imidazolinone moieties appears to be very similar, judging by the similarity of the $C_p$-$C_b$ and $C_t$-$C_b$ bond lengths in HBDI$^+$ and HBMPDI$^+$ (see Table 3.2). This is further supported by the charges on $O_p$ and $O_t$ (-0.58 and -0.47 in HBDI$^+$ and HBMPDI$^+$). The IE values of HBDI and HBMPDI are determined by the two competing effects, more extended resonance stabilization due to acylimine and electron-withdrawing character of this moiety (which contains several electro-negative atoms) and based on the computed IE values, the latter appears to be more important in this case.

**Quantifying resonance interaction by structural analysis**

The degree of resonance interactions in these species can be quantified by the representative geometric parameters collected in Table 3.2 (see Fig. 3.4 for atom labeling scheme).

Comparing the $C_p$-$C_b$ and $C_t$-$C_b$ bond lengths in reduced/oxidized HBDI and HBMPDI, we see that the bond length alternation is lower in the case of the ionized form of the neutral (HA$^+$) and the anionic form of the deprotonated (A$^-$) species, as expected.

We define $\Delta = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (r_i - \bar{r})^2}$, where $\bar{r}$ is the average bond lengths of a given type (e.g., C-N bond in imidazolinone), which quantifies the degree of bond length alternation and therefore, resonance stabilization. Likewise, $\sigma(C_s-N_s)$ is defined as the difference between the two $C_s$-$N$ bonds in acylimine. Comparing $\Delta$s computed for the phenolate and imidazolinone rings shows that HBMPDI$^+$ exhibits a similar degree of resonance stabilization as HBDI$^+$. HBMPDI-D has lower $\Delta$(imidazolinone) than
HBDI-D, while respective $\Delta$(phenolate) are similar. This is because the imidazolinone in HBMPDI-D is further stabilized due to the additional resonance structure (see Fig. 3.5).

We further analyze the degree of delocalization and resonance stabilization by comparing the relevant NBO charges and spin densities. The molecules can be divided into different parts as shown in Fig. 3.4. Table 3.3 shows the spin densities ($\rho_\alpha - \rho_\beta$) on the different moieties of the oxidized chromophores quantifying the location of the unpaired electron.

In the case of both forms of HBDI and HBMPDI, we observe that the ionization involves both the phenol and imidazolinone moieties, with phenol/phenolate playing the leading role in deprotonated species (the imidazolinone hosts a larger fraction of the spin-density in HBDI$^+$ and HBMPDI$^+$). The spin densities on acylimine are smaller in the case of HBMPDI/HBMPDI-D than in HHIMA/HHIMA-D. Therefore, acylimine plays a less important role in the red chromophore than in the blue one.

The highest occupied molecular orbitals (HOMOs) are shown in Fig. 3.6. Comparing the HOMOs for HBMPDI and HBMPDI-D we note that there is less electron density on the C$_r$-C$_s$ bond and acylimine in the case of the protonated species. The HOMO of HBMPDI is similar to that of HBDI — it is delocalized over phenol and imidazolinone, but does not have much density on acylimine. However, the HOMO of HBDI-D is somewhat different from that of HBMPDI-D showing different degree of delocalization.

### 3.3.2 Solvent effects

Table 3.4 shows solvation free energies for the relevant species as well as $\Delta\Delta G_{solv}$ in acetonitrile, the solvent contribution to the free energy of the oxidation reactions, Eqns.
Table 3.2: Selected geometric parameters (Å) of the model chromophores and the respective oxidized species\(^a\).

<table>
<thead>
<tr>
<th>Species</th>
<th>(C_p)-(C_b)</th>
<th>(C_i)-(C_b)</th>
<th>(O_p)-(C_p)</th>
<th>(O_i)-(C_i)</th>
<th>(\sigma(C_s\text{-}N_s))</th>
<th>(\Delta(N_i\text{-}C_i))</th>
<th>(\Delta(C_p\text{-}C_p))</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIMA</td>
<td>n.a.</td>
<td>1.490</td>
<td>n.a.</td>
<td>1.353</td>
<td>0.127</td>
<td>0.014</td>
<td>n.a.</td>
</tr>
<tr>
<td>HIMA(ionized)</td>
<td>n.a.</td>
<td>1.476</td>
<td>n.a.</td>
<td>1.297</td>
<td>0.171</td>
<td>0.029</td>
<td>n.a.</td>
</tr>
<tr>
<td>HIMA-D</td>
<td>n.a.</td>
<td>1.487</td>
<td>n.a.</td>
<td>1.245</td>
<td>0.045</td>
<td>0.021</td>
<td>n.a.</td>
</tr>
<tr>
<td>HIMA-D(ionized)</td>
<td>n.a.</td>
<td>1.477</td>
<td>n.a.</td>
<td>1.211</td>
<td>0.145</td>
<td>0.026</td>
<td>n.a.</td>
</tr>
<tr>
<td>HBDI</td>
<td>1.435</td>
<td>1.349</td>
<td>1.354</td>
<td>1.209</td>
<td>n.a.</td>
<td>0.049</td>
<td>0.008</td>
</tr>
<tr>
<td>HBDI(ionized)</td>
<td>1.397</td>
<td>1.390</td>
<td>1.319</td>
<td>1.197</td>
<td>n.a.</td>
<td>0.004</td>
<td>0.028</td>
</tr>
<tr>
<td>HBDI-D</td>
<td>1.393</td>
<td>1.385</td>
<td>1.248</td>
<td>1.236</td>
<td>n.a.</td>
<td>0.050</td>
<td>0.040</td>
</tr>
<tr>
<td>HBDI-D(ionized)</td>
<td>1.413</td>
<td>1.369</td>
<td>1.228</td>
<td>1.209</td>
<td>n.a.</td>
<td>0.038</td>
<td>0.046</td>
</tr>
<tr>
<td>HBMPDI</td>
<td>1.447</td>
<td>1.345</td>
<td>1.351</td>
<td>1.208</td>
<td>0.132</td>
<td>0.045</td>
<td>0.009</td>
</tr>
<tr>
<td>HBMPDI(ionized)</td>
<td>1.395</td>
<td>1.393</td>
<td>1.316</td>
<td>1.197</td>
<td>0.153</td>
<td>0.004</td>
<td>0.029</td>
</tr>
<tr>
<td>HBMPDI-D</td>
<td>1.385</td>
<td>1.394</td>
<td>1.234</td>
<td>1.226</td>
<td>0.093</td>
<td>0.033</td>
<td>0.045</td>
</tr>
<tr>
<td>HBMPDI-D(ionized)</td>
<td>1.404</td>
<td>1.377</td>
<td>1.226</td>
<td>1.208</td>
<td>0.134</td>
<td>0.028</td>
<td>0.048</td>
</tr>
</tbody>
</table>

\(^a\) \(\sigma(C_s\text{-}N_s)\) and \(\Delta\) quantify the degree of bond length alternation in acylimine and phenolate/imidazolinone moieties, respectively (see text). Atom labeling scheme is given in Fig. 3.4d.
Figure 3.6: The HOMOs of the model chromophores.
Table 3.3: Mulliken analysis of spin densities in the oxidized species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Phenol</th>
<th>Imidazolinone</th>
<th>Bridge</th>
<th>Acylimine</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBDI</td>
<td>0.41</td>
<td>0.59</td>
<td>0.00</td>
<td>n.a.</td>
</tr>
<tr>
<td>HBMPDI</td>
<td>0.39</td>
<td>0.60</td>
<td>-0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>HHIMA</td>
<td>n.a.</td>
<td>0.86</td>
<td>0.02</td>
<td>0.12</td>
</tr>
<tr>
<td>HBDI-D</td>
<td>0.73</td>
<td>0.51</td>
<td>-0.24</td>
<td>n.a.</td>
</tr>
<tr>
<td>HBMPDI-D</td>
<td>0.68</td>
<td>0.52</td>
<td>-0.25</td>
<td>0.05</td>
</tr>
<tr>
<td>HHIMA-D</td>
<td>n.a.</td>
<td>0.80</td>
<td>0.00</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Table 3.4: Free energies of solvation (kcal mol\(^{-1}\)) in acetonitrile for the model chromophores.

<table>
<thead>
<tr>
<th>Species</th>
<th>(\Delta G_{\text{red}})</th>
<th>(\Delta G_{\text{ox}})</th>
<th>(\Delta \Delta G_{\text{solv}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIMA</td>
<td>-11.85</td>
<td>-57.65</td>
<td>-45.80</td>
</tr>
<tr>
<td>HHIMA</td>
<td>-12.05</td>
<td>-58.01</td>
<td>-45.96</td>
</tr>
<tr>
<td>HBDI</td>
<td>-15.33</td>
<td>-52.52</td>
<td>-37.19</td>
</tr>
<tr>
<td>HBMPDI</td>
<td>-16.95</td>
<td>-54.63</td>
<td>-37.68</td>
</tr>
<tr>
<td>HIMA-D</td>
<td>-52.52</td>
<td>-10.78</td>
<td>+41.74</td>
</tr>
<tr>
<td>HHIMA-D</td>
<td>-53.12</td>
<td>-9.91</td>
<td>+43.21</td>
</tr>
<tr>
<td>HBDI-D</td>
<td>-57.87</td>
<td>-15.56</td>
<td>+42.31</td>
</tr>
<tr>
<td>HBMPDI-D</td>
<td>-51.63</td>
<td>-16.79</td>
<td>+34.84</td>
</tr>
</tbody>
</table>

(3.1) and (3.2). For most of the species, the \(\Delta \Delta G_{\text{solv}}\)s in acetonitrile follow an opposite trend relative to IE/DEs. This is because the solvent stabilization is larger for more localized charges. Thus, the greater is the stability of the charged species (due to charge delocalization), the lower is its solvation free energy. For example, charge distribution shows that HBMPDI-D has the most charge delocalization and HHIMA\(^+\) has the least charge delocalization (among the charged species). They have the lowest and highest \(\Delta G_{\text{solv}}\), respectively, the trend which is carried over to the \(\Delta \Delta G_{\text{solv}}\)s.

We observe similar trends for solvation energies in water (Table 3.5). We also note that difference between \(\Delta \Delta G_{\text{solv}}\) is very similar for acetonitrile and water in the case of
Table 3.5: Free energies of solvation (kcal mol\(^{-1}\)) in water for the model chromophores.

<table>
<thead>
<tr>
<th>Species</th>
<th>(\Delta G_{\text{red}})</th>
<th>(\Delta G_{\text{ox}})</th>
<th>(\Delta \Delta G_{\text{solv}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIMA</td>
<td>-10.96</td>
<td>-56.82</td>
<td>-45.86</td>
</tr>
<tr>
<td>HHIMA</td>
<td>-12.22</td>
<td>-58.06</td>
<td>-45.84</td>
</tr>
<tr>
<td>HBDI</td>
<td>-12.06</td>
<td>-47.65</td>
<td>-35.59</td>
</tr>
<tr>
<td>HBMPDI</td>
<td>-13.24</td>
<td>-49.50</td>
<td>-36.26</td>
</tr>
<tr>
<td>HIMA-D</td>
<td>-56.03</td>
<td>-8.07</td>
<td>+47.96</td>
</tr>
<tr>
<td>HHIMA-D</td>
<td>-57.82</td>
<td>-7.95</td>
<td>+49.87</td>
</tr>
<tr>
<td>HBDI-D</td>
<td>-60.01</td>
<td>-11.22</td>
<td>+48.79</td>
</tr>
<tr>
<td>HBMPDI-D</td>
<td>-52.68</td>
<td>-11.80</td>
<td>+40.88</td>
</tr>
</tbody>
</table>

Figure 3.7: Thermodynamic cycle

neutral species \((\approx 0.77\ \text{kcal mol}^{-1}\ \text{difference on average})\), but is somewhat shifted for the deprotonated species \((\approx 6.35\ \text{kcal mol}^{-1}\ \text{difference on average})\).
3.3.3 Redox potentials

From the energetics of ionization/electron-detachment and solvation processes, we can construct a thermodynamic cycle (Fig. 3.7) using Hess’s law:

\[
\Delta G_{\text{rxn}} = \Delta G_g + (\Delta G_{\text{ox}} - \Delta G_{\text{red}})
\]

\[
E_{\text{ox}}^0 = -\frac{\Delta G_{\text{rxn}}}{nF}
\]

where \(n\) is number of electrons involved in the redox reaction and \(F\) is Faraday’s constant. We note that gas-phase Gibbs free energy changes of the oxidation reactions, \(\Delta G_g\), are very close to ADEs (the differences do not exceed 0.04 eV, see Tables 3.1 and 3.9). Using this equation, we calculated the \(E_{\text{ox}}^0\) with respect to standard hydrogen electrode (SHE). Here, we have taken the \(\Delta G\) of SHE to be the recent value of -4.281 V.\(^6\) The reported values can easily be converted to the potentials relative to more commonly used reference electrodes, e.g., a ferrocene couple (Fc\(^+\)/Fc) used in Ref. 20. The calculated redox potentials of the FP model chromophores in acetonitrile and water are given in Table 3.6. Experimentally, anionic chromophores can only be prepared in water (at high pH), thus, the computed \(E^0\) in acetonitrile cannot be verified experimentally. However, these values are useful for theoretical analysis, as they allow us to compare neutral versus anionic chromophores in the same non-protic solvent, and to quantify the effect of solvent polarity on different species. Based on the previous study,\(^20\) the errors in absolute values of the \(E^0\)'s computed using this protocol were around 0.2 V; however, the differences between different chromophores were reproduced by theory more accurately (maximum error of 0.08 V). Thus, the differences in computed \(E^0\) for different chromophores are larger than anticipated error bars of the method employed.
The trends in redox potentials of the chromophores are dominated by IEs/DEs. However, since solvent stabilization ($\Delta \Delta G_{solv}$) follows an opposite trend from the IEs/DEs, it offsets the differences in IEs/DEs and can even reverse the overall energetics when the differences in IEs/DEs are small. Consequently, the variations in the redox potentials are smaller relative to the differences in the respective IEs/DEs. This also follows from the empirical equations for the calculation of redox potentials from VIEs.\(^6\)

We note that the redox potentials of aqueous HBDI and HBMPDI are close to $E^0$ of phenol (1.32 V);\(^5\) however, the potentials for the respective anionic species are somewhat lower than for phenolate (0.89 V, Ref. 50).

In our previous work on the ortho, meta, and para isomers of HBDI,\(^2\) we observed that the trend in redox potentials is dominated by the variations of IEs, since solvent effects for structurally similar chromophores are similar. In the present study, however, the chromophores are significantly different and have different solvation free energies. Because of these two opposing effects, the trends in the redox potentials sometimes differ from the IE/DE predictions, e.g., the redox potential of the neutral blue chromophore is lower than that of the red and green chromophores, although the IEs of the red and green
chromophores are lower than those of the blue one. Therefore, both the IEs/DEs as well as effects of solvation are important for understanding the trends in the redox potentials. Moreover, the observed solvent effects suggest that protein environment can strongly modulate the redox properties of the protein-bound chromophore by electrostatic interactions. For example, one can anticipate different redox potentials for families of FPs sharing the same chromophore but having different local environment. These effects will be investigated in future studies. As of today, the only available estimate of $E^0$ of a protein-bound chromophore is for eGFP. The value reported in Ref. 19 (0.47 V) was computed using $\Delta G=-4.36$ V for SHE. Thus, the corrected value using more recent value of SHE (-4.281 V), we arrive to 0.55 V. The computational protocol used in Ref. 19 was rather crude suggesting error bars of about 0.1-0.2 V. Within these error bars, the computed $E^0$ of the protein-bound anionic green chromophore is indistinguishable from $E^0$ of HBDI-D in aqueous solution. Thus, although protein as a whole is less polar than water, the nearby charged groups (such as argenine) provide strong stabilizing effect for the anionic chromophore making its electron-donating ability comparable to that of isolated chromophores in aqueous solutions. The comparison of the acetonitrile value and $E^0$ of the protein-bound chromophore shows that protein environment provides stronger stabilizing effects as compared to acetonitrile. Based on these comparisons, one can use the $E^0$ of a chromophore in water and acetonitrile as a very crude estimate bracketing the redox potential of a protein-bound chromophore. However, more data on the redox properties of FPs and their bare chromophores is necessary to understand the range of the effect of protein environment on $E^0$. 

98
3.4 Conclusions

We performed detailed computational study of the electron-donating abilities of the three model chromophores representing green, red, and blue FPs. The calculations reveal that the energetics of ionization/electron detachment processes invokes a delicate balance between resonance stabilization and electronegativity considerations. The main trends in IEs/DEs can be explained by the charge stabilization due to extended resonance. Since the effect of resonance stabilization is more important in charged species, the respective energetics follows opposite trends in the neutral and anionic chromophores. However, this trend can be offset by electronegativity of atoms comprising the chromophores. Somewhat counter-intuitively, the red chromophore has higher DE/IE than the green chromophore.

The solvation free energies follow the opposite trends than IEs/DEs. The redox potentials are predominantly driven by IEs/DEs; however, the difference in redox potentials between the species is much smaller than gas-phase energetics would imply. Moreover, solvent effects can even reverse the trend based on IEs/DEs. Thus, protein environment is expected to have significant effect on the redox properties of the chromophores.
3.5 Appendix A: Ionization energies calculated at different levels of theory

The ionization/detachment energies of the model FP chromophores calculated with DFT/$\omega$B97x-D and EOM-IP-CCSD are given in Table 3.7. The basis sets used are 6-31+G(d) and 6-311++G(2df,2pd). The EOM-IP-CCSD-DFT difference (IE$_{\text{EOM}}$-IE$_{\text{DFT}}$) is computed as the difference between the EOM-IP-CCSD and DFT/$\omega$B97x-D values in the 6-31+G(d) basis set.

Table 3.7: Vertical ionization energies (eV) calculated with $\omega$B97x-D and EOM-IP-CCSD

<table>
<thead>
<tr>
<th>Species</th>
<th>EOM-IP-CCSD</th>
<th>$\omega$B97x-D</th>
<th>IE$<em>{\text{EOM}}$-IE$</em>{\text{DFT}}$</th>
<th>$\omega$B97x-D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6-31+G(d)</td>
<td>6-31+G(d)</td>
<td>6-31+G(d)</td>
<td>6-311++G(2df,2pd)</td>
</tr>
<tr>
<td>HIMA</td>
<td>7.64</td>
<td>7.73</td>
<td>-0.09</td>
<td>7.70</td>
</tr>
<tr>
<td>HHIMA</td>
<td>7.77</td>
<td>7.87</td>
<td>-0.10</td>
<td>7.83</td>
</tr>
<tr>
<td>HBDI</td>
<td>7.33</td>
<td>7.39</td>
<td>-0.06</td>
<td>7.38</td>
</tr>
<tr>
<td>HBMPDI</td>
<td>7.59</td>
<td>7.64</td>
<td>-0.07</td>
<td>7.64</td>
</tr>
<tr>
<td>HIMA-D</td>
<td>2.43</td>
<td>2.74</td>
<td>-0.31</td>
<td>2.75</td>
</tr>
<tr>
<td>HHIMA-D</td>
<td>2.58</td>
<td>2.89</td>
<td>-0.31</td>
<td>2.90</td>
</tr>
<tr>
<td>HBDI-D</td>
<td>2.48</td>
<td>2.75</td>
<td>-0.27</td>
<td>2.79</td>
</tr>
<tr>
<td>HBMPDI-D</td>
<td>3.01</td>
<td>3.25</td>
<td>-0.24</td>
<td>3.27</td>
</tr>
</tbody>
</table>
3.6 Appendix B: Thermodynamic data used to compute Gibbs free energy for oxidation reaction in the gas phase

Entropy contribution (at T=298 K) and enthalpy corrections to the Gibbs free energy of the gas-phase oxidation reaction are summarized below. The thermodynamic corrections are calculated within the rigid rotor harmonic oscillator (RRHO) approximation using $\omega_{B97X-D/cc-pVTZ}$ frequencies computed at the respective optimized geometries. The enthalpy corrections ($\Delta H_{corr}$) are an order of magnitude smaller than the ZPE corrections in most cases. Overall, the $\Delta G$ values computed with the inclusion of $dH$ and $T\Delta S$ corrections differ from adiabatic IE/DE (with ZPE) by less than 0.04 eV.

**Table 3.8: Entropy change for gas-phase oxidation reaction.** The entropies are calculated within the rigid rotor harmonic oscillator (RRHO) approximation. The entropies are in cal/mol·K and $T\cdot \Delta S$ is in kcal/mol.

<table>
<thead>
<tr>
<th>Species</th>
<th>Entropy of reduced species</th>
<th>Entropy of oxidized species</th>
<th>Entropy change ($\Delta S$)</th>
<th>$T\cdot \Delta S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIMA</td>
<td>115.90</td>
<td>118.01</td>
<td>2.11</td>
<td>0.63</td>
</tr>
<tr>
<td>HHIMA</td>
<td>118.35</td>
<td>120.04</td>
<td>1.69</td>
<td>0.51</td>
</tr>
<tr>
<td>HBDI</td>
<td>121.62</td>
<td>123.21</td>
<td>1.59</td>
<td>0.48</td>
</tr>
<tr>
<td>HBMPDI</td>
<td>133.22</td>
<td>130.53</td>
<td>2.69</td>
<td>0.81</td>
</tr>
<tr>
<td>HIMA-D</td>
<td>114.24</td>
<td>115.35</td>
<td>1.11</td>
<td>0.33</td>
</tr>
<tr>
<td>HHIMA-D</td>
<td>109.16</td>
<td>112.14</td>
<td>2.98</td>
<td>0.89</td>
</tr>
<tr>
<td>HBDI-D</td>
<td>119.84</td>
<td>121.04</td>
<td>1.20</td>
<td>0.36</td>
</tr>
<tr>
<td>HBMPDI-D</td>
<td>132.28</td>
<td>134.58</td>
<td>2.30</td>
<td>0.69</td>
</tr>
</tbody>
</table>
Table 3.9: Free energy change (eV) of gas-phase oxidation reaction.

<table>
<thead>
<tr>
<th>Species</th>
<th>ADE (with ZPE)</th>
<th>( \Delta H_{corr} )</th>
<th>( T \cdot \Delta S )</th>
<th>( \Delta G )</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIMA</td>
<td>7.32</td>
<td>-0.0046</td>
<td>0.03</td>
<td>7.29</td>
</tr>
<tr>
<td>HHIMA</td>
<td>7.50</td>
<td>0.0043</td>
<td>0.02</td>
<td>7.48</td>
</tr>
<tr>
<td>HBDI</td>
<td>7.15</td>
<td>0.0031</td>
<td>0.02</td>
<td>7.13</td>
</tr>
<tr>
<td>HBMPDI</td>
<td>7.35</td>
<td>-0.0052</td>
<td>0.03</td>
<td>7.31</td>
</tr>
<tr>
<td>HIMA-D</td>
<td>2.35</td>
<td>0.0081</td>
<td>0.01</td>
<td>2.35</td>
</tr>
<tr>
<td>HHIMA-D</td>
<td>2.63</td>
<td>0.0046</td>
<td>0.04</td>
<td>2.59</td>
</tr>
<tr>
<td>HBDI-D</td>
<td>2.67</td>
<td>0.0003</td>
<td>0.01</td>
<td>2.66</td>
</tr>
<tr>
<td>HBMPDI-D</td>
<td>3.15</td>
<td>-0.0104</td>
<td>0.03</td>
<td>3.11</td>
</tr>
</tbody>
</table>

3.7 Appendix C: Resonance structures of the deprotonated chromophores

![Resonance structures](image1)

Figure 3.8: Resonance structures of the deprotonated model GFP chromophore

![Resonance structures](image2)

Figure 3.9: Resonance structures of the deprotonated model blue (TagBFP) chromophore
Figure 3.10: Resonance structures of the deprotonated model RFP chromophore
Chapter 3 References


[37] Subach, O.M.; Cranfill, P.J.; Davidson, M.W.; Verkhusha, V.V. An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore *PLoS One* 2011, 6, e28674.


Chapter 4: Turning on and off photoinduced electron transfer in fluorescent proteins by \( \pi \)-stacking, halide binding, and Tyr145 mutations

4.1 Introduction

Fluorescent proteins (FPs) from the green fluorescent protein family (GFP) enable detailed imaging of processes in live cells and even animals.\(^1\)\(^2\) The GFP chromophore is formed auto-catalytically, upon protein folding, and only requires ambient oxygen; thus the FP sequence can be genetically encoded such that a fluorescent label is produced by an organism along with a protein it is tagging. Hundreds of FPs have been engineered to suit various imaging applications.\(^2\) Among those, enhanced GFP (eGFP) and its yellow variant, eYFP, are considered standard general-purpose FPs.\(^3\)

The GFP chromophore (Fig. 4.1) features a conjugated \( \pi \)-system resembling cyanine dyes. The chromophore is buried inside a tight protein barrel that limits its range of motions and the accessibility to solvent and other species (ions, oxygen, etc). The
Figure 4.1: Structure of eGFP/eYFP. Left: $\beta$-barrel enclosing the chromophore. Right: eGFP and eYFP have the same anionic chromophore formed by cyclization and oxidation of the protein backbone at positions 65-67 (top). In YFPs, the chromophore is $\pi$-stacked with Tyr203 (bottom).

Protection of the barrel is essential for achieving high quantum yields (QY) and photo-stability, as compared to regular dyes. For example, the GFP chromophore in aqueous solutions is non-fluorescent, whereas QY in eGFP is 0.6. Typical QY of bleaching in FPs is $10^{-4}$-$10^{-5}$; it can be as low as $10^{-6}$ in buffered solutions when no oxidants are present. The solution content can strongly affect photostability, even when the dissolved species are too large to penetrate the barrel. For example, oxidized flavines in circa 1 mM decrease photostability of eGFP by up to an order of magnitude both in vitro and in cellulo. Thus, significant changes in protein photo-behavior may occur.
without the direct access to chromophore. Bleaching in some FPs is enhanced in the presence of oxygen and depends on oxygen accessibility to the chromophore.\textsuperscript{11–14}

Multiple excited-state processes competing with fluorescence\textsuperscript{4, 15} are responsible for reduced optical output and bleaching. These include radiationless relaxation, formation of triplet states, photooxidation/photoreduction, and production of reactive oxygen species which can further react with the protein. Importantly, each of these processes can initiate a chemical transformation of the chromophore leading to temporary or permanent loss of fluorescence (i.e., reversible or irreversible bleaching) or change of color (photoconversion). While in many situations these changes are regarded as parasitic processes, they are exploited in other techniques. For example, bleaching and photoswitching are utilized in super-resolution imaging,\textsuperscript{2, 4, 16–18} methods based on fluorescence loss and recovery are used to trace protein dynamics, photoconversions and photoswitching enable optical highlighting and timing of biochemical processes.\textsuperscript{19, 20}

In contrast to dyes, the photoinduced redox processes in FPs are not well understood. They came into a spotlight in 2009, when it was discovered that FPs can be efficient light-induced electron donors.\textsuperscript{21} Bogdanov et al. have reported that many FPs with an anionic GFP chromophore (such as one in Fig. 4.1) undergo photoconversion from green to red form upon irradiation in the presence of oxidants. This process, dubbed oxidative redding, may be exploited in various applications.\textsuperscript{2, 20, 22} Chemical steps leading to the red chromophore formation are initiated by photooxidation, photoinduced electron transfer (ET) from the chromophore to an external oxidant molecule.\textsuperscript{21} Another type of photoconversion (based on the stabilization of the anionic form of the chromophore relative to the protonated neutral one) also involves a photoinduced redox process — photoinduced ET from nearby Glu to the chromophore is believed to be
a gateway step leading to decarboxylation.\textsuperscript{23, 24} Recently, photoreduction of the chromophore was invoked to explain the formation of red-shifted transient species in red FPs.\textsuperscript{25} Photoinduced ET from the anionic chromophores to O$_2$ may lead to superoxide formation, which might be responsible for phototoxicity.\textsuperscript{26, 27} Photoinduced ET coupled with proton transfer has been invoked in the proposed mechanism of bleaching in IrisFP.\textsuperscript{28, 29}

Oxidative redding was observed in various FPs that share the anionic GFP-like chromophore;\textsuperscript{21} later, similar photoconversions were engineered in orange FPs in which the GFP-like chromophore is extended to include a conjugated acylimine tail.\textsuperscript{30, 31} Thus, redding appears to be a robust process characteristic of anionic chromophores that is not very sensitive to the details of the protein environment. No structural information about the red chromophore is available, although several hypotheses were put forward.\textsuperscript{21, 22, 32}

The formation of the red form occurs on second-to-minutes timescale\textsuperscript{21} and is likely to entail significant chemical transformation, such as extension of the conjugated $\pi$-system or breaking of the covalent bonds. These chemical steps are initiated by photoinduced ET from the chromophore (Chro) to an external oxidant molecule. Thus, one can describe redding as an effectively two-step process:

\textbf{Scheme 4.1:} Steps involved in the oxidative redding process leading to final red form.

\[ \text{Chro}^- \xrightarrow{\text{hv}} \text{Chro}^{-*} \xrightarrow{\text{fast, } -1e} \text{Chro}^* \xrightarrow{\text{slow, chemistry}} \text{Red form} \]

The rate-determining step is the second step involving slow chemical changes. The first step is fast, as it is limited by the excited-state lifetime (nanoseconds). It is a gateway step — no redding can occur if there is no ET. The yield of this step provides an upper bound for the yield of the red form.
Here, we investigated three YFPs derived from *A. victoria*: eYFP, Venus, and Citrin.\textsuperscript{7, 33, 34} These YFPs have the same anionic chromophore as eGFP; the change of color is due to $\pi$-stacking of the chromophore with a nearby tyrosine residue (Tyr203, Fig. 4.1). Surprisingly, we found that redding does not occur in these YFPs. However, in eYFP the redding can be turned on by halides, $\text{Cl}^-$, $\text{I}^-$, $\text{Br}^-$, $\text{F}^-$. (eYFP has a halide binding pocket and is used as a halide sensor\textsuperscript{33, 35}). This puzzling finding stimulated theoretical investigations and provided an opportunity to gain an insight into a mechanism of photoinduced ET in FPs. By using molecular dynamics (MD) and quantum-mechanics/molecular mechanics (QM/MM) simulations, we computed Gibbs free energies and electronic couplings for various ET pathways. The simulations suggested that photooxidation of the chromophore proceeds predominantly by hopping mechanism via Tyr145 residue and that Tyr203 affects this major pathway by modulating the ET rate between the chromophore and Tyr145 and by acting as a trap site for ET. The effect on the rate is explained by electronic factors (changes in chromophore’s oxidation potential due to $\pi$-stacking) and structural variations (changes in the Chro-Tyr145 distance). The theoretical predictions were validated by point mutations, which showed that replacing Tyr145 by less efficient electron acceptors results in highly photostable FPs. These results represent the first step towards developing detailed mechanistic understanding of photoinduced ET in FPs and its role in bleaching and photostability.

### 4.2 Experimental and computational details

The experimental measurements were performed as follows. His-tagged proteins were expressed in *E. coli* and purified by a metal-affinity resin. The resin beads with immobilized proteins were placed into phosphate-buffered saline (PBS) with 0.5 mM potassium ferricyanide as an oxidant and illuminated with strong blue light using a fluorescence
microscope. Changes of fluorescence in green/yellow and red channels were monitored during illumination. In addition, *in cellulo* measurements have been performed.

**Microscopy.** For widefield fluorescence microscopy, a Leica AF6000 LX imaging system with Photometrics CoolSNAP HQ CCD camera was used. Green and red fluorescence images were acquired using 63x 1.4NA oil immersion objective and standard filter sets: GFP (excitation BP470/40, emission BP525/50) and TX2 (excitation BP560/40, emission BP645/75). For confocal microscopy, a Leica laser scanning confocal inverted microscope DMIRE2 TCS SP2 with an 63x 1.4NA oil objective and 125 mW Ar laser was used. Live HEK293 cells expressing target proteins in cytoplasm were imaged and bleached using the following settings: 512x512 points, zoom 16 (15x15 mkm field of view), 488 nm laser intensity 5% (1.5 mkW) for detection and 100% (120 mkW) for bleaching, fluorescence detection at 500-550 nm. Photobleaching and redding were monitored in time-lapse imaging in the green and red channels at low light intensity combined with exposures to blue light of maximum intensity (GFP filter set or 100% 488 nm laser). Images were acquired and quantified using Leica LAS AF and Leica Confocal software.

**Protein expression and purification.** eYFP, Venus, Citrine, eGFP as well as eGFP-Y145L, eGFP-Y145F, eYFP-Y145L, eYFP-Y145F mutants were cloned into the pQE30 vector (Qiagen) with a 6His tag at the N-terminus, expressed in Escherichia coli XL1 Blue strain (Invitrogen), and purified using TALON metal-affinity resin (Clontech). For mammalian cells expression eGFP-N1 vector backbone (Clontech) was used. eYFP, its mutants as well as eGFP mutants were cloned into eGFP-N1 instead of eGFP. HEK293T cells (ATCC) were transfected with the above listed constructs to obtain transient protein expression.
Mammalian cell culture and transfection. Human embryonic kidney 293 (HEK293T), cell line was used. Cells were transfected with eGFP-N1 (Clontech) and derived plasmids (see protein expression and purification) using FuGene6 reagent (Promega) and growth in DMEM (Paneco) containing 10% FBS (Sigma). Live cells in the same medium were imaged 36 h after transfection using the Leica AF6000 LX fluorescence microscope and Leica SP2 confocal microscope at room temperature.

Site-directed mutagenesis. The eGFP Y145L, eGFP Y145F, eYFP Y145L, eYFP Y145F mutants were generated using overlap-extension PCR technique with the following oligonucleotide set containing the appropriate substitutions: forward 5’-ATGCGGATCCATGGTGAGCAAGGGCGAG-3’, reverse 5’-ATGCAAGCTTTTACTTTGACAGCTCGTC-3’ and forward 5’-GAGTACAAACTTCAACAGCCAC-3’, reverse 5’-GTGGCTGTGAAGTTGTACTC-3’ for eYFP and eGFP Y145F; forward 5’-ATGCGGATCCATGGTGAGCAAGGGCGAG-3’, reverse 5’-ATGCAAGCTTTTACTTTGACAGCTCGTC-3’ and forward 5’-GAGTACAAACTTCAACAGCCAC-3’, reverse 5’-GTGGCTGTGAAGTTGTACTC-3’ for eYFP and eGFP Y145L. For bacterial expression, a PCR-amplified BamHI/HindIII fragment encoding an FP variant was cloned into the pQE30 vector (Qiagen). For mammalian expression, a PCR-amplified (with 5’-CAGTACCGGGTGGCCACCATTGGTAGCAAGGGCGAGGAGCTG-3’ an-d 5’-GATCGCGGGCCGGCTACTTGTACAGCTCGTCATGCGG-3’) AgeI/NotI fragment encoding an FP variant was cloned into eGFP-N1 vector (Clontech) instead original eGFP gene.

Computational. PDB structures 1F09 and 1F0B were used to represent YFP with and without halide. For GFP, 1EMA structure was used. The details of the model
system setup and protonation states of the key residues around the chromophore are described in chapter 2. To identify possible binding sites for an outside oxidant, we performed docking calculations using AutoDock.\textsuperscript{37} These calculations were followed by the MD simulations (10 ns). We performed semi-empirical calculations of tunneling probabilities between the chromophore and various possible electron acceptors using the \textit{Pathways} model\textsuperscript{38} in which the tunneling probability between specified donor and acceptor is computed as a product of tunneling probabilities via all possible pathways. The model considers tunneling via covalent bonds, hydrogen bonds, and through space. The tunneling through the covalent bonds is assigned highest probability, followed by tunneling through hydrogen bonds, and through space. Thus, \textit{Pathways} model accounts for the distances and the connectivity (covalent and hydrogen bonds) between the donor and acceptor moieties.

To evaluate the feasibility of various mechanisms, we performed detailed calculations of the rates of ET between different sites using the Marcus rate expression:\textsuperscript{39, 40}

\[
k_{ET} = \frac{2\pi}{\hbar} |H_{DA}|^2 \frac{1}{\sqrt{4\pi \lambda k_B T}} \exp \left\{ -\frac{(\Delta G + \lambda)^2}{4\lambda k_B T} \right\},
\]

where $\Delta G$, $\lambda$, and $H_{DA}$ are the free energy change, reorganization energy, and coupling between the electronic states involved in ET. Relevant free energies and electronic couplings were computed using QM/MM. Thermodynamic averaging was performed using Warshel’s linear response approximation.\textsuperscript{41} In this approach, $\Delta G$ and $\lambda$ for the oxidation process are computed as:

\[
\Delta G_{ox} = \frac{1}{2} \left( \langle E_O - E_R \rangle_R + \langle E_O - E_R \rangle_O \right),
\]
\[
\lambda_{ox} = \frac{1}{2} \left( \langle E_O - E_R \rangle_R - \langle E_O - E_R \rangle_O \right),
\]
where \( E_O \) and \( E_R \) are electronic energies of the oxidized and reduced states of the chromophore (or tyrosine) and the brackets indicate thermodynamic averaging (subscripts \( R \) and \( O \) correspond to the averaging on the reduced and oxidized states). We used the following protocol to compute these quantities. First, we performed MD for the initial (Chro\(^-\)) and oxidized (Chro\(^+\)) states of the protein to generate equilibrium sampling (for tyrosine, the two states corresponded to Tyr and Tyr\(^-\)). We then followed with the QM/MM calculations of \( E_O - E_R \) on both states. To calculate the energetics for ET between the chromophore and selected residues, instead of \( E_O - E_R \) we computed the energy differences between the initial (Chro\(^-\)...ResX) and charge-transfer (Chro\(^+\)...ResX\(^-\)) states.

Figs. 4.2 and 4.3 show QM/MM schemes used in the calculations. In calculations of the ionization energy of the chromophore, the QM system contained the chromophore. For computing electron attachment energies of tyrosines (Tyr145 or Tyr203), the QM system contained the respective residues. In CDFT-CI calculations, the QM system contained both the chromophore and the accepting tyrosine moiety.

Following protocols validated in our previous calculations of the redox potentials, in QM/MM calculations\(^{42,43} \) we used the \( \omega \)B97X-D functional, which includes exact long-range exchange and dispersion correction.\(^ {44,45} \) The detailed protocol is described in chapter 2.

To understand the trends in the computed ET rates, we analyzed relevant structural parameters along equilibrium trajectories for various systems.

MD simulations were performed using NAMD.\(^ {46} \) Electronic structure and QM/MM calculations were performed using Q-CHEM.\(^ {47} \) CDFT-CI was used for calculations of couplings.\(^ {48} \) CHARMM27 parameters for standard protein residues\(^ {49} \) and the parameters derived by Reuter \textit{et al.} for the anionic GFP chromophore were used in the MD
Figure 4.2: QM/MM schemes for eYFP (a) and eGFP (b) used in the calculations of the ionization energies of the chromophore. The black dotted lines denote the boundary between the QM (blue) and MM parts. The MM part in which point charges were set to zero is denoted by green and red (note that green atoms are part of the chromophore).
Figure 4.3: QM/MM scheme for eYFP and eGFP used in the calculations of electron attachment energies of tyrosines. The black dotted lines denote the boundary between the QM (blue) and MM parts. The MM part in which point charges were set to zero is denoted by green and red (note that green atoms are part of tyrosine).
4.3 Results and discussion

Fig. 4.4a shows the normalized yields of bleaching (green/yellow form disappearance) and redding (red form appearance) as a function of the oxidant concentration for eGFP and eYFP (in the presence of chloride). When chloride is present, eYFP behaves similarly to eGFP. As one can see, the bleaching of the green form and the yield of the red form depend strongly on the concentration of the oxidant. Thus, under these conditions, the bleaching is mainly due to photooxidation. Therefore, the disappearance of the green/yellow form, which describes bleaching kinetics, can be loosely correlated with the ET step from Scheme (4.1) — as shown previously, one-electron oxidation leads to the formation of radical with strongly blue-shifted absorption. The rise of the red signal is related to the second step, the red chromophore formation. The upper bound for the total yield of the red form is given by the yield of the one-electron oxidized form of the chromophore, a precursor of the red form. As shown below, mutations and variations in experimental conditions affect the two signals differently, e.g., some strongly affect bleaching kinetics whereas others have no effect on bleaching but lead to changes in the red form buildup.

As mentioned in the Introduction, this study was motivated by a drastically different behavior of YFPs relative to eGFP. When no halides are present, no red signal is observed in any of the three YFPs. It is known that eYFP’s fluorescence is sensitive to Cl\(^-\), and Venus and Citrin lack this sensitivity. So, we tested the influence of Cl\(^-\) on eYFP’s redding. Indeed, we found that eYFP undergoes yellow-to-red
Figure 4.4: eGFP and eYFP oxidative photoactivation. (a) The effect of potassium ferricyanide concentration on the main (green/yellow) fluorescent state bleaching (green full squares/yellow full triangles) and the red fluorescence increase (red open squares/magenta open triangles) in the oxidative redding of immobilized eGFP and eYFP. After one activating irradiation cycle with GFP filter set, remaining green fluorescence (normalized according to initial value) and originating red fluorescence (normalized according to maximal value) were measured and shown in the graph. (b) The red fluorescence appearance in eYFP during irradiation. Immobilized eYFP was irradiated with arc-lamp (GFP filter set, 0.6 W/cm$^2$) in phosphate buffer (black squares), in the presence of oxidant (blue triangles), in the presence of sodium chloride (red circles), and in the presence of both oxidant and chloride (green triangles). Redding efficiency is normalized according to initial yellow fluorescence. Each data point is an average of three independent experiments. Error bars, s.d.
photoconversion only in the presence of chloride (Fig. 4.4b). As Cl$^{-}$ quenches eYFP’s fluorescence due to electrostatic stimulation of chromophore protonation,\textsuperscript{33,51} we also tested photoconversion of eYFP at different pH. In the absence of Cl$^{-}$, eYFP reddening was not detected even at low pH leading to complete chromophore protonation (see Fig. S3 in supporting information of Ref. 52). Thus, we concluded that the effect of Cl$^{-}$ is not related to chromophore’s protonation state. Next we tested the influence of different halide ions on eYFP’s reddening and found that efficiency of photoconversion decreases in the series: F$^{-}$ >Cl$^{-}$ >Br$^{-}$ >I$^{-}$; thus, it can be correlated with the size of the halide; this is shown in Figs. S1 and S2 in supporting information of Ref. 52.

To understand the different behavior of YFPs relative to eGFP, we turn to the analysis of possible mechanisms of photoinduced ET, drawing from the extensive studies of ET in proteins.\textsuperscript{53–57}

ET in proteins can proceed through large distances, up to 20 Å; the rates between $10^2$-$10^8$ s$^{-1}$ have been observed.\textsuperscript{53} The rates decay exponentially with the donor-acceptor distance. In many redox-active proteins the ET proceeds between well defined redox sites; in such systems the discussion of the mechanism focuses on identifying dominant pathways for ET (or the absence of thereof) and discrimination between the direct ET (one-step transport via coherent tunneling or flickering resonance) and hopping (multi-step ET via intermediate electron acceptors) mechanisms. One-step ET can proceed through space (if the donor and acceptor residues are sufficiently close) or can be mediated by covalent or hydrogen bonds (bridge-mediated superexchange). Multi-step hopping entails transient localization of charge carriers, i.e., formation of reduced or oxidized intermediates along the ET pathway.
In the case of FPs, the location of the electron-accepting oxidant molecule is not known. The distance between the chromophore and the closest solvent-accessible surface sites is about 8-10 Å. Thus, the oxidant cannot form a close contact with the chromophore. In order to identify an access point which is most favorable for ET, we investigated possible docking sites with an aim to identify those corresponding to the shortest chromophore-oxidant distances. Note that efficient redding was observed\textsuperscript{21} using a variety of oxidant molecules including rather bulky ones, such as cytochrome-c, that cannot penetrate the tight GFP barrel. In our simulations, we used para-benzoquinone, BQ, as a model oxidant. Docking calculations revealed several docking sites on the surface of the barrel. Among those, we identified a cluster of structures corresponding to the shortest chromophore-BQ distance; these structures for eGFP and eYFP are shown in Fig. 4.5. To verify the results of the docking simulations, we performed MD simulations for the docked structure with the shortest chromophore-BQ distance. We observed that the distance between Tyr145 and BQ stays mostly within 3.9-5.4 Å throughout a 10 ns long MD trajectory. The detailed discussion of the docking and MD simulations is shown in section 2.7.

The distance between Chro and BQ in these structures is about 6 Å, which is sufficiently short to consider direct tunneling. We also considered a possibility of ET by a hopping mechanism via residues with aromatic groups such as tryptophan, tyrosine, phenylalanine, or histidine. A similar mechanism involving aromatic residues serving as “stepping stones” for charge transfer in respiratory complex I has been introduced to explain the experimentally observed fast rates for ET.\textsuperscript{58} On the basis of their electron affinities, we identified tryptophan and tyrosine as the most likely acceptors. We analyzed the structures of eGFP and eYFP identifying those residues in the vicinity of the chromophore. In addition to structural analysis, we also performed semi-empirical
Figure 4.5: Mechanism of photoinduced ET in FPs. An oxidant molecule (represented by para-benzoquinone, BQ) docked to eGFP (left) and eYFP (right) and the relevant distances. The direct tunneling and two-step hopping (via Tyr145) mechanisms for ET are shown by dashed arrows.

calculations using the Pathways model that allows one to compare tunneling probabilities ($T_{DA}$) between different sites and to identify the residues that mediate ET. These calculations identified Tyr145 as the most probable electron acceptor both in eYFP and eGFP ($T_{DA} = 1.7 \times 10^{-2}$ and $1.9 \times 10^{-2}$, respectively). In eYFP, the tunneling probability to Tyr203 was of the same magnitude as for Tyr145 ($2.3 \times 10^{-2}$). For other tyrosines, the computed tunneling probabilities were at least an order of magnitude lower. $T_{DA}$ for the direct ET (from Chro to BQ) was $4.6 \times 10^{-3}$ (in eGFP); this pathway is mediated by Tyr145.

Thus, based on docking and tunneling calculations we put forward two mechanistic hypotheses (see Fig. 4.5): (i) direct tunneling to the outside oxidant (docked in the vicinity of 145 and mediated by it) and (ii) 2-step hopping mechanism in which the electron is first transferred to Tyr145 forming a transient radical-anion and then is picked up by the oxidant. Other competing channels may be operational, e.g., in eYFP, ET to
Tyr203 might occur; Phe165 or Tyr92 may also be involved. Importantly, Tyr145 is much closer to the surface than Tyr203. Thus, based on the docking and the *Pathways* calculations, Tyr145 might be an efficient intermediate electron acceptor mediating the ET to an outside oxidant, whereas Tyr203 (or other residues buried deeply inside the barrel) are inaccessible to the oxidants and are likely to be trap sites leading to either permanent bleaching (via chemical reactions of the resulting radical) or quenching (by back ET to the chromophore).

To evaluate the feasibility of these mechanisms, we performed detailed calculations of the rates of ET between different sites using the Marcus rate expression,\(^4\) Eq. (4.1), and QM/MM calculations of relevant free energies and electronic couplings using high-level electronic structure methods and Warshel's linear response approximation\(^4\) for thermodynamic averaging (see section 2.2 for details). We then analyzed the differences between eGFP, eYFP with and without chloride, as well as mutants. To make quantitative comparisons between different systems and to compare with the experiment, we focus on evaluating QY of the precursor of the red form, the product of one-electron oxidation of the chromophore.

For the direct tunneling mechanism, QY of bleaching is determined by the competition between the two channels — radiative or/and radiationless decay of the excited state (characterized by the combined rate, \(r_f\)) restoring the ground-state chromophore and ET (\(r_{et}\)). The QY of the red-form precursor is then:

\[
Y_r = \frac{r_{et}}{r_{et} + r_f} \approx \frac{r_{et}}{r_f}
\]  

(4.4)

This expression allows us to estimate an anticipated order of magnitude for ET rates. Using typical fluorescence lifetime (nanoseconds, \(r_f \sim 10^9 \text{ s}^{-1}\)) and a typical QY of bleaching\(^4,7,8\) \(Y_{bl} \sim 10^{-5}\), the estimated ET rate is then \(10^4 \text{ s}^{-1}\) (slower rates will result...
in lower bleaching yields). Because the yield of the precursor can be higher than of the bleached form, this estimate provides a lower bound to the ET rate.

![Kinetic model of photoinduced ET via hopping mechanism]

**Figure 4.6:** Kinetic model of photoinduced ET via hopping mechanism. The excited state can decay to the ground state, either radiatively or non-radiatively. This channel is characterized by $r_f$ which is inversely proportional to the excited-state lifetime ($r_f \sim 10^9 \text{ s}^{-1}$). Alternatively, the excited state can decay via ET from the chromophore to either Tyr145 or another acceptor, ResX (this could be Tyr203 in eYFP). ET to Tyr145 or ResX results in anion-radical (e.g., Tyr$^{-}$) formation that can lead to permanent bleaching ($r_b$). ET to Tyr145 can also lead to ET to an outside oxidant ($r_2$) forming a precursor of the red form. The observed bleaching is the sum of the yields of the red form precursor and of permanently bleached states. Based on our rates calculations, $r_3$ and $r_6$ are slow; $r_5$ is slow for Tyr203.

The kinetic model for the hopping mechanism is shown in Fig. 4.6. It comprises 5 states: ground-state and electronically excited chromophore, oxidized chromophore (red-form precursor), and two intermediate states in which the chromophore is oxidized and the electron resides on one of the protein residues (Tyr145 or a trap site, TyrX). $r_2$, the rate of ET between Tyr145 and an outside oxidant, is expected to be very fast, as this is an exothermic step. The upper bound is given by the diffusion-limited rate, $r_2 = 2 \times 10^{10} \text{ s}^{-1}$. We consider the following mechanism for photoinduced ET via hopping. We assume that in eGFP, there is a direct ET pathway from Chro$^{-}$ to Tyr145,
the rate is given by \( r_1 \). Once the electron reaches Tyr145, it can either go back (\( r_{-1} \) and \( r_3 \)) restoring the anionic chromophore, or initiate some chemistry (potentially leading to bleaching), or irreversibly tunnel out (\( r_2 \), fast), to an outside oxidant forming a red-form precursor. There is a competing channel, \( r_4 \), to ResX; this channel can lead to either permanent bleaching (\( r_b \)) or to restoring the chromophore (\( r_{-4} \) and \( r_5 \)). In eYFP, ResX\( \equiv \)Tyr203, in eGFP, ResX might be Tyr92 or another acceptor. As illustrated by the Pathways model and docking calculations, Tyr203 is buried inside the barrel and the pathway for ET from Tyr203 to Tyr145 involves the chromophore thus increasing the probability of quenching. Therefore, \( r_6 \) is expected to be slow making Tyr203 a dead-end for photooxidation.

The detailed analysis of this kinetic model is described in section 4.5; the main result is:

\[
Y_r \approx \frac{r_1}{r_f(1 + \frac{r_b}{r_2})} \approx \frac{r_1}{r_f} \quad (4.5)
\]

\[
Y_{totb} \approx \left( \frac{r_1}{r_f} + \frac{r_4}{r_f} \right)(1 + \frac{r_b}{r_2}) \approx \frac{r_1}{r_f} + \frac{r_4}{r_f} \quad (4.6)
\]

We note that \( \frac{r_b}{r_2} \) term is likely to be small (since \( r_2 \) is expected to be much faster than the rate of chemical reactions leading to permanent bleaching) and is, therefore, neglected in the present analysis. Thus, the trend in the yield of red form is dominated by the \( \frac{r_1}{r_f} \) ratio; as in the direct tunneling mechanism, the lower bound for \( r_1 \) is \( 10^4 \) s\(^{-1} \). The total yield of the bleached form, \( Y_{totb} \) is roughly equal the sum of yields of the red-form precursor and a permanently bleached form produced via a competing channel (ET to ResX). In order for this channel to have a noticeable effect on the yield, rate \( r_4 \) should be comparable to (or larger than) \( r_1 \).
In both mechanisms (direct or hopping) Tyr145 may play a role, either as a mediating residue or as a transient electron acceptor; thus, in the calculations below we consider the effect of mutation of this residue on the computed rates and yields.

Table 4.1: Redox properties of the ground-state and electronically excited chromophores of eGFP, eGFP-Y145L, eYFP and halide-bound eYFP at T=298 K.

<table>
<thead>
<tr>
<th>System</th>
<th>$\Delta G_{ox}^{gs}$, eV</th>
<th>$\lambda_{ox}$, eV</th>
<th>$\Delta G_{ox}^{ex}$, eV</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFP</td>
<td>4.551</td>
<td>1.599</td>
<td>2.111</td>
</tr>
<tr>
<td>eYFP</td>
<td>4.697</td>
<td>1.400</td>
<td>2.347</td>
</tr>
<tr>
<td>eYFP +Cl$^-$</td>
<td>4.274</td>
<td>1.686</td>
<td>1.924</td>
</tr>
<tr>
<td>eGFP-Y145L</td>
<td>4.548</td>
<td>1.528</td>
<td>2.108</td>
</tr>
</tbody>
</table>

Table 4.1 shows the key quantities related to the redox properties of FPs in the ground and electronically excited states. For the oxidation process to be thermodynamically favorable, $\Delta G_{ox}(Chro) + \Delta G_{red}(OX)$ should be negative. The original GFP redding study$^{21}$ reported that eGFP can be oxidized by various oxidizing agents with $E^0$ up to -0.114 V relative to the standard hydrogen electrode (SHE), which translates into $\Delta G_{red} = -4.167$ eV at pH=0, using $\Delta G(SHE)=4.281$ eV.$^{59}$ Thus, the computed energetics is consistent with estimated $\Delta G_{red}$: oxidation of the ground-state chromophore is not thermodynamically favorable, however, it becomes possible upon electronic excitation. We also observe that eYFP is more difficult to oxidize relative to eGFP, whereas chloride binding reduces $\Delta G_{ox}$. This is due to $\pi$-stacking with Tyr203, which increases ionization energy. The effect of chloride binding is twofold: it upsets $\pi$-stacking and also decreases the ionization energy due to electrostatic interactions. The Y145L mutation does not affect $\Delta G_{ox}$ of the chromophore. Using the data from Table 4.1 and estimated $\Delta G_{red}$ and $\lambda_{red}$ for BQ, we can estimate the rates (and $Y_r$) for the direct tunneling mechanism; these data are presented in Table 4.7. The main result of these calculations is that despite the variations in $\Delta G_{ox}$, the direct ET mechanism predicts similar rates in
eGFP, eYFP, chloride-bound eYFP. Thus, it does not explain the experimental findings. This direct tunneling model also predicts that the Y145L mutation will have no effect on the ET rate (because the the free energy of oxidation of the chromophore is not affected, as can be seen from Table 4.1).

The Gibbs free energies, electronic couplings and the rates for ET via hopping mechanism are collected in Table 4.2 (more details are given in chapter 2).

Table 4.2: Relevant Gibbs free energy differences, reorganization energies, and couplings and Marcus rates for ET at 298 K. Energy and coupling values are given in eV and eV^2, respectively, and the rate constants are in s^-1.

| System     | Final state | ∆G_{CT}^{sex} | λ_{CT} | |H_{da}|^2 | r_1 or r_4 |
|------------|-------------|----------------|--------|-----------|--------------|-----------------|
| eGFP       | CT 145      | 0.452          | 0.846  | 0.214     | 1.5×10^7     |
| eYFP       | 0.783       | 0.704          | 0.141  | 1.5×10^2  |
| eYFP + Cl^-| 0.561       | 0.787          | 0.590  | 2.0×10^6  |
| eYFP       | CT 203      | 0.564          | 0.287  | 0.180     | 1.2×10^5     |
| eYFP-Y145L | 0.636       | 0.394          | 0.097  | 1.1×10^4  |
| eYFP + Cl^-| 0.857       | 0.235          | 0.067  | 8.4×10^-7 |

The computed rates show that ET to Tyr145 is strongly affected by π-stacking and by halide binding. π-stacking completely shuts down the main channel and opens up another ET channel, to Tyr203. The halide binding opens up the main channel and shuts down ET to Tyr203. The computed Y_r are: Y_(r)(eGFP)=1.5 %, Y_(r)(eYFP)=2×10^-5 %, and Y_(r)(eYFP+Cl^-)=0.2 %. We also performed calculations using strong coupling regime (see supporting information of Ref. 52); the computed rates are slower (giving rise to lower QY), but the main trend remains the same. Thus, the hopping model reproduces the observed differences between the three proteins. We note that the computed rate for ET to Tyr203 in eYFP is sufficiently large to have a noticeable effect on the total yield of bleaching and that chloride binding completely shuts down this competing channel. Other residues, such as Tyr92 or Phe165 may, in principle, contribute to this
channel (their possible roles will be investigate in future study). The calculations suggest that the observed bleaching in eYFP without halide is due to permanent bleaching via ET to Tyr203, whereas in the presence of halide most of the bleaching results from forming the red-form precursor.

To understand the differences in the computed ET rates in eGFP, eYFP, and chloride-bound eYFP, we analyzed relevant structural parameters along the equilibrium MD trajectories. We focus on the distance between the chromophore and Tyr145 and between the chromophore and Tyr203 (in eYFP). Table 4.3 summarizes the results. $d_1$ is defined as the distance between chromophore’s and Tyr145 phenolic oxygens; the variations in this distance are expected to modulate the energetics and couplings defining $r_1$. To quantify the $\pi$-stacking between the chromophore and Tyr203, we computed the distances between the edges of the respective aromatic rings; these are denoted by $d_2$ and $d_3$. The definitions of these structural parameters are shown in Fig. 4.7. The distance between the two phenolic rings is given by $D = \frac{d_2 + d_3}{2}$ and the deviations from a perfectly parallel arrangement is given by $\Delta = |d_2 - d_3|$. As one can see, $d_1$ is about 1.2 Å longer in eYFP than in eGFP, but it shrinks upon chloride binding. Further analysis of the trajectories reveals that in eYFP there are two interconverting hydrogen-bond patterns: one in which there is a hydrogen bond between Tyr145 and the chromophore (in this structure, the relative orientation of Tyr145 and the chromophore is similar to eGFP and $d_1$ is small) and one in which Tyr145 moves away and forms a hydrogen bond with His169 (see Fig. 4.11). In the course of 12 nanosecond equilibrium dynamics, the first structure is populated $\sim$63% of the time (see Fig. 4.12). The presence of the two structures is responsible for the larger average value and large standard deviation of $d_1$ (see Table 4.3). Chloride binding suppresses the second structure, which leads to shorter $d_1$; it also affects $\pi$-stacking. As one can see from Table 4.3, in eYFP the phenolic rings
of the chromophore and Tyr203 are closer (shorter $D$) and more parallel (smaller $\Delta$) than in chloride-bound eYFP. As illustrated in Fig. S18 in supporting information of Ref. 52, $\pi$-stacking controls the delocalization of electronic density between the chromophore and Tyr203, which, in turn, controls electronic couplings and affects orbital energies. Thus, on the basis of the structural analysis, we conclude that the rate of ET between the chromophore and Tyr145 ($r_1$) is suppressed in eYFP because of the: (i) increase of the chromophore’s electron detachment energy due to $\pi$-stacking with Tyr203; and (ii) the presence of the conformation in which hydrogen-bond between the chromophore and Tyr145 is broken. Chloride binding suppresses the structural fluctuations, which leads to shorter Chro-Tyr145 distances and increased $r_1$. Chloride binding also distorts $\pi$-stacking of the chromophore with Tyr203, which shuts down this competing ET channel ($r_4$).

![Figure 4.7: Relevant structural parameters. The distance between the phenolic oxygens of the chromophore and Tyr145 ($d_1$) affects the main ET channel ($r_1$). The extent of $\pi$-stacking can be quantified by $D \equiv \frac{d_2+d_3}{2}$ and $\Delta \equiv |d_2 - d_3|$.

Within the hopping model, the magnitude of $r_1$ determines the yield, see Eq. (4.11). Thus, the model predicts that the photooxidation efficiency can be controlled by the mutations of residue 145. The effect of mutation of Tyr145 to phenylalanine and leucine is expected to increase $\Delta G_{CT}$ by about 0.07 eV or more. This would result in $r_1$ decrease...
by a factor of 8-10 in eGFP and eYFP+Cl\(^{-}\) reducing the yields proportionally. For the Y145F mutant, we estimated changes in free energies and couplings from snapshots of MD simulations (see section 4.5) and found that the rate for ET drops by at least a factor of 2, the main effect being the decrease in the coupling because of the lack of H-bond with the chromophore. For the eYFP-Y145L mutant, we also computed the rate for ET to Tyr203 \( (r_4) \). The mutation results in the one order of magnitude drop of \( r_4 \) (see Table 4.2), which suggests the increased photostability of the eYFP-Y145L mutant at non-oxidative conditions (i.e., without halide binding and in the absence of oxidants).

As one can see from Table 4.2, the ET to Tyr145 is endothermic and is expected to slow down at low temperature. This trend may be partially offset by the increased electronic couplings and small increase in fluorescence lifetime. The calculations predict a moderate drop in \( r_1 \) and, consequently, in \( Y_r \), e.g., in eGFP \( Y_r(273)/Y_r(298) = 0.23 \).

To test the theoretical prediction of the role of Tyr145 as an intermediate electron acceptor in the two-step hopping mechanism, we conducted mutagenesis studies. The mutants of eGFP and eYFP were constructed by mutating residue 145 to phenylalanine and leucine. As illustrated in Fig. 4.8, mutating Tyr145 to a less favorable electron acceptor led to a significantly reduced bleaching. The effect was stronger for leucine — both the eGFP-Y145L and eYFP-Y145L mutants were very photostable (Fig. 4.8a,c). Although

<table>
<thead>
<tr>
<th>System</th>
<th>( d_1, \text{Å} )</th>
<th>( D, \text{Å} )</th>
<th>( \Delta, \text{Å} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFP</td>
<td>3.77 (0.46)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>eYFP</td>
<td>5.03 (1.28)</td>
<td>3.97</td>
<td>0.25</td>
</tr>
<tr>
<td>eYFP+Cl(^{-})</td>
<td>2.89 (0.25)</td>
<td>4.18</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Table 4.3: Average values of relevant structural parameters for eGFP, eYFP, and eYFP+Cl\(^{-}\). The standard deviations are shown in parenthesis. See Fig. 4.7.
mutants with Leu145 have decreased extinction coefficients (Table S1 in supporting information of Ref. 52), 80- and 25-fold increased photostabilities of eGFP-Y145L and eYFP-Y145L can not be attributed solely to 3-5-fold decrease of their extinction coefficients (compared to eGFP and eYFP, respectively). As reported in a recent study, the fluorescence QY and lifetimes in Y145L and Y145F mutants are very similar to those in eYFP. Thus, strikingly different photostability can be attributed to the ET channel, and not to the changes in radiative and radiationless population decay of the excited state. In the eGFP mutants, the rate of red form appearance was also suppressed (Fig. 4.8b). At the same time, mutants of eYFP showed no significant changes of the redding rate (Fig. 4.8d) suggesting that Y145L/F mutations affect both steps in Scheme (4.1).

To test whether the increased photostability of the mutants persists at the conditions relevant to imaging studies, we conducted in cellulo measurements. As illustrated in Fig. 4.8(e,f), the eGFP-Y145L and eYFP-Y145L mutants expressed in mammalian cells demonstrated several-fold increased photostabilities relative to respective parental proteins in both laser scanning confocal and widefield fluorescence microscopy.

As discussed above, the hopping model via Tyr145 predicts reduced rates of photooxidation at low temperature. To verify this prediction, we compared oxidative photoconversion of eGFP in vitro in the presence of 0.5 mM potassium ferricyanide at 273, 295 and 310K (Fig. S4 in supporting information of Ref. 52). In agreement with theory, eGFP’s bleaching and redding rates are slightly enhanced at elevated temperature.

Obviously, photostability represents one of the most important characteristic of a fluorescent protein. Unfortunately, mechanisms of FP photobleaching are poorly understood. Some amino acid substitutions (mainly found by chance) were shown to strongly enhance photostability of FPs, especially low-photostable ones. For example, photostability of EBFP was dramatically (two orders of magnitude) increased by V150I
Figure 4.8: Bleaching and redding kinetics in the eGFP and eYFP mutants. (a-d) Photoconversion of immobilized proteins in vitro in PBS in the presence of 0.5 mM potassium ferricyanide. (PBS contains potassium chloride.) Graphs show the main form bleaching (a,c) and simultaneous appearance of red fluorescence (b, d) in eGFP, eGFP-Y145L, eGFP-Y145F (a, b), and eYFP, eYFP-Y145L, eYFP-Y145F (c, d). Green/yellow and red fluorescence intensities were background subtracted and normalized to the maximum values. Standard deviation values (n = 15-20 measurements in a representative experiment out of five independent experiments) are shown. (e) Bleaching of eGFP, eGFP-Y145L, eYFP, and eYFP-Y145L in live HEK293 cells induced by 488 nm laser in a confocal microscope. (f) Increase of photostability (time to half-bleaching) of the eGFP-Y145L and eYFP-Y145L mutants compared to eGFP and eYFP, respectively, under confocal and widefield microscopy of live HEK293 cells. Standard deviation values for 50-60 cells in three independent experiments are shown.
plus V224R mutations.\textsuperscript{61,62} Single substitution S158T (corresponding to position 165 in GFP) strongly improved photostability of TagRFP.\textsuperscript{63} In chloride-sensitive variant of yellow fluorescent protein ClsM, mutation S205V substantially suppressed photobleaching.\textsuperscript{64} A common feature of these mutants is the insertion of bulkier residues. It results in a decrease or full elimination of fast initial phase of bleaching, which is thought to represent cis-trans chromophore isomerization and/or protonation-deprotonation events.\textsuperscript{65–67} Also, a possible reason for photostability enhancement is lowering the accessibility of the chromophore to molecular oxygen by bulky residues. There are a few crystallographic studies directly demonstrated chromophore destruction\textsuperscript{68–70} or oxidation of chromophore-adjacent Met and Cys residues\textsuperscript{29} in photobleached FPs. The latter possibly explains a key role of the mutation M163Q (position 167 in GFP) in a high photostability of mCherry.\textsuperscript{63}

In contrast to the previous investigations of photostability, our study provides concrete mechanistic suggestion about bleaching via photoinduced ET and therefore furnishes a design principle for rational engineering of more photostable FPs. We identified Tyr145 as a key residue controlling ET and demonstrated that its substitution by less effective electron acceptors leads to the increased photostability. The residues that we selected are less bulky than the original one (Phe/Leu versus Tyr). The full mechanistic picture of bleaching and photooxidation in FPs is likely to be more complex than the 3-state model from Fig. 4.6. For example, we anticipate that other residues may also play a role and that additional ET pathways may be operational (or may become operational upon further mutations). Furthermore, in order to fully understand oxidative redding in FPs, details of red chromophore formation need to be elucidated. Particularly interesting question concerns catalytic role of various residues in the second step of Scheme 4.1. Thus, although open questions remain, our results represent the first step towards
developing molecular-level picture of photoinduced ET in FPs and provide motivation for future investigations of this fascinating phenomenon.

4.4 Conclusion

By combination of theory and experiment, we identified a dominant pathway for photoinduced ET in FPs by a hopping mechanism via Tyr145. Photooxidation can be efficiently suppressed by disrupting hydrogen bonding between the chromophore and Tyr145 and by \( \pi \)-stacking with Tyr203 (Tyr203 can also serve as an electron acceptor leading to permanent bleaching). The quenching can be controlled by the halide binding. The quenching is explained by (i) changes in energetics of ET between the chromophore and Tyr145 and by (ii) the competitive ET channel to Tyr203, which serves as a trap site. The halide binding affects structures, energetics, and electronic couplings. Our mechanism does not exclude possible involvement of other channels — additional pathways for ET and the role of other residues on ET rates will be investigated in the future studies. To further advance our understanding of oxidative redding photoconversion, structural information about the red form is needed. Better understanding of photooxidation mechanism is important for engineering FPs with desired properties optimal for a particular application. Our findings suggest design principles for controlling photoconversions and bleaching via \( \pi \)-stacking and targeted mutations around Tyr145 residue aiming to speedup or slowdown ET. We conclude by saying that FPs provide an exciting model for studying mechanism of ET in complex systems such as proteins.
4.5 Appendix A: Kinetic model for ET via hopping mechanism

In this section, we discuss the hopping mechanism for ET. Herein, we present relevant rates and introduce a kinetic model. In section 4.7, we discuss an alternative mechanism via direct ET. Fig. 4.6 shows our kinetic model of ET via the hopping mechanism. Table 4.2 summarizes the computed energetics and relevant rates at 298 K. We note that typical \( r_f \sim 10^9 \text{s}^{-1} \). The rate of the second step, \( r_2 \), of the hopping mechanism is expected to be very fast, as this is an exothermic step. The upper limit is set by the diffusion-limited rate, which we estimated as \( r_2 = 2 \times 10^{10} \text{s}^{-1} \). If this step is diffusion limited and/or dominated by tunneling, \( r_2 \) should be temperature-independent.

We consider the following mechanism for photoinduced ET via hopping; we assume that there is a direct ET pathway from Chro\(^{-*}\) to Tyr145 in the eGFP. The rate of this ET process is given by \( r_1 \). Once the electron reaches Tyr145, it can either go back (\( r_{-1} \) and \( r_3 \)) restoring Chro, or initiate some chemistry (potentially leading to bleaching) or irreversibly tunnel out (\( r_2 \), fast), to an outside oxidant forming a red-form precursor. There is a competing channel, \( r_4 \), to ResX; this channel can lead to either permanent bleaching (\( r_6 \)) or to restoration of the chromophore (\( r_{-4} \) and \( r_5 \)). In eYFP, ResX\(\equiv\)Tyr203, in eGFP, ResX might be Tyr92 or another acceptor, but it is not as competitive as Tyr203. As illustrated by the Pathways and docking calculations, Tyr203 is buried inside the barrel and the pathway for ET from Tyr203 to Tyr145 involves the chromophore, thus increasing the probability of quenching. Therefore, \( r_6 \) is expected to be slow. As one can see from Table 4.2, the \( \pi \)-stacking with Tyr203 affects the energetics of ET from the chromophore to Tyr145, suppressing the main channel for ET (\( r_1 \)). The anions affect
this scheme by modulating the couplings and energetics ($\Delta G$). The analysis reveals that the anions upset $\pi$-stacking by changing the orientation of Tyr203.

Note that in our calculations, we neglect possible proton transfer that may occur following ET. Proton transfer is expected to stabilize the accepting sites (Tyr145 or ResX), thus lowering the reverse rates.

By computing the first-passage time, the model gives the following expression for the yield of red form precursor:

$$Y_r = \frac{r_1 r_f (r_2 + r_1 + r_b)}{r_f (r_2 + r_1 + r_b) (r_4 + r_b) + r_1 (r_2 + r_b) (r_4 + r_b) + r_4 r_b (r_2 + r_1 + r_b)}.$$  \hspace{1cm} (4.7)

The total yield of bleaching is

$$Y_{totb} = \frac{1}{1 + \frac{r_f (r_2 + r_1 + r_2)}{r_1 (r_2 + r_1 + r_b) + r_4 r_b (r_2 + r_1 + r_b) / (r_4 + r_b)}}.$$  \hspace{1cm} (4.8)

These bulky expressions can be simplified under the following assumptions: $\frac{r_1}{r_2} \ll 1$ and $\frac{r_4}{r_b} \ll 1$, leading to

$$Y_r \approx \frac{r_1}{(r_f + r_1 + r_4) (1 + \frac{r_b}{r_2})},$$  \hspace{1cm} (4.9)

$$Y_{totb} \approx \frac{r_f + r_4}{(1 + \frac{r_2}{r_4}) + r_1 + r_4}.$$  \hspace{1cm} (4.10)

We can further simplify these expressions by using the fact that $r_f$ is much larger than other rates. We also note that the $\frac{r_b}{r_2}$ term is likely to be small (since $r_2$ is expected to be
much faster than the rate of chemical reactions leading to permanent bleaching, \( r_b \) and can be neglected in the present analysis. Under these conditions:

\[
Y_r \approx \frac{r_1}{r_f (1 + \frac{r_b}{r_2})} \approx \frac{r_1}{r_f}, \quad (4.11)
\]

\[
Y_{totb} \approx \left( \frac{r_1}{r_f} + \frac{r_A}{r_f} \right) (1 + \frac{r_b}{r_2}) \approx \frac{r_1}{r_f} + \frac{r_A}{r_f}. \quad (4.12)
\]

As expected, the yield of the red-form precursor is predominantly determined by \( \frac{r_1}{r_f} \). The yield of total bleaching is approximately equal to the sum of the yield of permanent bleaching via ResX and forming the red-form precursor.

**Implications of the hopping model**

Using Eqns. 4.11–4.12 and the rates from Table 4.2, we obtain \( Y_r(eGFP) = 1.5\% \), \( Y_r(eYFP) = 2 \times 10^{-5}\% \), and \( Y_r(eYFP+Cl^-) = 0.2\% \). Thus, the hopping model describes the observed differences between the three proteins correctly. The contributions to the \( Y_{totb} \) from the Tyr203 channel are \( 1.2 \times 10^{-2}\% \) (eYFP), \( 1.1 \times 10^{-3}\% \) (eYFP-Y145L), and \( \approx 0 \) in the presence of the halide. Therefore, in the absence of NaCl and the oxidant, the observed eYFP bleaching may be attributed to ET to the trap site, Tyr203. In the strong coupling limit, the computed rates are slower leading to reduced QY, i.e., 0.01\% in eGFP, which is still feasible for redding. Thus, using the Marcus theory in the strong coupling regime leads to the same conclusions.

The mutation of Tyr145 to phenylalanine and leucine is expected to increase \( \Delta G_{CT} \) by at least 0.07 eV (phenylalanine). This would result in a decrease of \( r_1 \) by a factor of 8-10 in eGFP and eYFP+Cl\(^-\), which will reduce the yields proportionally.

For a more quantitative evaluation of the effect of the mutation, we performed the following calculation: using a ground-state trajectory for eGFP, we replaced Tyr145 by Phe and computed the energy of the CT state using CDFT-CI energies and the couplings...
using CDFT-CI/ωB97X-D/cc-pVDZ. We found that $\Delta E_{CT}$ does not change (difference of about 0.04 eV) but the coupling drops by a factor of 2.2 (because of the absence of an H-bond), thus resulting in an $r_1$ that is twice as slow. This calculation yields an upper bound for the coupling — if one performs a proper equilibrium simulation of the mutant, we expect to observe large structural fluctuation of the Phe that will lead to even smaller coupling.

4.6 Appendix B: Temperature dependence for hopping model of ET.

Within the hopping model, the magnitude of $r_1$ determines the yield (see Eq. 4.11). As one can see from Table 4.2, this step is endothermic and is expected to slow down at low temperature. This trend may be partially offset by increased electronic couplings and a small increase in the fluorescence lifetime. Table 4.4 shows calculation of rates and yields at different temperatures using the data from Table 4.2.

Thus, the hopping model predicts a moderate decline in yields at lower temperature. Electronic couplings slightly increase at lower temperatures, which partially offsets this trend. For example, for the eGFP at 273 K – and taking into account the temperature dependence of the couplings – yields are $Y_r(273)/Y_r(298) = 0.23$ (compared to 0.18 from Table 4.4). Slight increase in fluorescence lifetime at low temperature may be an additional factor but we do not expect it to be large.

To summarize, the hopping model predicts a modest decrease of $Y_r$ at low T, which is in agreement with the experimental observation of a slight decrease of the bleaching
Table 4.4: Temperature dependence of the computed rates and yields for eGFP and eYFP+Cl\(^{-}\) assuming T-independent couplings.

<table>
<thead>
<tr>
<th>T</th>
<th>(r_1)</th>
<th>(Y_r, %)</th>
<th>(\frac{r_1}{r_1(298)})</th>
<th>(\frac{Y_r}{Y_r(298)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>310</td>
<td>3.09\times10^7</td>
<td>3.1</td>
<td>2.074</td>
<td></td>
</tr>
<tr>
<td>298</td>
<td>1.49\times10^7</td>
<td>1.5</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>288</td>
<td>7.71\times10^6</td>
<td>0.8</td>
<td>0.517</td>
<td></td>
</tr>
<tr>
<td>278</td>
<td>3.81\times10^6</td>
<td>0.4</td>
<td>0.256</td>
<td></td>
</tr>
<tr>
<td>273</td>
<td>2.63\times10^6</td>
<td>0.3</td>
<td>0.177</td>
<td></td>
</tr>
<tr>
<td>eYFP+Cl(^{-})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>310</td>
<td>4.52\times10^6</td>
<td>0.5</td>
<td>2.342</td>
<td></td>
</tr>
<tr>
<td>298</td>
<td>1.93\times10^6</td>
<td>0.2</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>288</td>
<td>1.30\times10^6</td>
<td>0.1</td>
<td>0.674</td>
<td></td>
</tr>
<tr>
<td>278</td>
<td>3.97\times10^5</td>
<td>&lt;0.1</td>
<td>0.206</td>
<td></td>
</tr>
<tr>
<td>273</td>
<td>2.58\times10^5</td>
<td>&lt;0.1</td>
<td>0.134</td>
<td></td>
</tr>
</tbody>
</table>

yield at low T. The large increase of the red chromophore formation can be explained by the T-dependence of the slow chemistry step of the red chromophore formation.

4.7 Appendix C: ET via direct tunneling

Based on the docking and Pathways calculations, we also considered the possibility of the direct ET/tunneling mechanism shown in Fig. 4.5. Docking calculations reveal that the lowest-energy docked structures with the closest BQ-Chro distance correspond to the BQ docked closely to Tyr145. The Pathways calculations confirm that the direct ET in this structure is mediated by Tyr145. The computed \(T_{DA}\) is 4.3\times10^{-3} for eGFP when the chromophore and BQ are about 6 Å apart, which is 10 times smaller than the Chro-Tyr145 value. Thus, based on this calculation alone, the direct ET is feasible. To investigate the effect of mutations on ET, we constructed a mutant, eGFP-Y145L and repeated docking calculations. We found a similar docking site for this mutant. The
shortest distance between the docked BQ and the chromophore for eGFP-Y145L is \( \approx 6.5 \, \text{Å} \) (Fig 4.9).

![Figure 4.9: BQ docked in the vicinity of residue 145 in the eGFP-Y145L mutant.](image)

We then computed tunneling probabilities for the mutant. The results are summarized in Table 4.5. The probabilities are one order of magnitude smaller than those for eGFP. The difference is due to the H-bond between Tyr145 and the chromophore in eGFP (in the Pathways model, H-bonds increase the tunneling probabilities relative to the through-space pathway, as described in section 2.5). Thus, docking and Pathways calculations predict a decrease of the ET rates via direct ET/tunneling in the 145 mutants.

**Table 4.5: Direct tunneling probabilities from the chromophore to the closest docked BQ in eGFP and eGFP-Y145L.**

<table>
<thead>
<tr>
<th>System</th>
<th>( T_{DA} )</th>
<th>Mediated by</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFP</td>
<td>( 4.6 \times 10^{-3} )</td>
<td>Tyr145</td>
</tr>
<tr>
<td>eGFP-Y145L</td>
<td>( 3.1 \times 10^{-4} )</td>
<td>His148</td>
</tr>
<tr>
<td>eGFP-Y145L</td>
<td>( 1.4 \times 10^{-4} )</td>
<td>Leu145</td>
</tr>
</tbody>
</table>

However, the Pathways model only captures the effect of DA distance and the connectivity e.g., hydrogen bonding and covalent bonding networks and is not sensitive to the details of electronic structure. To take these effects into account, we computed electronic couplings using CDFT-CI for the two states:

1. \( Chro^– – Tyr145/Leu145 – BQ \)
2. Chro – Tyr145/Leu145 – BQ$^-$

In these calculations, residue 145 acts as a mediating residue. We used a similar protocol as in the CDFT-CI calculations of the Chro-Tyr145 couplings. The QM part in the CDFT-CI calculation comprises all three residues. In the ground and CT states, the constraints were applied to the chromophore and BQ, respectively. The remainder of the protein was included in the MM region as point charges. We repeated the same calculation for eGFP-Y145L (here, the mediator was Leu145).

**Table 4.6: Electronic couplings for direct ET from the chromophore to the closest docked BQ in eGFP and eGFP-Y145L.**

| System          | $|H_{DA}|^2$, eV$^2$ |
|-----------------|--------------------|
| eGFP            | $2.6 \times 10^{-5}$ |
| eGFP-Y145L      | $2.0 \times 10^{-5}$ |

Surprisingly, the coupling values are much smaller (4-5 orders of magnitude) than the couplings between the chromophore and Tyr145 (0.214 eV$^2$). Moreover, the difference between eGFP and eGFP-Y145L is rather small. Thus, contrary to the Pathways calculations, CDFT-CI calculations suggest that (i) couplings are considerably smaller (so the rates would be slower too, although more favorable energetics — exothermic $\Delta G$ — may offset and even reverse that), and (ii) couplings in eGFP and eGFP-Y145L are very similar. These results strongly argue against the direct ET mechanism.

We did not compute all the relevant energetics and ET rates for this mechanism which would require very extensive calculations. Instead, what is presented below is a simple analysis using the redox potentials that are already computed. Based on BQ EA, the respective Gibbs free energy change is negative. For aqueous BQ, $\Delta G_{red} = -4.30$ eV. When BQ is docked on the protein surface, we expect this value to be less negative.
(less efficient solvation of BQ\(^-\) by the protein surface relative to bulk water). Thus, \(\Delta G\) for ET from the chromophore to BQ in the eGFP is \(\Delta G = -2.189\ \text{eV}\). The maximum rate for ET is achieved when \(\lambda = -\Delta G\). Using the computed Chro-BQ couplings (Table 4.6), the maximal possible ET rate is \(k_{max} = 3 \times 10^{10} \text{s}^{-1}\). Thus, direct ET might be possible.

The rate depends very strongly on \(\lambda\). For example, using \(\lambda = 0.85\ \text{eV}\) (largest reorganization energy for Chro-Tyr ET), \(k = 5.8 \times 10^1 \text{s}^{-1}\) (which is too slow for the excited-state ET). In order for the rate to be equal to the rate of Chro→Tyr145 ET, \(\lambda\) should be 1.20 eV. In order to attain the estimated lower-bound of the rate, \(10^4 \text{s}^{-1}\), \(\lambda\) should be 0.96 eV.

To obtain a more realistic estimate of the rate via direct ET, we computed \(\lambda\) for BQ in aqueous solution using MD and AIMD (B-LYP/6-31+G*) trajectories and \(\omega\)B97X-D/6-31+G(d,p) for \(\Delta E\). \(\lambda\) was computed as a variance (which often overestimates \(\lambda\) relative to the so-called Stokes \(\lambda\) that we are calculating in LRA). The resulting values were 1.79 eV (for the AIMD sampling) and 2.74-2.94 eV for the MD sampling. Table 4.7 lists the rates for the direct ET computed using different \(\lambda\) values and the thermo-dynamic quantities for the chromophores from Table 4.1. The following estimates were used

\[
\Delta G = \Delta G_{ox}(Chro^{-*}) + \Delta G_{red}(BQ) \tag{4.13}
\]
\[
\lambda = \lambda_{ox}(Chro^{-*}) + \lambda_{red}(BQ). \tag{4.14}
\]

These expressions assume that the donor and acceptor are sufficiently far apart for their interaction to be neglected (this is clearly not the case in Chro→Tyr145 calculations). Note that the lower bound for \(\lambda\) is given by \(\lambda_{ox}(Chro^{-*})\), in this approach.
### Table 4.7: Rates for direct ET (Chro$^\ast- \rightarrow$BQ) at T=298 K.

<table>
<thead>
<tr>
<th>System</th>
<th>$\Delta G$, eV</th>
<th>$\lambda$, eV</th>
<th>$H_{DA}^2$, eV$^2$</th>
<th>$k$, s$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$(BQ)=2.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eGFP</td>
<td>-2.189</td>
<td>4.439</td>
<td>2.6E-05</td>
<td>3.2x10$^5$</td>
</tr>
<tr>
<td>eYFP</td>
<td>-1.953</td>
<td>4.240</td>
<td>2.6E-05</td>
<td>1.3x10$^5$</td>
</tr>
<tr>
<td>eYFP+Cl$^-$</td>
<td>-2.376</td>
<td>4.526</td>
<td>2.6E-05</td>
<td>1.0x10$^6$</td>
</tr>
<tr>
<td>eGFP-Y145L</td>
<td>-2.192</td>
<td>4.368</td>
<td>2.0E-05</td>
<td>4.3x10$^5$</td>
</tr>
</tbody>
</table>

| $\lambda$(BQ)=1.79 |                |               |                   |              |
| eGFP              | -2.189         | 3.389         | 2.6E-05           | 3.8x10$^8$   |
| eYFP              | -1.953         | 3.190         | 2.6E-05           | 2.3x10$^8$   |
| eYFP+Cl$^-$       | -2.376         | 3.476         | 2.6E-05           | 8.0x10$^8$   |
| eGFP-Y145L        | -2.192         | 3.318         | 2.0E-05           | 4.5x10$^8$   |

As one can see, the computed rates contradict the experimental observations. The rates computed from the BQ data are very similar in all four proteins. The rates computed using a smaller value of $\lambda$ (1.20 eV) show a faster rate in the eYFP than in the eGFP and a slower rate in the eYFP+Cl$^-$. The rate the in eGFP-Y145L is almost the same as in the eGFP. Thus, these calculations provide a strong argument against the direct ET mechanism.

Using $\lambda = 1.20$ eV, the anticipated T-dependence for the direct ET is shown in Fig. 4.10. As one can see, even though this process is exothermic, the Marcus model predicts slower rates at lower T, i.e., $r(273)/r(298) = 0.5$ (again, this trend can be partially offset by the increased couplings and excited-state lifetime).

One can attempt to estimate the trends in rates using a simpler expression that does not depend on $\lambda$ and only takes into account trends in $\Delta G_{ox}$ from Table 4.1. As one
can see, the redox potentials show the same trend as the ET rates for the eGFP and eYFP/eYFP+Cl\(^-\) (but not for the mutant). The π-stacking with Tyr203 increases Δ\(G_{ox}\) by 0.15 eV, which may slow down the rate by about two orders of magnitude at room T. The chloride decreases Δ\(G_{ox}\), thus making the oxidation process feasible.

One can estimate the changes in the rate using these energies and the linear free energy approach within a simple one-step model, Chro\(^-\)\* → Chro\(^-\). In this approach, the activation energy of a reaction is assumed to be proportional to the change in Gibbs free energy. The rate constant can be then calculated as

\[
k \approx \exp\left(-\frac{\alpha \Delta G_{ox}}{k_B T}\right),
\]

(4.15)
where $\alpha$ is a constant between 0 and 1. Using $\alpha = 0.5$, we calculate relative rate constants for the four proteins:

$$k_{eGFP} : k_{eYFP} \approx 99 : 1$$ (4.16)

$$k_{eYFP} : k_{eYFP+Cl^-} \approx 1 : 3750$$ (4.17)

$$k_{eGFP} : k_{eYFP+Cl^-} \approx 1 : 38$$ (4.18)

$$k_{eGFP-Y_{145}L} : k_{eGFP} \approx 1.1 : 1.$$ (4.19)

As one can see, the rate of oxidation in the eYFP is about 120 times slower than in the eGFP. Thus, the yield of redding should also drop proportionally. This is consistent with the experiment. However, since the Tyr145→Leu mutation does not affect $\Delta G_{ox}$ of the chromophore, the resulting rate in the mutant is almost the same as in the eGFP. Thus, these calculations also argue against the direct ET mechanism.

### 4.8 Appendix D: Structural analysis

Fig. 4.11 shows the representative snapshots from the eYFP’s equilibrium trajectory: the snapshot in which the relative arrangement of the chromophore and Tyr145 is similar to that in the eGFP and the second, in which Tyr145 moves away from the chromophore and forms a hydrogen bond with the His169. Fig. 4.12 shows the distance between the chromophore and Tyr145 along a 12 ns equilibrium trajectory. The relative population of the conformation in which the hydrogen bond between Tyr145 and the chromophore is broken (third peak in the histogram) is $\sim 37\%$. 
Figure 4.11: Two snapshots along the eYFP equilibrium trajectory illustrating two interconverting hydrogen-bond patterns. In the dominant conformation (left), the chromophore forms a hydrogen bond with Tyr145, similar to the eGFP. In the second conformation, Tyr145 flips and forms a hydrogen bond with the His169.
Figure 4.12: Left: Distance between the chromophore and Tyr145 along a 12 ns trajectory. Right: Chro-Tyr145 distance distribution.
Chapter 4 References


Chapter 5: Future work

The research presented in this thesis investigates the mechanism of ET in fluorescent proteins and provides insight into factors controlling the photostability of fluorescent proteins. This insight provides guidelines for designing fluorescent proteins with desired photostability. For example, the oxidative redding process can be controlled using the mechanistic insight uncovered by our research. In this chapter we discuss remaining issues and outline future directions.

Recall that the $\pi$-conjugation is identical in the chromophores of eGFP and eYFP. The chromophores differ because of one of the amino acids forming the chromophore, residue 65. Position 65 is occupied by Thr in GFP and by Gly in YFP. Residue 65 is connected to the imidazolinone part of the chromophore (see Fig. 4.2). It turns out that oxidative redding in eYFP can be enabled (without the halide) by introducing a threonine-like side chain, achieved by G65T mutation (Lukyanov, private communication). Unlike eYFP, the eYFP-G65T mutant undergoes redding in the absence of halides. Further work is needed to understand this experimental observation, including calculations of the ET rates in eYFP-G65T and a structural analysis focusing on the relative distances between the chromophore and Tyr145.

In this thesis, we presented a detailed and accurate protocol for computing redox properties of the GFP chromophore. We also reported a protocol for computing ET
rates from the GFP chromophore to several acceptors. In chapter 3, we presented computed redox potentials of model blue, green and red fluorescent protein chromophores. We note that none of these model chromophores are actually fluorescent in solution, i.e., they are only fluorescent inside the protein barrel. Although the redox potentials of these model chromophores have been measured experimentally in solution, the redox potentials of the protein-bound chromophores are still not known. The change in fluorescence when a residue in GFP is oxidized has been investigated and exploited in the design of redox-sensitive fluorescent proteins\textsuperscript{1,2} but no quantitative measurements of the redox potential of the protein-bound GFP chromophore have been reported thus far. In addition, the rates of ET in GFP has not been measured, and so we can only predict the consequence of ET qualitatively, using a kinetic model. For example, in chapter 4, we predict the trend in the yield of bleaching and redding using our kinetic model and computed ET rate, but these computed rates cannot be compared to the experiment. This motivates the need for more experimental work in this direction.

Recently, in their study of myoglobin, Chergui and co-workers\textsuperscript{3} explored ET from an electronically excited tryptophan to the heme cofactor. Figure 5.1 shows the relevant amino-acid residues in the horse myoglobin that served as the model protein for this study. By using ultra-broadband ultrafast 2D spectroscopy, they directly measured the rate of ET from Trp14\textsuperscript{*} to the heme cofactor and found it to be $\sim 2.5 \times 10^{10} \text{ s}^{-1}$. The results obtained using the experimental protocol reported by Chergui and co-workers\textsuperscript{3} can be used to test the robustness of our protocol for computing ET rates; we can compute the rate of ET from Trp14\textsuperscript{*} to the heme in this system and directly compare it with the experimental results. This intriguing system brings some new challenges. An obvious challenge is that the size of the QM region required to model ET in horse myoglobin is much larger than the QM region used to study ET in fluorescent proteins. The heme
Figure 5.1: Structure of horse myoglobin, with relevant residues shown using spherical atom representation. Reproduced from Ref. 3.

cofactor comprises an iron atom and a porphyrin ring, satisfying the square-planar co-ordination sites of Fe. The 5th and 6th coordination sites are occupied by a histidine and a small molecule (or ion), such as water, hydroxyl ion, etc. Inclusion of a transition metal atom, such as Fe, in electronic structure calculations calls for careful description of the system. Often, effective core potentials are used to describe the core electrons. In the ground and CT states of this system, Fe assumes +3 and +2 oxidation states, respectively, but its spin states are unclear. For example, in +2 oxidation state, Fe$^{+2}$ can be in a low-spin (spin multiplicity = 1, closed-shell) or a high-spin (spin multiplicity = 5, open-shell) state, but in a complex environment such as myoglobin, Fe$^{+2}$ can adopt other intermediate spin states as well.\textsuperscript{4–7} So a careful analysis using the spin-flip approach\textsuperscript{8, 9} is warranted to correctly identify the lowest energy spin state of the relevant ground and CT states.

ET occurring within the heme-chains of membrane proteins is yet another area of active research where the protocol established in this thesis could be applied. X-ray structures of several membrane proteins have been resolved in recent years, including
MtrF and OmcA\textsuperscript{10, 11} of a metal reducing bacteria, \textit{Shewanella oneidensis}. This bacterium is capable of extending its outer membrane to form nanowires composed of proteins stacked on top of each other.\textsuperscript{12, 13} These nanowires can transport electrons that are essential for the respiration process of bacteria. This process is very efficient, as one can estimate from respiration rates of the bacteria. The current passing through the entire nanowire have been also measured experimentally.\textsuperscript{13}

\textbf{Figure 5.2:} Relative orientation of decaheme unit found in MtrF. Reproduced from Ref. 14.

El-Naggar and co-workers\textsuperscript{12} applied a real-time imaging techniques to observe how these nanowires grow from the extracellular membrane of \textit{Shewanella oneidensis}. Several computation studies\textsuperscript{14, 15} attempted to explain the mechanism of ET through the decaheme channel of MtrF by computing the rate of ET between the heme units. However, the free energy and reorganization energy that they obtained for each step of ET were inaccurate, because they were computed using a classical non-polarizable force field. In addition, the calculated coupling values were much smaller than the experimental estimates. A combined theoretical and experimental study\textsuperscript{16} using kinetic Monte
Carlo (KMC) simulations and single-molecule scanning tunneling microscopy (STM), demonstrated the need for more accurate computation of electronic coupling. Although KMC simulations can estimate the current through a very long nanowire (20 nm in this study), reliable Marcus parameters are needed (i.e., free energy, reorganization energy and electronic coupling) for each ET step. Using our protocol, it should be possible to accurately compute the ET parameters, which can then be incorporated into a KMC simulation of the entire membrane. This is yet another direction of research that can exploit computational protocols developed in this thesis.
Chapter 5 References


