Quantum Chemical and QM/MM Calculations on Flavin and Flavin-Binding Fluorescent Proteins

Mohammad Pabel Kabir

Follow this and additional works at: https://scholarworks.gsu.edu/chemistry_diss

Recommended Citation
doi: https://doi.org/10.57709/35481479

This Dissertation is brought to you for free and open access by the Department of Chemistry at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Chemistry Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
Quantum Chemical and QM/MM Calculations on Flavin and Flavin-Binding Fluorescent Proteins

by

Mohammad Pabel Kabir

Under the Direction of Samer Gozem, PhD

A Dissertation submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2023
ABSTRACT

Flavin is the cofactor for the large and diverse family of proteins called flavoproteins. Upon absorption of blue light, the flavin cofactor may undergo one of several potential photophysical processes, depending on the surrounding environment. Possible photophysical processes include fluorescence from the first singlet excited state ($S_1$), intersystem crossing to a long-lived triplet state ($T_1$), or photoreduction to form a radical species ($D_1$). Each of those processes has been exploited for applications such as biosensing, bioimaging, FRET process, optogenetics, and singlet oxygen generation. However, the dependence of flavin’s photophysics (mechanism and kinetics) on its protein environment is not well understood. As a step towards developing this understanding, we simulated UV-vis and FT-IR spectra of flavin in a different polar environment to understand the effect of hydrogen bonding interaction on absorption and fluorescence energies and the vibrational frequencies of flavin. We also generated electrostatic spectral tuning maps (ESTMs) for each of flavin’s redox and protonation states to see how electrostatics influence their energetics. We also developed an automated quantum mechanical/molecular mechanical (QM/MM) protocol for simulating flavoproteins. With the guide of ESTMs and the QM/MM calculations, we focus on studying the spectral tuning of iLOV, an engineered flavin-binding fluorescent protein. Specifically, we suggest a novel mutant iLOV-Q430E that provides a red-shifted absorption and fluorescence maximum wavelength, which has been experimentally verified.

INDEX WORDS: Flavin, Flavoproteins, Photoreceptors, Quantum mechanical/Molecular mechanical (QM/MM) simulations, Molecular Dynamics (MD), Spectroscopy.
Quantum Chemical and QM/MM Calculations on Flavin and Flavin-Binding Fluorescent Proteins

by

Mohammad Pabel Kabir

Committee Chair: Samer Gozem

Committee: Giovanni Gadda
            Ivaylo Ivanov
            Gary Hastings

Electronic Version Approved:

Office of Graduate Services
College of Arts and Sciences
Georgia State University
May 2023
DEDICATION

To my beloved parents

For their unconditional love and support

&

To my family

For their encouragement and sacrifice throughout this journey
ACKNOWLEDGEMENTS

I am deeply grateful to my supervisor, Dr. Samer Gozem, for his unwavering support throughout my research journey. Not only did he provide invaluable guidance and motivation within the academic realm, but he also became a friend who made a profound impact on my life. Dr. Gozem's expertise in computational chemistry was a major source of inspiration for me and without his guidance, involvement, and persistent support, my dissertation would not have been possible.

I would like to express my sincere gratitude to our experimental collaborators, Dr. Giovanni Gadda and Dr. Gary Hastings, for their guidance and support. Their discussions and advice throughout my research work were invaluable and provided me with knowledge and experience that will stay with me forever.

I would also like to thank my committee members, Dr. Giovanni Gadda, Dr. Gary Hastings, and Dr. Ivaylo Ivanov, for their constructive suggestions and discussions during my research work. Their input was extremely helpful, and I am grateful for their presence on my committee.

I am also thankful to our former postdoctoral fellow, Dr. Yoelvis Orozco Gonzalez, who taught me a great deal about computational methods and provided guidance throughout this journey. Additionally, I would like to express my gratitude to my lab mates, Dr. Paulami Ghosh, Jorge Garcia-Alvarez, Stephen Ajagbe, Ngan Le, and Erica Browne, for their helpful discussions and valuable input. It has been a pleasure to work with each and every one of them.

I am filled with gratitude for my parents, Md. Siraj Miah and Mrs. Rokeya Begum, who have always been there for me, supporting me through thick and thin. Their belief in me gave me the freedom to pursue my dreams. I will always be grateful to my mother for her selfless love, care, and sacrifices in educating and preparing me for my future. My deepest thanks go to my
beloved wife, Bilkis Mehrin Moni, for her love, understanding, sacrifices, and constant support throughout this journey.

I am also incredibly grateful to my siblings, Md. Faysal Kabir, Md. Jewel Kabir, Mrs. Rita Kabir, Md. Rubel Kabir and Mrs. Sania Kabir, for their unconditional love, support, and encouragement. Their support has been a constant source of strength for me. I would also like to thank my in-laws for their support and valued prayers.

Finally, I would like to acknowledge the support of the National Science Foundation, the Molecular Basis of Diseases, Provost Dissertation Fellowship, and the Department of Chemistry at Georgia State University. I am grateful for the contributions of all those who have made this dissertation a reality and who have supported me in ways both large and small.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................ V

LIST OF TABLES .................................................................................................................. XIII

LIST OF FIGURES ................................................................................................................ XV

1  INTRODUCTION ......................................................................................................................... 1

1.1 Overview ................................................................................................................................. 1

1.2 Flavoprotein Photoreceptors ................................................................................................. 3

1.2.1 Cryptochrome .................................................................................................................... 3

1.2.2 LOV domain ...................................................................................................................... 4

1.2.3 BLUF domain ................................................................................................................... 4

1.3 LOV domain Photocycle ....................................................................................................... 5

1.4 Outline of Chapters and Goal of the Research ..................................................................... 6

1.5 References ............................................................................................................................. 9

2  ELECTROSTATIC SPECTRAL TUNING MAPS FOR BIOLOGICAL CHROMOPHORES ........................................................................................................ 13

2.1 Abstract ................................................................................................................................. 13

2.2 Introduction .......................................................................................................................... 14

2.3 Methodology ......................................................................................................................... 18

2.4 Result and Discussion .......................................................................................................... 22

2.4.1 Chromophores of Photoreceptor proteins ..................................................................... 22
2.4.2 Rhodopsins.............................................................................................................. 24

2.4.3 Photoactive Yellow Protein (Xanthopsins) ............................................................... 25

2.4.4 Cyanobacteriochromes (Phytochromes) ................................................................. 25

2.4.5 CRY, LOV, BLUF, and other flavoproteins............................................................... 26

2.4.6 GFP derivatives ...................................................................................................... 28

2.4.7 Aromatic amino acids ............................................................................................ 31

2.4.8 Phenylalanine........................................................................................................... 31

2.4.9 Tyrosine................................................................................................................... 32

2.4.10 Tryptophan ........................................................................................................... 32

2.4.11 Effect of charge magnitude, sign, distance, and plurality on the ESTMs............. 33

2.4.12 Effect of charge magnitude on the ESTM............................................................... 34

2.4.13 Effect of charge sign on the ESTM ....................................................................... 35

2.4.14 Effect of charge distance on the ESTM ................................................................. 35

2.4.15 Effect of multiple charges .................................................................................... 36

2.5 Conclusions.................................................................................................................. 44

2.6 References .................................................................................................................. 46

3 ELECTRONIC SPECTRA OF FLAVIN IN DIFFERENT REDOX AND
PROTONATION STATES: A COMPUTATIONAL PERSPECTIVE ON THE
EFFECT OF THE ELECTROSTATIC ENVIRONMENT ............................................. 57

3.1 Abstract....................................................................................................................... 57
3.2 Key words ................................................................................................................. 58
3.3 Abbreviations ............................................................................................................. 58
3.4 Introduction ............................................................................................................... 58
3.5 Theoretical Background .......................................................................................... 62
3.6 Computational Approach ......................................................................................... 63
3.7 Results and Discussion ............................................................................................ 67
  3.7.1 The Fl UV/vis spectrum and ESTMs................................................................. 68
  3.7.2 Comparison with experimental data for Fl......................................................... 71
  3.7.3 Comparison with other computational studies for Fl......................................... 71
  3.7.4 The FlH∙ UV/vis spectrum and ESTM. ............................................................... 73
  3.7.5 Comparison with experimental data for FlH∙ ..................................................... 75
  3.7.6 Comparison with other computational studies for FlH∙.................................... 76
  3.7.7 The Fl⁻ UV/vis spectrum and ESTMs ................................................................. 76
  3.7.8 Comparison with experimental data for Fl⁻ ......................................................... 78
  3.7.9 Comparison with other computational studies for Fl⁻ ....................................... 78
  3.7.10 The FlH₂ UV/vis spectrum and ESTMs ............................................................. 78
  3.7.11 Comparison with experimental data for FlH₂ .................................................... 80
  3.7.12 Comparison with other computational studies for FlH₂.................................. 81
  3.7.13 The FlH⁻ UV/vis spectrum and ESTMs ............................................................. 81
  3.7.14 Comparison with experimental data for FlH⁻ .................................................... 83
3.7.15 Comparison with other computational studies for FlH .................................................. 83

3.7.16 Calculation of Franck-Condon factors............................................................................. 84

3.7.17 Simulation of experimental spectra using broadening and normalization..... 86

3.8 Conclusion.......................................................................................................................... 88

3.9 References .......................................................................................................................... 90

4 THE EFFECT OF HYDROGEN-BONDING ON FLAVIN’S INFRARED
ABSORPTION SPECTRUM........................................................................................................ 100

4.1 Abstract ............................................................................................................................... 100

4.2 Key words ............................................................................................................................. 101

4.3 Introduction ......................................................................................................................... 101

4.4 Methodology ....................................................................................................................... 104

4.5 Results and Discussion......................................................................................................... 108

4.6 Conclusion ............................................................................................................................ 123

4.7 References ............................................................................................................................ 124

5 AN ALTERNATIVE STRATEGY FOR SPECTRAL TUNING OF FLAVIN-
BINDING FLUORESCENT PROTEINS .................................................................................... 130

5.1 Abstract ............................................................................................................................... 130

5.2 Keywords .............................................................................................................................. 130

Error! Bookmark not defined.

5.3 Introduction ......................................................................................................................... 131

5.4 Methods ............................................................................................................................... 134
5.4.1 Computational approach ................................................................. 134
5.4.2 Bacterial strains and plasmids .......................................................... 139
5.4.3 Protein expression and purification ..................................................... 139
5.4.4 Flavin reconstitution ........................................................................ 140
5.4.5 UV-Visible Absorption and Fluorescence Spectroscopy ...................... 141
5.5 Results and Discussion ........................................................................ 141
5.5.1 ESTMs and charge analysis ............................................................... 142
5.5.2 QM/MM simulations of iLOV and iLOV-Q430E ................................. 146
5.5.3 Experimental absorption and emission spectra of iLOV and iLOV-Q430E . 148
5.6 Conclusion ............................................................................................ 154
5.7 References ............................................................................................. 156

6 EXCITED-STATE BENCHMARK STUDIES FOR FLAVIN COFACTOR IN
DIFFERENT REDOX AND PROTONATION STATES .............................. 169

6.1 Abstract ................................................................................................. 169
6.2 Introduction ............................................................................................. 170
6.3 Computational details ............................................................................ 171
6.4 Results and Discussion ........................................................................ 172
6.4.1 Oxidized flavin (Fl) ........................................................................ 173
6.4.2 Other redox states ............................................................................ 177
6.4.3 Potential Energy Surface (PES) Scan: ............................................. 179
7 MODULATION OF LOW-LYING EXCITED STATE ENERGETICS IN THE FLAVIN COFACTOR ........................................................................................................ 196

7.1 Abstract ........................................................................................................ 196

7.2 Keywords ...................................................................................................... 196

7.3 Introduction .................................................................................................. 196

7.4 Computational details .................................................................................. 198

7.5 Methodology ................................................................................................ 199

7.6 Result and Discussion .................................................................................. 201

7.6.1 The effect of hydrogen bonding interactions ....................................... 203

7.6.2 The effect of geometry distortion .............................................................. 205

7.7 Conclusion ................................................................................................... 206

7.8 Reference .................................................................................................... 208

8 CONCLUSION .................................................................................................. 210
LIST OF TABLES

Table 3.1 Computed vertical excitation energies and oscillator strengths (intensities) as well as adiabatic excitation energies for bright excited states in Fl, FlH-, Fl-, FlH2, and FlH-. The number of the excited state in the TD-B3LYP/cc-pVTZ calculation is labeled for easy referencing in the text. ................................................................. 67

Table 3.2 Experimental $\lambda_{\text{max}}$ values from literature for oxidized flavin ........................................... 87

Table 4.1 Computed bond distance ($q$), local mode force constant ($k_a$), and bond order ($n$) for various hydrogen-bonding interactions in 9W. ................................................................. 108

Table 4.2 Computed bond orders ($n$) for all hydrogen-bonding interactions in 9W computed with different basis sets........................................................................................................... 118

Table 4.3 Computed frequencies (in cm$^{-1}$) in the gas-phase and solvent models, scaled by 0.964. ........................................................................................................................................ 121

Table 4.4 Computed frequencies (in cm$^{-1}$) using gas-phase and different solvent models, scaled by such that the C=C vibrational mode matches the experimental value (1548 cm$^{-1}$). .. 122

Table 4.5 Computed bond orders ($n$) for non-covalent hydrogen bonding interactions for each of the models in Figure 4.3 ........................................................................................................... 122

Table 5.1 Experimental absorption and fluorescence properties of iLOV proteins......................... 150

Table 5.2 CASPT2//CASSCF/ANO-L-VDZP computed ground ($S_0$) and first singlet excited ($S_1$) state energies. Each column refers to a step of the ASEC-FEG protocol. In addition to absolute energies (in Hartree), the $S_0$-$S_1$ vertical excitation energies (in eV and nm) are also shown, as well as the shift (in nm) caused by the Q430E mutation. ........................................ 152

Table 5.3 CASPT2//CASSCF/ANO-L-VDZP computed ground ($S_0$) and first singlet excited ($S_1$) state energies in the presence of 4 additional Na$^+$ and Cl$^-$ ions. Each column refers to a step
of the ASEC-FEG protocol. In addition to absolute energies (in Hartree), the $S_0-S_1$ vertical excitation energies (in eV and nm) are also shown, as well as the shift (in nm) caused by the Q430E mutation.

**Table 5.4** CASPT2//CASSCF/ANO-L-VDZP computed ground ($S_0$) and first singlet excited ($S_1$) state energies at the $S_1$ minimum geometry. Each column refers to a step of the ASEC-FEG protocol. Note that the calculations needed 3 steps to reach self-consistency, so results from Steps 1-3 were discarded. In addition to absolute energies (in Hartree), the $S_0-S_1$ vertical excitation energies (in eV and nm) are also shown, as well as the shift (in nm) caused by the Q430E mutation.

**Table 6.1** Nonparallelity errors (NPEs) in eV for the ground and excited states along with $S_0-S_1$ and $S_1-S_2$ path.

**Table 7.1** Bond lengths of lumiflavin, in Angstroms, in the presence of water at different positions (indicated at the top of the table). The atomic labels are shown in Figure 7.1.

**Table 7.2** Vertical excitation energies of FMN in the presence of external water positions, heavy atoms, and distorted geometry.
LIST OF FIGURES

Figure 1.1 Structure of flavin in different redox and protonation states........................................ 1

Figure 1.2 Structure of Lumiflavin (LF), riboflavin (RF, vitamin B2), flavin mononucleotide (FMN), and flavin adenine dinucleotide (FAD). ................................................................. 2

Figure 1.3 Electronic states involved in LOV’s photocycle and their applications......................... 6

Figure 2.1 Schematic figure presenting the electrostatic spectral tuning mechanism of all-trans rPSB. Introduction of a positive electrostatic potential near the Schiff base nitrogen results in a red-shift of the absorption energy, while a positive potential near the β-ionone segment would result in a blue-shift. The opposite is true for a negative electrostatic potential. Adapted from refs 18-19 with permission. ................................................................. 16

Figure 2.2 The approach used to generate the ESTMs, shown here for lumiflavin (panel 1). We use a vdW surface that is 2 vdW radii from all atoms of the molecule (panel 2). We then locate points on the vdW surface at a density of one point per Å² (panel 3). We place a point charge at each position and, for each individual point, run a single-point excited-state calculation to compute the excitation energy of the chromophore in the presence of that point charge. The points are then represented using a color reflecting the spectral shift they induce in the chromophore (panel 4). Finally, a color-coded ESTM is generated using these points (panel 5). More details are provided in the text. ................................................................. 19

Figure 2.3 Structures and ESTMs of model chromophores of biological photoreceptors. The systems shown are: A. Me-rPSB11; B. MeS-pCA; C. PCB; and D. LF. For LF, the top ESTM is for the first bright excited state (corresponding to the experimental peak at ~450 nm), and bottom ESTM is for the second bright excited state (corresponding to the experimental peak at ~370 nm). The legends indicate the magnitude of the shifts relative
to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses). Relevant atoms are labeled for reference. The corresponding figure with the individual points before the generation of the surfaces are shown in Figure 2.7 for added clarity.

**Figure 2.4** Structures and ESTMs for deprotonated (top) and protonated (bottom) forms of p-hydroxybenzylidene-imidazolinone (pHBDI), a model chromophore of the green fluorescent protein. The legends indicate the magnitude of the shifts relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses). The corresponding figure with the individual points before the generation of the surfaces are shown in Figure 2.8 for added clarity.

**Figure 2.5** Structures and ESTMs for the side chains of aromatic amino acids: A. phenylalanine; B. tyrosine; C. tryptophan $^1L_a$ state (top), and tryptophan $^1L_b$ state (bottom). The legends indicate the magnitude of the shift relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses). The corresponding figure with the individual points before the generation of the surfaces are shown in Figure 2.9 for added clarity.

**Figure 2.6** Plots showing the effect of changes in the charge magnitude, sign, position, and plurality on the spectral shift induced in the LF model chromophore. Each point in the plots corresponds to a point on the ESTM surface. Specifically, we investigate the effect of: A. changing the charge magnitudes from +0.1e to +0.2e; B. changing the charge sign from +0.1e to -0.1e; C. Changing the charge positions from 2 vdW radii to 1 vdW radius (blue), 3 vdW radii (green), and 4 vdW radii (yellow). The inset shows a plot of the slopes from the plot against “hypothetical slopes” that are anticipated from a simple Coulomb model between two points varied by the same distances; D. recomputing the effect of each point charge in the presence of four other permanent point charges. In all plots, linear trendlines
are shown in black and the associated regression equation and $R^2$ values are shown in the top left of each graph.

**Figure 2.7** Structures and Electrostatic spectral tuning maps (ESTMs) for model chromophores of biological photoreceptors. The systems shown are: A. Me-rPSB11; B. MeS-pCA; C. PCB; and D. LF. For LF, the top ESTM is for the first bright excited state (corresponding to the experimental peak at ~450 nm), and bottom ESTM is for the second bright excited state (corresponding to the experimental peak at ~370 nm). The legends indicate the magnitude of the shifts relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses).

**Figure 2.8** Structures and ESTMs for deprotonated (top) and protonated (bottom) forms of p-hydroxybenzylidene-imidazolinone (pHBDI), a model chromophore of the green fluorescent protein. The legends indicate the magnitude of the shifts relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses).

**Figure 2.9** Structures and ESTMs for the side chains of aromatic amino acids: A. phenylalanine; B. tyrosine; C. tryptophan $^1L_a$ state (top), and tryptophan $^1L_b$ state (bottom). The legends indicate the magnitude of the shift relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses).

**Figure 2.10** ESTMs for lumiflavin generated using A. +0.1e point charges; B. +0.2e point charges; and C. -0.1e point charges. The legends indicate the magnitude of the shift relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses).

**Figure 2.11** ESTMs for lumiflavin generated using +0.1e point charges at A. 1 vdW distance; B. 2 vdW distance; C. 3 vdW distance; and D. 4 vdW distance. The legends indicate the
Figure 2.12 ESTMs for lumiflavin generated using A. the normal protocol with +0.1e point charges; and B. the same protocol but in the presence of four permanent +0.1e point charges. The legends indicate the magnitude of the shift relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses). Note that for panel B the reference has been shifted due to the already existing point charges. Panel C. shows the positions and shift magnitudes of the four chosen permanent point charges.

Figure 3.1 Structure of flavins in different redox and protonation states. The atoms are labeled for the oxidized structure.

Figure 3.2 Schematic figure showing the difference between vertical, adiabatic, and vibronic excitations. The vertical excitation (black arrow) is the electronic energy difference between the ground and excited state obtained at the ground state geometry. The adiabatic energy accounting for zero-point vibrational energy correction (green arrow) is the energy difference between the lowest vibrational states of the ground and excited electronic state (v=0 to v’=0). Vibronic excitation energies (green and orange arrows) are energy differences between all ground and excited vibrational states. We assume that the ground state is always in the lowest vibrational state (i.e., we compute 0 K spectra where higher ground vibrational states are not considered). The intensities of vibronic transitions are determined by the overlaps of the ground state wave function (red) and excited state wave functions (blue).

Figure 3.3 Computed (solid lines) and experimental (dashed lines) UV/vis spectra of the oxidized flavin (Fl). The vibronic excitation energies computed in a water PCM model are shown.
as impulse lines in blue. The corresponding broadened peaks are also shown in blue. The solvent shift (computed from differences in vertical excitation energies) and changes in relative intensity associated with different solvents are also shown for gas (light red) and cyclohexane (bright red). The experimental UV/vis spectrum corresponds to aqueous FMN. ESTMs corresponding to the two lowest bright excited states are shown to the right. The legend indicates the magnitude of the shifts relative to the gas-phase reference excitation energy in eV (and in nm in parentheses). The inset shows the first excited state peak simulate with a FWHM of 0.1 eV to show the vibronic band structure in case of smaller broadening.

**Figure 3.4** Computed (solid lines) and experimental (dashed lines) UV/vis spectra of the neutral semiquinone form of flavin (FIH\textperiodcentered). The vibronic excitation energies computed in a methyl propanoate PCM model are shown as impulse lines in magenta. The corresponding broadened peaks are also shown in magenta. The solvent shift (computed from differences in vertical excitation energies) and changes in intensity associated with different solvents are also shown for gas (light red), cyclohexane (bright red), and water (blue). The experimental UV/vis spectrum corresponds to FADH\textperiodcentered in PaNMO. The ESTM corresponding to the lowest bright excited state is shown to the right. The legend indicates the magnitude of the shifts relative to the gas-phase reference excitation energy in eV (and in nm in parentheses).

**Figure 3.5** Computed (solid lines) and experimental (dashed lines) UV/vis spectra of the anionic semiquinone form of flavin (F\textperiodcentered\textperiodcentered). The vibronic excitation energies computed in a methyl propanoate PCM model are shown as impulse lines in magenta. The corresponding broadened peaks are also shown in magenta. The solvent shift (computed from differences
in vertical excitation energies) and changes in intensity associated with different solvents are also shown for gas (light red), cyclohexane (bright red), and water (blue). The experimental UV/vis spectrum corresponds to FAD$^-$ in PaNMO.$^{17}$ ESTMs corresponding to the two lowest bright excited states are shown to the right. The legend indicates the magnitude of the shifts relative to the reference gas-phase excitation energy in eV (and in nm in parentheses).

**Figure 3.6** Computed (solid lines) and experimental (dashed lines) UV/vis spectra of the neutral hydroquinone form of flavin (FlH$_2$). The vibronic excitation energies computed in a water PCM model are shown as impulse lines in blue. The corresponding broadened peaks are also shown in blue. The solvent shift (computed from differences in vertical excitation energies) and changes in intensity associated with different solvents are also shown for gas (light red) and cyclohexane (bright red). The experimental UV/vis spectrum corresponds to aqueous FADH$_2$.52 ESTMs corresponding to the two bright excited states near the 300-400 nm range of the spectrum are shown to the right. The legend indicates the magnitude of the shifts relative to the gas-phase reference excitation energy in eV (and in nm in parentheses).

**Figure 3.7** Computed (solid lines) and experimental (dashed lines) UV/vis spectra of the anionic hydroquinone form of flavin (FlH$^-$). The vibronic excitation energies computed in a water PCM model are shown as impulse lines in blue. The corresponding broadened peaks are also shown in blue. The solvent shift (computed from differences in vertical excitation energies) and changes in intensity associated with different solvents are also shown for gas (light red) and cyclohexane (bright red). The experimental UV/vis spectrum corresponds to aqueous FADH$^-$ in aqueous solution$^{52}$ (black) and in photolyase$^{109}$ (grey). ESTMs
corresponding to the two bright excited states near the 300-450 nm range of the spectrum are shown to the right. The legend indicates the magnitude of the shifts relative to the gas-phase reference excitation energy in eV (and in nm in parentheses).

**Figure 4.1** The tricyclic structure common to FMN, FAD, riboflavin, and lumiflavin.

**Figure 4.2** Approach used to find potential lumiflavin-water interactions. We first optimized lumiflavin in the gas phase (panel 1). We then use a surface that is 3.4 Å from all atoms of the molecule (panel 2) and locate points on the surface at a density of 1 point per 5 Å² (panel 3). Note that 3.4 Å is slightly longer than the length of a typical hydrogen bond distance between heavy atoms. We place an oxygen atom at each point and add two hydrogen atoms to each oxygen to get a total of 106 water positions (panel 4). For lumiflavin in the presence of each one of these waters, we run an unconstrained geometry optimization and frequency calculation (a total of 106 geometry optimization and frequency calculations were performed, each with lumiflavin and one single water molecule). Of the 106 calculations, 92 converged to a structure where the water is hydrogen-bonded or non-covalently bound to the lumiflavin.

**Figure 4.3** Lumiflavin + water structures employed in this study. Structures in orange and blue regions were found using the approach outlined in **Figure 4.2**. Structures in the green region include two water molecules, while structures in the magenta region include six to nine water molecules. The labels, which will be used throughout this work, are indicated next to each structure.

**Figure 4.4** Experimental (bottom panel) and computed (all other panels, each panel representing a different model) IR spectra in the 1750 – 1450 cm⁻¹ region. Four bands dominate the spectra and are assigned to νC = C (blue), νC = N (green), νC2 = O (grey), and νC4 = O.
(yellow). Colored dashed lines are used to help compare calculated and experimental spectra. The background is colored in a way that corresponds to the color scheme in Figure 4.3 to indicate the model used. In part A, calculated frequencies were scaled by 0.964, while in part B the frequencies were scaled so that the $\nu C = C$ calculated frequency is equal to 1548 cm$^{-1}$. The scale factors are indicated on the right side of panel B.

**Figure 4.5 A.** The frequencies of the $\nu C = C$ (blue), $\nu C = N$ (green), $\nu C2 = O$ (grey), and $\nu C4 = O$ (yellow) modes computed with different models. The experimental reference$^{32}$ is indicated with a dashed line, and the corresponding frequency on the right vertical axis. A single scaling factor of 0.964 is used. The background is colored coded as in Figure 4.3 and Figure 4.4 to indicate the model used. **B.** The same data but with the frequencies scaled so that the calculated $\nu C = C$ frequency matches the experiment.

**Figure 4.6** The relative intensities of the $\nu C = C$ and $\nu C = N$ peaks (blue) and $\nu C2 = O$ and $\nu C4 = O$ peaks (green) computed with different models. The experimental references are indicated with a dashed line with corresponding relative intensity shown on the right (obtained by finding the relative areas under the peaks from the experimental FTIR spectrum in ref $^{34,32}$).

**Figure 4.7 A.** The frequencies of the $\nu C = C$ (blue), $\nu C = N$ (green), $\nu C2 = O$ (grey), and $\nu C4 = O$ (yellow) modes computed with different basis sets with the PCM (circles), 6W atomistic (triangles), and 9W atomistic (squares) models. The experimental reference is indicated with a dashed line and the corresponding frequency on the right. Frequencies were scaled such that the C=C computed frequency matches the experiment. **B.** The relative intensities of the $\nu C=C$ and $\nu C=N$ peaks (blue) and $\nu C2=O$ and $\nu C4=O$ peaks (green) computed with different basis sets with the PCM (circles), 6W atomistic (triangles), and 9W atomistic (squares).
models. The experimental reference, indicated with a dashed line and the corresponding relative intensity on the right, is obtained by integrating to find the relative areas under the peaks from the experimental FTIR spectrum from ref. 32.

**Figure 4.8** Experimental FTIR absorption spectra for FMN or FAD in H$_2$O or D$_2$O. The spectra from references 1-7, as labeled on the right for each spectrum.

**Figure 4.9** The normal mode vectors corresponding to the four prominent FTIR peaks discussed in this study.

**Figure 4.10** Computed (for lumiflavin) and experimental$^{2,7}$ (for FMN) FTIR absorption spectra in both deuterated (top four panels) and undeuterated (bottom four panels) solvent. In the deuterated PCM model, only the ionizable N$_3$ hydrogen is deuterated, while in the cluster (6W and 9W) deuterated models the explicit waters are also deuterated.

**Figure 4.11** Left. The normal mode vectors for the additional intense peak that appears in PCM calculations (but that is not intense in other atomistic calculations or the experiment). Right. The normal mode vectors appearing at around 1624 cm$^{-1}$ in water cluster models. This peak is close in frequency to the C$_2$=O vibrational mode, which may explain why the experimental C$_2$=O vibrational peak is broader than the C$_4$=O peak, even appearing to have a shoulder in some D$_2$O experimental spectra.

**Figure 5.1** The isoalloxazine ring of FMN and atom number labels. R=CH$_3$ for lumiflavin and R=ribose-5′-phosphate for FMN.

**Figure 5.2** Steps in the automated ASEC-FEG protocol for flavoproteins. Ovals with solid borders indicate required user input, while ovals with dashed borders indicate optional user input.
Figure 5.3 FMN inside the binding pocket of iLOV from PDB 4EES. The dashed lines indicate the distance in Angstroms between the Q489 and Q430 glutamine side chain nitrogen atoms and the closest FMN nitrogen atom. The red circle indicates the hydrogen Link Atom (LA), which separates the QM subsystem (the lumiflavin) and the MM subsystem (the ribose-5′-phosphate group, protein, and solvent). The figure was prepared using PyMol.

Figure 5.4 A. An ESTM map reproduced from ref. 48. The map indicates the change in the vertical excitation energy between the ground state ($S_0$) and the first singlet excited state ($S_1$) introduced by a +0.1-probe charge placed at the van der Waals surface of lumiflavin in its $S_0$ equilibrium geometry. The map suggests mutations that redshift the excitation energy (see arrows and labels). The legend indicates the magnitude of the excitation energy shifts relative to the gas-phase reference excitation energy in eV (and in nm in parentheses).

B. The same ESTM map was computed at the $S_1$-optimized lumiflavin geometry, corresponding to the fluorescent minimum. The legend indicates the magnitude of the emission energy shifts relative to the gas-phase reference emission energy in eV (and in nm in parentheses).

Figure 5.5 Center: LoProp charge population analysis (with hydrogen atom charges summed onto the heavy atoms they are connected to) for lumiflavin in its ground ($S_0$) and first singlet excited ($S_1$) state. Red circles indicate negative charge density, and blue circles indicate positive charge density. The area of the circles is directly proportional to the charge on the corresponding atom. The $S_1$ and $S_0$ states have slightly different charge distributions that are difficult to discern without close inspection. Therefore, the difference in the atomic charges ($S_1$–$S_0$) is also shown in the middle. In the $S_1$–$S_0$ difference plot, the areas of the
circles are proportional to the magnitude of the charge difference between \( S_0 \) and \( S_1 \), with red circles indicating reduced charge (higher electron density) on the atom after excitation from \( S_0 \) to \( S_1 \) and blue circles indicating increased charge (lower electron density). **Left:** A scheme illustrating the effect of placing a positively charged lysine side chain close to the flavin N5/C4a, which would result in a redshift (see text for details). **Right:** A scheme illustrating the effect of placing a negatively charged glutamate side chain close to the flavin N1, which should also result in a redshift (see text for details).

**Figure 5.6** This figure shows three representatives QM/MM optimized snapshots to show the distance between the flavin N1 and the Q430 (iLOV) or E430 (iLOV-Q430E) side chain. The labeled distances are reported averages from one of the production runs of the MD simulations. Specifically, we label the distance between the flavin N1 and the Q430 side chain nitrogen in iLOV (3.5 Å), the E430 side chain carboxylate carbon in the ground-state optimized system (E430\( ^{\text{abs}} \), 8.9 Å), and the E430 side chain carboxylate carbon in the excited-state optimized system (E430\( ^{\text{fl}} \), 8.4 Å). The figure was prepared using PyMol.

**Figure 5.7** UV-visible absorption spectra of iLOV (panel A, blue), iLOV-Q430E (panel B, red), and free FMN (Panel C, black). The spectra were recorded in 20 mM pH 7.0 phosphate buffer solution and 15 °C.

**Figure 5.8** The excitation and emission spectra of iLOV (panel A, blue), iLOV-Q430E (panel B, red), and free FMN (panel C, black). The spectra were recorded in 20 mM pH 7.0 phosphate buffer and 15 °C.

**Figure 5.9** A. Difference excitation spectra of iLOV and iLOV-Q430E minus that of free FMN. B. Difference emission spectra of the iLOV and iLOV-Q430E minus that of free FMN.
The reference line in panels A and B represents the ∆EX (free FMN-Free FMN) and ∆EM (free FMN-Free FMN), respectively................................................................. 150

**Figure 5.10** A plot of first excited state excitation energy vs active space for B3LYP and CASPT2 methods. The excitation energies were calculated from MP2-optimized geometry with OpenMolcas and Gaussian software................................................................. 154

**Figure 6.1** Orbital entanglement diagram of oxidized flavin............................................ 174

**Figure 6.2** (a) π-bonding and (b) π-anti-bonding orbitals for oxidized flavin............... 175

**Figure 6.3** SA-CASSCF, SS-CASPT2, MS-CASPT2, XMS-CASPT2, XDW-CASPT2, MC-PDFT, and TD-B3LYP excitation energies of oxidized quinone with different active spaces and the number of states included in the state averaging. (a) two states included in the state averaging, (b) three states included in the state averaging, (c) four states included in the state averaging, (d) five states included in the state averaging, (e) six states included in the state averaging, and (f) seven states included in the state averaging............................... 177

**Figure 6.4** CASSCF, SS-CASPT2, MS-CASPT2 XMS-CASPT2, XDW-CASPT2, MC-PDFT, and TD-B3LYP excitation energy of (a) neutral semiquinone radical, (b) anionic semiquinone radical, (c) neutral hydroquinone, and (d) anionic hydroquinone with different active spaces and three number of states in the state averaging............................... 179

**Figure 6.5** TD-DFT/cc-pVTZ PESs of flavin: (a) from S₀ minimum to the S₁ minimum. (b) from S₁ minimum to the S₂ minimum. ................................................................. 180

**Figure 6.6** MS-CASPT2 PESs of flavin: (a) from S₀ minimum to the S₁ minimum. (b) from S₁ minimum to the S₂ minimum................................................................. 181

**Figure 6.7** Comparison of shape of excited states of linearly interpolated geometries between S₀ minimum to the S₁ minimum of flavin at different level of theories such as SA-CASSCF,
Figure 6.8 Comparison of shape of excited states of linearly interpolated geometries between $S_1$ minimum to the $S_2$ minimum of flavin at different level of theories such as SA-CASSCF, CASPT2, MC-PDFT, DFT, EOM-EE-CCSD, SOS-CISD: (a) $S_0$ (b) $S_1$ ($\pi\pi^*$) (c) $S_2$ ($\pi\pi^*$) (d) $T_1$ ($3\pi\pi^*$) (e) $T_2$ ($3\pi\pi^*$).

Figure 6.9 Effect of different methods on ISC between $S_1$ ($\pi\pi^*$) and $T_2$ ($\pi\pi^*$) states of flavin.

Figure 6.10 Orbital entanglement diagram of (a) neutral semiquinone (b) anionic semiquinone, (c) neutral hydroquinone, and (d) anionic hydroquinone.

Figure 6.11 $\pi$-bonding and $\pi$-anti-bonding orbitals of the (a) neutral semiquinone radical, (b) anionic semiquinone radical, (c) neutral hydroquinone, and (d) anionic hydroquinone.

Figure 6.12 CASSCF, SS-CASPT2, MS-CASPT2, XMS-CASPT2, XDW-CASPT2, MC-PDFT, and TD-B3LYP excitation energies of neutral semiquinone radical with different active spaces and the number of states included in the state averaging. (a) two states included in the state averaging, (b) three states included in the state averaging, (c) four states included in the state averaging, (d) five states included in the state averaging, (e) six states included in the state averaging, and (f) seven states included in the state averaging.

Figure 6.13 CASSCF, SS-CASPT2, MS-CASPT2, XMS-CASPT2, XDW-CASPT2, MC-PDFT, and TD-B3LYP excitation energies of anionic semiquinone radical with different active spaces and the number of states included in the state averaging. (a) two states included in the state averaging, (b) three states included in the state averaging, (c) four states included...
in the state averaging, (d) five states included in the state averaging, (e) six states included in the state averaging, and (f) seven states included in the state averaging...

Figure 6.14 CASSCF, SS-CASPT2, MS-CASPT2, XMS-CASPT2, XDW-CASPT2, MC-PDFT, and TD-B3LYP excitation energies of neutral hydroquinone with different active spaces and the number of states included in the state averaging. (a) two states included in the state averaging, (b) three states included in the state averaging, (c) four states included in the state averaging, (d) five states included in the state averaging, (e) six states included in the state averaging, and (f) seven states included in the state averaging.

Figure 6.15 CASSCF, SS-CASPT2, MS-CASPT2, XMS-CASPT2, XDW-CASPT2, MC-PDFT, and TD-B3LYP excitation energies of anionic hydroquinone with different active spaces and the number of states included in the state averaging. (a) two states included in the state averaging, (b) three states included in the state averaging, (c) four states included in the state averaging, (d) five states included in the state averaging, (e) six states included in the state averaging, and (f) seven states included in the state averaging.

Figure 7.1 The isoalloxazine ring of FMN and atomic labels. R=CH₃ for lumiflavin and R= ribose-5'-phosphate for FMN.

Figure 7.2 Lumiflavin with water structures was employed in this study.

Figure 7.3 Interpolated geometry along a) twisting out of the plane b) bond length alternation (BLA).

Figure 7.4 Relevant frontier molecular orbitals of lumiflavin at its S₀ optimized geometry.
1 INTRODUCTION

1.1 Overview

In 1879, A. W. Blyth, an English chemist, extracted a bright yellow pigment from cow’s milk named riboflavin.\(^1\) Riboflavin, also known as vitamin B2, is present in food and sold as a dietary supplement. The name “riboflavin” originated from “ribose,” a reduced sugar of ribitol, and ‘flavin’, which is an isoalloxazine ring that functions as a cofactor in flavoproteins. The flavin cofactor can be present in flavoproteins either as flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD) (Figure 1.2). Furthermore, it can exist in one of the three redox states: oxidized form quinone (Fl), one-electron/one-proton coupled reduced from semiquinone radical (FlH\(^{+}\)), and two-electron/two-proton coupled fully reduced hydroquinone (FlH\(_2\)). Since the pKa values for quinone, semiquinone and hydroquinone are 10.3, 8.3, and 6.7 in an aqueous solution, respectively,\(^2-4\) semiquinone and hydroquinone can have anionic forms under physiological conditions: anionic semiquinone radical (Fl\(^{+}\)) and anionic hydroquinone (FlH\(^{+}\)).

![Figure 1.1 Structure of flavin in different redox and protonation states.](image-url)
Overall, flavin can be present in flavoproteins one of the five different redox and protonation states at physiological conditions, making it highly versatile. Of these states, two redox pairs, neutral quinone/anionic semiquinone (Fl/Fl{-}) and neutral semiquinone/anionic hydroquinone (FlH'/FlH{-}), are frequently involved in electron-transfer reactions (Figure 1.1).

Flavoproteins are found in all forms of life, where they play crucial roles in a variety of enzymological reactions and metabolic processes. These processes include photosynthesis, aerobic and anaerobic respiration, and denitrification. Additionally, flavoproteins aid in the metabolism of carbohydrates, lipids, and proteins, synthesis of vitamin B12 and pyridoxal 5’-phosphate, as well as DNA repair and regulation of caspase-independent apoptosis.

Figure 1.2 Structure of Lumiflavin (LF), riboflavin (RF, vitamin B2), flavin mononucleotide (FMN), and flavin adenine dinucleotide (FAD).
1.2 Flavoprotein Photoreceptors

Flavin acts as a chromophore in light-responsive proteins, including cryptochromes (CRY),\textsuperscript{15} light-oxygen-voltage (LOV) sensing domains,\textsuperscript{16} blue-light sensing (BLUF) domains,\textsuperscript{17} and DNA photolyases.\textsuperscript{18,19} The flavin chromophore absorbs blue light (in the range of \textasciitilde450 nm) and undergoes various photochemical changes. Generally, photoexcitation of the flavin induces covalent bond formation in LOV, hydrogen-bond (HB) rearrangements in BLUF, and redox reactions in CRY. The photochemical events initiate with the structural change of the chromophore followed by conformational changes of the surrounding microenvironment, ultimately resulting in distal conformational changes at the protein surface leading to the biological response.

1.2.1 Cryptochrome

Cryptochromes are ancient and highly conserved flavoproteins that exist in all kingdoms of life. They are two-chromophore proteins containing either a pterin or a deazaflavin (8-hydroxy-8-demethyl-5-deazariboflavin, 8-HDF) and FAD.\textsuperscript{20} Cryptochromes have a very similar structure to photolyases enzymes involved in repairing UV-induced DNA damage.\textsuperscript{20,21} In contrast, cryptochromes lack DNA repair activity except for cry-DASH (\textit{Drosophila}, \textit{Arabidopsis}, \textit{Synechocystis}, and \textit{Homo} cryptochrome).\textsuperscript{22} Cryptochromes play a pivotal role in the generation and maintenance of circadian rhythms in animals and plants.\textsuperscript{23} In \textit{Drosophila}, cryptochrome (dCRY) acts as a blue-light photoreceptor that directly modulates light input into the circadian clock,\textsuperscript{24} while in mammals, cryptochromes (CRY1 and CRY2) act as transcription repressors within the circadian clockwork.\textsuperscript{25} Some insects, such as the monarch butterflies, possess both a mammal-like and a \textit{Drosophila}-like version of cryptochrome, demonstrating an ancestral clock mechanism involving both light sensing and transcriptional repression by cryptochrome.\textsuperscript{26,27}
1.2.2 LOV domain

After a few years of cryptochrome identification, phototropin (phot) came into the scene. Phot is one of the blue light (BL) receptor proteins used by many higher plants, microalgae, fungi, and bacteria to sense environmental conditions. In higher plants, they control phototropism, chloroplast movement, stomatal opening, and leaf expansion. In fungal organisms, they adjust the circadian temporal organization of the cells to the daily and seasonal periods. Most plants contain two types of phot, namely phot1 and phot2.

The plant photos comprise ~950–1000 amino acid residues and two non-covalently bound flavin mononucleotide (FMN) molecules folded into three functional domains. The N-terminal half contains two LOV (light-oxygen-voltage-sensing) domains, designated as LOV1 and LOV2, which belong to the PAS (Per-Arnt-Sim) superfamily. The LOV domain usually contains ~110 amino acids. The C-terminal half of phot is a serine-threonine kinase (STK) domain classified into group VIII of the AGC family. Each LOV domain receives BL via the FMN chromophore and undergoes photochemical reactions.

1.2.3 BLUF domain

A third class of flavin-containing blue-light receptors domain has been identified in *Euglena gracilis*. This domain also absorbs blue light through flavin adenine dinucleotide (FAD) chromophore, and is therefore called BLUF (blue light sensing using FAD). It is comparable in size to the LOV domain at ~100-110 amino acids. The BLUF domain has been found in proteins from several branches of the phylogenetic tree of Bacteria and in photosynthetic lower Eukarya. The BLUF domain is also identified in the AppA protein, which acts as a blue light and redox sensor in regulating photosynthesis genes in *Rhodobacter sphaeroides*. In the dark state, BLUF domains show typical features of an oxidized flavin, and their very peculiar
photocycle involves the reversible formation of a ~10 nm redshifted intermediate (BLUF\textsubscript{Red}) without apparent variations in the flavin redox state.\textsuperscript{39}

1.3 LOV domain Photocycle

The photophysical and photochemical properties of flavoproteins are controlled by the energetics of the flavin chromophore. Upon absorption of blue light (~447nm), the flavin chromophore behaves differently in different protein environments. In the \textit{Chlamydomonas reinhardtii} LOV domain, the flavin cofactor excites to the singlet excited state and then rapidly decays to a long-lived triplet state via an intersystem crossing (ISC) that Magerl and co-workers have experimentally proved.\textsuperscript{40, 41} They also reported that a small portion (~20\%) of the triplet state flavin returns to the ground state and gives phosphorescence. The rest of the triplet state flavin (~80\%) reacts with the conserved cysteine residue in the active site and forms flavin-cysteinyln-thiol photo-adduct.\textsuperscript{40, 41} Interestingly, Zhu and co-workers found in \textit{Chlamydomonas reinhardtii} LOV2 that the adduct formation can also happen from the singlet excited state instead of the triplet state.\textsuperscript{42} Finally, the photo adduct reverts to its dark state (dissociated form) in a few seconds to hours (\textbf{Figure 1.3}). Besides, recently engineered flavin binding proteins, \textit{Arabidopsis thaliana} iLOV, and miniSOG do not have cysteine residue in the vicinity of the flavin cofactor.\textsuperscript{43-45} Therefore, these proteins lose their ability to form photo adducts at the excited states. As a result, the iLOV protein gains fluorescence properties with a quantum yield of 0.24, while miniSOG acquires a longer triplet state lifetime of 33.6 µs.\textsuperscript{46} The triplet state lifetime of miniSOG is even more increased when the electron-rich amino acids are replaced by less electron-rich amino acids.\textsuperscript{47, 48}
Figure 1.3 Electronic states involved in LOV’s photocycle and their applications.

1.4 Outline of Chapters and Goal of the Research

The focus of this dissertation is on flavin and flavin-binding fluorescent proteins. Flavin, a component of flavoproteins, exists in five different redox and protonation states under physiological conditions. These states can be used to probe the microenvironment of proteins through simple redox and pH changes, making them a useful tool in biotechnology. However, there is a lack of information on how the spectral properties of flavin are affected by the protein or solvent environment.

To better understand this, we developed a computational protocol to generate electrostatic spectral tuning maps (ESTMs) in chapter 2. This protocol allows us to understand how a biological chromophore will spectroscopically behave in different electrostatic environments. In chapter 3,
we simulated the near-UV/vis spectra of flavin in its five redox/protonation states, including its vertical, adiabatic, and vibronic excitations. We also modeled the effects of different solvents on the spectral shifts and relative intensities.

In chapter 4, we used explicit solvent and continuum solvation models to reproduce the vibrational frequencies and intensities of prominent peaks in flavin's IR spectrum in an aqueous solution. We also explored how intermolecular interactions with the solvent, such as hydrogen bonding, affect the vibrational frequencies and intensities of the isoalloxazine ring of flavin.

In chapter 5, we aimed to engineer a red-shifting variant of the flavin-binding fluorescent protein iLOV. This was done by introducing a negatively charged amino acid via a single-point mutation (Q430E). To understand the effect of the mutation, we conducted hybrid quantum mechanical/molecular mechanical calculations using the average solvent electrostatic configuration free energy gradient approach. This was followed by laboratory expression of iLOV-Q430E and measurement of its absorption, excitation, and fluorescence spectra to verify if the spectra were indeed red-shifted relative to iLOV.

To accurately describe the relative energies between the ground and excited states of flavin, we employed the complete active-space self-consistent field (CASSCF) method, followed by second-order perturbative energy correction (CASPT2). Choosing a suitable active space for this method was critical as the computational cost increased with the size of the active space. Hence, a benchmark study was performed in Chapter 6 to determine the vertical excitation energies of flavin in different redox states.

In Chapter 7, we carried out a study on flavin model systems to understand the influence of non-bonding interactions like hydrogen bonding on the energetics of the low-lying excited states
of flavin. This was done with the goal of understanding the differences in the behavior of the flavin chromophore in different protein environments.

Overall, our research outcomes will enhance our understanding of how proteins, specifically when bound to the oxidized form of the flavin cofactor, impact the absorption and fluorescence energies of flavins. This will aid in the development of new flavoprotein color derivatives for biotechnological applications. The aim of the study was to engineer a red-shifted variant of iLOV and to comprehend the effects of the mutation and the interactions between flavin and its protein environment. The results of the study have potential applications in multicolor bioimaging and FRET-based biosensing.
1.5 References


2 ELECTROSTATIC SPECTRAL TUNING MAPS FOR BIOLOGICAL CHROMOPHORES

Yoelvis Orozco-Gonzalez, Mohammad Pabel Kabir, and Samer Gozem


Copyright © The American Chemical Society 2019

My contribution to this chapter was data curation, formal analysis, and co-writing with Drs. Orozco-Gonzalez and Gozem.

2.1 Abstract

The mechanism by which the absorption wavelength of a molecule is modified by a protein is known as spectral tuning. Spectral tuning is often achieved by electrostatic interactions that stabilize/destabilize or modify the shape of the excited and ground-state potential energy surfaces of the chromophore. We present a protocol for the construction of three-dimensional “electrostatic spectral tuning maps” that describe how excitations in a chromophore are influenced by nearby charges. The maps are built by moving a charge on the van der Waals surface of the chromophore and calculating the change in its excitation energy. The maps are useful guides for protein
engineering of color variants, for interpreting spectra of chromophores that act as probes of their
environment, and as starting points for further quantum mechanical / molecular mechanical
(QM/MM) studies. The maps are semi-quantitative and can approximate the magnitude of the
spectral shift due to a point charge at a given position with respect to the chromophore. We
generate and discuss electrostatic spectral tuning maps for model chromophores of photoreceptor
proteins, fluorescent proteins, and aromatic amino acids. Such maps may be extended to other
properties such as oscillator strengths, absolute energies (stability), ionization energies, and
electron affinities

2.2 Introduction

Natural systems have evolved different spectral tuning mechanisms to modulate the color of
absorption of biological pigments. In some cases, it is achieved through chemical modification of
the chromophore, while in others it is through protein-chromophore interactions that modify the
absorption properties of the chromophore. The understanding of spectral tuning has been of broad
interest; not only does it help explain the diversity of colors observed in nature, it also provides
principles for bioengineering of new color variants of pigments for practical applications.
Examples include the engineering of color variants of fluorescent proteins, which has been
instrumental for multicolor and deep-tissue fluorescent imaging, as well as engineering of color
variants of optogenetics tools for multiwavelength optogenetics and tools for deeper penetration
depth.

In cases where chemical modification of the chromophore is not easily achievable or desirable, the
only approach for spectral tuning is through the modification of the protein to affect protein-
chromophore interactions. Such interactions are either steric or electronic. Steric interactions
imposed by the protein may modify the geometry of the chromophore, and thereby modulate its
absorption spectrum. This usually involves imposing torsional strains that modify the conjugation length in the chromophore.\textsuperscript{11-12} Direct electronic effects include modification of the energetics of the ground- and excited-state energetics through $\pi$-stacking (as in the yellow fluorescent protein\textsuperscript{13-14}) or hydrogen bonding. However, in the absence of such interactions, or in addition to them, spectral tuning is often achieved by electrostatic protein-chromophore interactions. These electrostatic interactions could either modulate the relative energies of the ground and excited states or they can affect the shape/curvature of the excited state and thereby modulate the vertical excitation energies.\textsuperscript{15} Spectral tuning through electrostatic interactions is perhaps best exemplified in photopsins, the pigments responsible for color (blue, green, and red) vision. All three types of photopsins employ the same chromophore, a retinal protonated Schiff base (rPSB), but absorb at very different wavelengths ($\lambda_{\text{max}} \approx 430$ nm, 530 nm, and 560 nm, respectively).\textsuperscript{16} Recently, Wang et al. have demonstrated that rPSB can be spectrally tuned over absorption wavelengths covering the entire range of the visible spectrum by engineering rhodopsin analogues exhibiting $\lambda_{\text{max}}$ ranging from 425 nm to 644 nm.\textsuperscript{17} This was largely achieved by exploiting a simple electrostatic spectral tuning principle for the first bright excited state of rPSB, summarized in Fig. 1.\textsuperscript{18-19} The $S_0 \rightarrow S_1$ excitation in rPSB is accompanied by a partial transfer of the positive charge, which is initially localized on the Schiff base nitrogen in the ground state, towards the $\beta$-ionone ring of the molecule.\textsuperscript{20-21} The different electron distribution in the rPSB excited state compared to the ground state means that the ground and excited state have different electrostatic interactions with the protein environment. Specifically, a positive charge introduced near the protonated Schiff base nitrogen would destabilize the ground state (which has a positive charge localized at the nitrogen) more than the excited state (which has a more delocalized positive charge), thereby red-shifting the absorption wavelength. The opposite is true for a positive charge introduced in the
vicinity of the β-ionone, or for a negative charge near the Schiff base (see Figure 2.1). This is primarily the spectral tuning principle that guided the spectral tuning of rPSB in the work by Wang et al.\textsuperscript{17} However, the electrostatic spectral tuning mechanism summarized in the schematic Figure 2.1 is the result of numerous experimental and computational studies aimed at understanding the detailed spectral tuning mechanism rPSB in different environments. An incomplete list of such studies includes refs. 18-52.

Figure 2.1 Schematic figure presenting the electrostatic spectral tuning mechanism of all-trans rPSB. Introduction of a positive electrostatic potential near the Schiff base nitrogen results in a red-shift of the absorption energy, while a positive potential near the β-ionone segment would result in a blue-shift. The opposite is true for a negative electrostatic potential. Adapted from refs 18-19 with permission.

Computational studies, and in particular the hybrid quantum mechanical / molecular mechanical (QM/MM) approach,\textsuperscript{53} have played an important role in elucidating the spectral tuning mechanism in retinal proteins as well as in other biological photoreceptors. In such studies, the chromophore is treated at a quantum mechanical (QM) level of theory capable of properly describing its spectroscopy, while the effect of the protein environment is usually accounted for using a molecular mechanics (MM) force field.\textsuperscript{54} Electrostatic embedding is used to account for polarization of the QM chromophore by the protein MM environment.\textsuperscript{55} In many models, van der Waals (vdW) QM-MM interactions are included, but typically these vdW QM-MM interactions
are treated simply with MM force fields (e.g., Lennard Jones \(^5\) potential),\(^5\) and therefore do not have an effect on the excitation energy of the chromophore because they are equal for the ground and excited states in these models. More advanced models include other interaction terms, such as polarization of the MM atoms, dispersion interactions, exchange, and charge transfer interactions.\(^5\) However, spectral tuning is often dominated by protein-chromophore electrostatic interactions, which is why even minimal QM/MM models that include only electrostatic QM-MM interactions have found considerable success in the study of spectral tuning in different systems.

Often, spectral tuning studies employ QM/MM model of a specific system and then performing mutations \textit{in silico} or turn off charges of specific amino acids to understand the effect of these mutations/amino acids on the absorption properties of the chromophore. However, spectral tuning through electrostatic interactions can often be qualitatively understood simply from intrinsic properties of the chromophore. This has been repeatedly demonstrated for rPSB, where the spectral tuning due to electrostatics can be summarized using a single Figure (\textbf{Figure 2.1}) for a large number of retinal-binding proteins.\(^1\),\(^5\) In fact, the same spectral tuning principle applies for scotopic and color visual rhodopsins,\(^2\),\(^24\),\(^34\)-\(^37\),\(^43\)-\(^44\),\(^51\)-\(^52\) microbial rhodopsins,\(^1\),\(^8\),\(^24\)-\(^25\),\(^35\),\(^38\),\(^45\) and melanopsins,\(^46\)-\(^47\) even though these subfamilies may host different stereoisomers of rPSB.

The same spectral tuning mechanism even applies in artificial retinal analogues synthesized from human cellular retinol-binding protein II,\(^1\),\(^7\),\(^48\)-\(^50\),\(^52\) which are proteins with a completely different secondary structure than rhodopsins. That is because the charge transfer character of the first excited state in rPSB is an intrinsic property of rPSB that is common to all these proteins, and is therefore general.
A rudimentary understanding of the tendency of a chromophore for electrostatic spectral tuning can be obtained from its transition dipole moment, which is a vector representing the change in electronic structure of the molecule upon excitation. However, this only gives a coarse-grained account of how a system will respond to its environment, without really revealing the details of the interactions of an external charge with the molecule. A more informative approach would be to compute the electrostatic potential of ground and excited state and find the difference, but this process of plotting electrostatic potential differences is not readily available in many QM software packages, nor does it account for perturbations in the QM wave function due to the external point charges. Here, we present an approach to investigate an intrinsic property of a system; how its gas-phase excitation energy is affected by a point charge introduced in its vicinity. The resulting “map” can be used to understand the electrostatic effect of an environment on the chromophore in general, as opposed to looking at a specific mutation in a given system. We use the resulting “Electrostatic Spectral Tuning Maps” (ESTMs) to explore the potential for spectral tuning through electrostatic interactions in biological chromophore models of photoreceptor proteins, the green fluorescent protein (GFP), and of aromatic amino acids. These ESTMs are intuitive, fast to generate, and therefore very useful for providing a preliminary understanding of the electrostatic spectral tuning potential of molecules. They can be used as starting point for further QM/MM spectral tuning studies.

2.3 Methodology

Figure 2.2 summarizes the computational protocol used to generate the ESTM for chromophores in this study. First, the geometry of each molecular system was optimized in the gas phase using density functional theory (DFT) with the B3LYP functional and 6-31+G* basis set, which includes diffuse and polarization functions. In the case of flexible chromophores that may adopt
several stereoisomers or configurations, a structure is optimized that is representative of a common biological role. For example, a for the rPSB model an 11-cis isomer is used, representing the dark state configuration of visual rhodopsins, while the 5-Z,syn-10-Z,syn-15-Z,anti isomer of phycocyanobilin is chosen, representing the typical dark state configuration of bilins in proteins from the phytochrome family.\textsuperscript{64-65} After the molecule is optimized, the Cartesian coordinates of points lying on a vdW surface at a distance of 2 vdW radii are determined (see second and third panels of Figure 2.2). The vdW surface of a molecule is a sum of the vdW surfaces of all individual atoms excluding the overlapping regions. Bondi’s radii were used to define atomic vdW radii.\textsuperscript{66} We chose a surface that is two vdW radii because this is the closest distance that an external atom can be from the chromophore before repulsive forces dominate. The Cartesian coordinates of points on the vdW surfaces were generated using the open-source python library “pyvdwsurface”.\textsuperscript{67} The density of points used is one point per Å\textsuperscript{2}.

**Figure 2.2** The approach used to generate the ESTMs, shown here for lumiflavin (panel 1). We use a vdW surface that is 2 vdW radii from all atoms of the molecule (panel 2). We then locate points on the vdW surface at a density of one point per Å\textsuperscript{2} (panel 3). We place a point charge at each position and, for each individual point, run a single-point excited-state calculation to compute the excitation energy of the chromophore in the presence of that point charge. The points are then
represented using a color reflecting the spectral shift they induce in the chromophore (panel 4). Finally, a color-coded ESTM is generated using these points (panel 5). More details are provided in the text.

Next, we place a positive charge of $0.1e$ at each of the previously defined points, and a QM calculation was performed for each point charge to compute the effect of that individual point charge on the absorption spectrum of the chromophore. For example, 300 QM calculations were performed for phenylalanine, the smallest system, and 926 QM calculations for phycocyanobilin, the largest system. Six excited states were computed in all cases using time-dependent DFT (TD-DFT)$^{68}$ at the B3LYP/6-31+G* level, except for phycocyanobilin, where just the first excited state was computed to reduce the computational cost. From the six excited states computed, the lowest energy root with an oscillator strength above 0.01, corresponding to the first bright excited state, was chosen to generate the ESTM. Note that most of the chromophores investigated in this work have a first bright excited state that is well separated from other bright states, and so can be identified unambiguously with TD-B3LYP. The only exceptions were flavin, which experimentally has two prominent peaks in the near-UV/vis range and for which we generated two ESTMs (the first root and the fourth root, corresponding to the experimental peaks at ~450 nm and ~370 nm, respectively), and tryptophan, which has two nearly degenerate low-lying absorbing states and for which we also generated two ESTMs. In a few cases, i.e. for a few specific point charges in some of the molecules, root switching or state mixing occurred where the relevant excited state moved to a different TD-DFT root or mixed with another state. In those cases, the energy of the relevant excited state is computed as an average of the mixing states weighted by their oscillator strengths compared to the gas-phase system (the reference).
After completing the single-point calculations, the point charges were plotted around the chromophore and colored according to their impact on the spectral properties of the chromophore, as shown in the fourth panel of Figure 2.2; red indicates that the point charge induced a red shift in the chromophore absorption, while blue indicates that the point charge induced a blue shift. The color saturation is related to the magnitude of the shift, as indicated in the legend.

To fill in the gaps between the points and generate a surface as shown in the last panel, we performed a Delaunay triangulation for all the points on the vdW surface. The triangles are colored using linear interpolation of the point colors. The Mayavi python library was used to achieve this.

To allow for direct comparison between different chromophores, the ESTMs of all the chromophores were produced in the same way, as described above, using a 2 vdW radii surface, 1 point/Å² density, and a charge magnitude of +0.1e. However, additional tests were also performed for the first excited state of lumiflavin. These tests include using a different sign and magnitude for the point charges, computing the surface in the presence of multiple point charges, and using vdW surfaces with different radii. To keep the number of point charges the same using different radii, surfaces were generated at 1 vdW, 3 vdW, and 4 vdW radii using 3.74, 0.365 and 0.175 point/Å² densities, respectively. The results of these tests are discussed in the Results and Discussion Section (Figures 2.10-2.12).

All DFT and TD-DFT calculations accounting for the effect of point charges in this work were performed with Gaussian.
2.4 Result and Discussion

2.4.1 Chromophores of Photoreceptor proteins

Van der Horst and Hellingwerf classify photoreceptor proteins into six main families. Three of these families have distinct chromophores; rhodopsins, xanthopsins, and phytochromes, which employ rPSB, p-coumaric acid, and bilins as their chromophores, respectively. Note that the photopsins discussed in the introduction belong to the rhodopsin family of proteins. The three other families, cryptochromes (CRY), light-oxygen-voltage sensing domains (LOV), and blue-light using flavin adenine dinucleotide (BLUF), are all flavoproteins that employ either flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD). Both FMN and FAD share the flavin group that is responsible for their spectral properties.

Using the approach described in the Methodology Section, we constructed ESTMs for gas-phase models of each of those biological chromophores (see Figure 2.3). For rhodopsins, we used a N-methylated 11-cis rPSB model (Me-rPSB11). The 11-cis isomer is more specifically a model for visual rhodopsins. For xanthopsins, we employ thiomethyl-p-coumaric acid (MeS-pCA) as a model chromophore of the photoactive yellow protein (PYP). Different phytochrome subfamilies may employ different bilins, but all subfamilies host a chromophore with the same tetrapyrrole core structure, and so we use phycocyanobilin (PCB) as the model in this work. Note that PCB has a similar conjugated backbone structure as both biliverdin (BV) and phytochromobilin (PΦB) with the only difference being in the substituents at the A and D rings. Therefore, PCB may be considered a model system for all three bilins. Finally, for the flavoprotein photoreceptors (CRY, LOV, BLUF, and other flavoproteins), the oxidized form of lumiflavin (LF) is the model used. The results of these calculations are shown in Figure 2.3 and are discussed below.
Figure 2.3 Structures and ESTMs of model chromophores of biological photoreceptors. The systems shown are: A. Me-rPSB11; B. MeS-pCA; C. PCB; and D. LF. For LF, the top ESTM is for the first bright excited state (corresponding to the experimental peak at ~450 nm), and bottom ESTM is for the second bright excited state (corresponding to the experimental peak at ~370 nm). The legends indicate the magnitude of the shifts relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses). Relevant atoms are labeled for reference.
The corresponding figure with the individual points before the generation of the surfaces are shown in Figure 2.7 for added clarity.

2.4.2 Rhodopsins

The spectral tuning mechanism of rPSB has been discussed at length in literature.\textsuperscript{18-19, 22-52} The ESTM obtained for Me-rPSB11 is in agreement with accounts in literature (compare Figure 2.3A to Figure 2.1). The excitation of rPSB from S\textsubscript{0} to S\textsubscript{1} is accompanied by a partial transfer of the positive charge from the nitrogen atom towards the β-ionone ring, and a large part of that transferred positive charge in the excited state resides on the tertiary C\textsubscript{5} and C\textsubscript{9} carbons.\textsuperscript{79} Interestingly, although the positive charge on C\textsubscript{9} increases in the excited state, the blue-shifting part of the spectral tunings surface only begins near C\textsubscript{7}, and the most blue-shifting points are near C\textsubscript{4} or C\textsubscript{5} which are on the β-ionone ring. This is also despite the fact that the β-ionone ring is ~40º twisted compared to the rest of the retinal backbone in the gas-phase optimized structure, which reduces conjugation of the C\textsubscript{5}=C\textsubscript{6} bond with the other double bonds of the chromophore. Positive point charges near C\textsubscript{9} and C\textsubscript{8} have no effect on the absorption of Me-rPSB11 or are even slightly red shifting. This is consistent with computational studies on bovine Rh which have shown that deprotonation of Glu181, which lies near the middle of the 11-cis rPSB chromophore, results in a blue shift in the rPSB absorption.\textsuperscript{79-81} Indeed, Glu181 is near the red region of Figure 2.3A and introduction of a negative charge (due to deprotonation) in that region is blue-shifting.

The legend associated with the Me-rPSB11 ESTM shows that the most blue-shifting points shift the absorption of Me-rPSB11 by 0.057 eV, while the most red-shifting points have result in a -0.072 eV shift. These magnitudes are associated with introduction of +0.1e point charge and would be larger if the charge magnitude is larger, such as for the mutation of a neutral residue to a charged one (the effect of the charge magnitude on the ESTM is explored in a later section). Out of all the
ESTMs shown in Figures 2.3, 2.4, and 2.5, that of Me-rPSB11 shows the largest spectral range due to the +0.1e point charges (a range of 0.129 eV). This explains why it is possible to tune the absorption of rPSB across the entire visible spectrum (a range of ~1.3 eV) simply through electrostatic interactions, without chemical modification of the chromophore.17

2.4.3 Photoactive Yellow Protein (Xanthopsins)

The ESTM for the first excited state of deprotonated MeS-pCA, the chromophore of PYP in the dark state,82 is shown in Figure 2.3B. The ESTM shows that positive charges near the thioester region would red-shift the absorption, while positive charges near the phenolate oxygen result in a blue shift. The ESTM is perfectly consistent with the computational results of Groenhof et al.,83 which show that excitation of MeS-pCA involves reduction of the electron density at the phenolate oxygen and an increased electron density at C7 and the carbonyl oxygen, resulting in a transition dipole moment pointing towards the thioester end of the molecule. Most other experimental and computational studies in literature focus on the effect of hydrogen-bonding interactions on the absorption of the PYP chromophore.15, 84-86 Specifically, experiments have demonstrated that hydrogen bonding at the phenolate oxygen results in a blue shift in the absorption,15, 85 while hydrogen bonding at the carbonyl of the thioester results in a red shift.86 Our model would only capture the electrostatic contribution introduced by hydrogen-bonding. In that sense, the ESTM is fully consistent with the reports in literature; a positive charge (e.g., corresponding to a partially positive hydrogen from a hydrogen-bond) would blue-shift the absorption at the phenolate oxygen and red-shift it at the carbonyl carbon.

2.4.4 Cyanobacteriochromes (Phytochromes)

The ESTM for the first excited state of PCB is shown in Figure 2.3C. Note that PCB is positively charged due to both nitrogen atoms in rings B and C being protonated. The D ring of PCB is not
pointing inwards like the rest of the pyrrole rings because the C14-C15 single bond is in the \emph{anti} configuration, as in the dark (P\text{r}) state of phytochromes and cyanobacteriochromes.\textsuperscript{64-65} The positive charge of PCB is initially delocalized over the B and C rings of the tetrapyrrole moiety, which are co-planar. Durbeej et al. have shown in a reduced computational bilin tetrapyrrole model that electronic excitation to the first excited state does not lead to a large change in dipole moment,\textsuperscript{87} while Altoe et al. have also shown in a similar reduced model that excitation of the chromophore involves a HOMO to LUMO transition with both orbitals being localized on the B and C rings.\textsuperscript{88} However, both authors indicate a small transfer of the positive charge from rings B and C to A and D. This is consistent with the ESTM of PCB in Figure 2.3; positive charges closer to the B and C rings have a red-shifting effect and those near A and D rings have a blue-shifting effect, although the magnitude of the shifts is small (see the legend), especially compared to the similarly cationic rPSB, because the magnitude of the charge transfer is also small. Therefore, electrostatic protein-chromophore interactions in PCB (and by extension in other in bilins with similar structures) is not an effective spectral tuning strategy. A similar conclusion was recently made by Wiebeler et al.\textsuperscript{89}, who computationally investigated the effect of turning off the protein environment in the PCB-binding cyanobacteriochrome Sir1393g3 and found that it has little effect on the absorption properties of PCB. Instead, they propose that spectral tuning in Sir1393g3 is primarily mechanical, due to steric effects of the protein controlling the conjugation of the D ring with the rest of the chromophore.

\subsection*{2.4.5 CRY, LOV, BLUF, and other flavoproteins}

We employ the oxidized form of LF as a model system for the chromophores of flavoproteins. Oxidized flavins are characterized by two prominent absorption bands centered at \textasciitilde450 nm and \textasciitilde375 nm, respectively, corresponding to the first two bright excited states. We generated an ESTM
for each of these two excitations, shown in Figure 2.3D. Flavin’s first excited state is a notoriously difficult case for spectral tuning. This can be demonstrated by several computational and experimental attempts in literature to predict or synthesize a red-shifted mutant of iLOV, a FMN-binding fluorescent protein, for the purpose of multicolor imaging. However, experimental mutagenesis of iLOV so far has only yielded variants with blue-shifted absorptions. The ESTM in Figure 2.3D top may help explain why this has not been easy; while the legend shows a sizable spectral shift range (a range of 0.055 eV), most of the ESTM features white regions and faded colors compared to other chromophores. The large spectral shifts can only be achieved in a small region of the chromophore surface, meaning that mutations attempting to shift the absorption wavelength of flavin must be pointed at certain regions of the chromophore; near the C4, N5, or N1 atoms. Otherwise, point charges near other regions of the flavin chromophore would not have a substantial effect on the absorption wavelength of flavin. This is also clear from Figure 2.7 which displays the effect of the individual points charges.

One way to red shift the absorption of the first excited state of LF, according to the ESTM, is by introducing a positive charge near the N5. This has been the strategy employed recently by Khrenova et al., who have predicted that introduction of a protonated lysine near N5 would result in a considerably red-shifted variant of iLOV. A second strategy, not yet explored in iLOV, would be the introduction of negative amino acids or dipoles in the vicinity of N1.

The second bright excited state of LF has a substantially different ESTM than the first excited state. While the first excited state would not be influenced by charges near the hydrophilic benzene ring of LF, the second excited state is considerably more sensitive to the electrostatic environment there. This has also been observed in experimental spectra of riboflavin, a LF derivative, in
different solvents; a much higher bathochromic shift is observed for the second bright excited state compared to the first in solvent environments of increasing polarity.\textsuperscript{96}

It is worth noting that because of the different ESTM of the first and second peak in flavin, flavin makes for a very good probe of its environment. For example, charges introduced near the benzene ring would be expected to shift its 370 nm peak and not have a big effect on the first peak, while charges near the C\textsubscript{4} carbonyl group would shift both peaks in the same direction. On the other hand, charges near the N\textsubscript{1} nitrogen would shift the peaks in opposite directions.

![Image](image_url)

**Figure 2.4** Structures and ESTMs for deprotonated (top) and protonated (bottom) forms of p-hydroxybenzylidene-imidazolinone (pHBDI), a model chromophore of the green fluorescent protein. The legends indicate the magnitude of the shifts relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses). The corresponding figure with the individual points before the generation of the surfaces are shown in Figure 2.8 for added clarity.

### 2.4.6 GFP derivatives

We investigate the electrostatic spectral tuning in both the deprotonated (anionic) and protonated (neutral) forms of p-hydroxybenzylideneimidazolinone (pHBDI), a model compound of the fluorophore of wild-type GFP (see **Figure 2.4**). Spectral tuning strategies in GFP have primarily
focused on chemically modified chromophores, π-stacking, and the protonation state of the phenolate group.\textsuperscript{97-101} As a result, there have been relatively few investigations into the electrostatic spectral tuning mechanism in GFP.\textsuperscript{102-103} A recent computational study by Kaila et al. has investigated the effect of the protein on the absorption spectrum of pHBDI compared to the gas phase.\textsuperscript{104} They found that the protein significantly red shifts the absorption of pHBDI relative to the gas phase, and the effect is largely electrostatic. By turning off the charges of individual amino acids in GFP, they pinpoint the amino acids responsible for the red shift in GFP. A comparison with the ESTM in Figure 2.4 reveals that indeed the red-shifting amino acids found by Kaila et al. introduce positive electrostatic interactions in the vicinity of the imidazolinone ring, consistent with our ESTMs. Interestingly, the ESTMs for the protonated and deprotonated forms shown in Figure 2.4 look markedly different. Specifically, the deprotonated form of the chromophore is mostly affected by charges in the vicinity of C$_2^\cdot$, while the protonated form is less sensitive to charges near the imidazolinone ring and is more perturbed by charges near the exocyclic atoms near C$_7$. This is again consistent with the study by Kaila et al.,\textsuperscript{104} who have found that several amino acids have different effects on the spectra of the protonated and deprotonated form of pHBDI.
Figure 2.5 Structures and ESTMs for the side chains of aromatic amino acids: A. phenylalanine; B. tyrosine; C. tryptophan $^1L_a$ state (top), and tryptophan $^1L_b$ state (bottom). The legends indicate the magnitude of the shift relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses). The corresponding figure with the individual points before the generation of the surfaces are shown in Figure 2.9 for added clarity.

Finally, it is worth noting that the magnitude of the spectral shifts for both protonated and deprotonated pHBDI introduced by point charges are not large (compare the legend to those of
Fig. 3). This explains why electrostatic spectral tuning is not the main strategy used in color tuning of GFP derivatives.

### 2.4.7 Aromatic amino acids

Most proteins have a characteristic absorbance at ~280 nm due to their aromatic amino acids.\(^{105-106}\) Tyrosine and tryptophan in particular contribute to that peak, while phenylalanine also absorbs in that energy range but has a much smaller extinction coefficient. Therefore, the spectral properties of tyrosine, tryptophan, and—to a lesser degree—phenylalanine play an important role in the quantitation and characterization of proteins. Moreover, aromatic amino acids can be employed as probes of their native protein conformation, e.g. using difference UV/vis spectroscopy. In such cases, it might be useful to understand how the spectra of aromatic amino acids are influenced by their electrostatic environments. Therefore, we employ gas-phase models of the side chains of tyrosine (p-cresol), tryptophan (3-methylindole), and phenylalanine (toluene) to generate ESTMs, as shown in Figure 2.5.

#### 2.4.8 Phenylalanine

Phenylalanine (modeled here using toluene) only has a weak absorption in the region of ~240-270 nm. Because of lack of electronegative/polar groups in toluene, electrostatic charges have only a limited effect on its absorption spectrum. This is clear from the ESTM of toluene (Figure 2.5A), which is mostly faded and has many white regions. The reddest regions are ortho and para to the methyl group, so there appears to be potential for some spectral tuning by introducing charges that interact with these positions. Note that the ESTM is somewhat asymmetric due to the methyl hydrogens, which break the symmetry across the molecule.
2.4.9 Tyrosine

Tyrosine (modeled here using p-cresol) has a strong absorption in the range of ~250-290 nm. There is a well-known characteristic change in tyrosine absorption upon deprotonation, but we will focus our discussion only on the protonated form since that is the protonation state of tyrosine at neutral pH. The ESTM of p-cresol is shown in Figure 2.5B. Excitation of tyrosine involves a partial charge transfer from the hydroxyl oxygen to the aromatic group, and therefore the most prominent feature of the ESTM is a blue region near the phenol group, while the red regions are at the ortho and meta positions to the phenol. While the absorption spectrum of tyrosine (and its phenol and p-cresol models) in different solvents have been discussed extensively in literature, fewer studies have focused on the effect of specific site-directed interactions on the absorption of tyrosine. However, the effect of hydrogen-bonding at the phenol oxygen have been discussed in literature. For example, Yanari and Bovey report that in cases where the phenol group is the hydrogen donor, this leads to a red shift in the absorption, while if the phenol oxygen is the hydrogen acceptor this leads to a blue-shift in the absorption. This is consistent with the ESTM of Figure 2.5B.

2.4.10 Tryptophan

Tryptophan (modeled here using 3-methylindole) has an absorption spectrum that largely overlaps with that of tyrosine, ranging over ~250-300 nm. Just as in the case of tyrosine, the spectral properties of tryptophan (both absorption and fluorescence) in different proteins and solvents has been studied extensively. For instance, Vivian et al. have discussed fluorescence shifts of tryptophan in different proteins from QM/MM models. In addition, Serrano-Andrés and Roos have modeled the spectral properties of indole in gas-phase and solvent. Indole has two low-lying excited states, $^1L_b$ and $^1L_a$. Both states have transition moments pointing from the benzene
to the pyrrole ring,\textsuperscript{111} although they are not parallel, which also reflected in their slightly different ESTMs in \textbf{Figure 2.5C}. Note that while \(^1\text{L}_b\) is lower in energy than \(^1\text{L}_a\) in indole, in 3-methylindole the two states are nearly degenerate.\textsuperscript{109} TD-B3LYP predicts that \(^1\text{L}_a\) is 0.27 eV lower than \(^1\text{L}_b\), but this is close to the uncertainty associated with TD-B3LYP.\textsuperscript{113} The ESTM legends predict that the \(^1\text{L}_a\) is significantly more sensitive to the environment than \(^1\text{L}_b\), which indeed has been demonstrated in literature both experimentally and theoretically.\textsuperscript{109, 111}

In an experimental study by Andrews and Foster, the effect of site-directed introduction of positive charges in the vicinity of indole was investigated by synthesizing a series of indole derivatives with positively charged centers at different positions.\textsuperscript{109} Their study found that while solvents always shift the \(^1\text{L}_b\) and \(^1\text{L}_a\) in a consistent way, site-specific charges may shift \(^1\text{L}_b\) and \(^1\text{L}_a\) in opposite directions. Remarkably, in agreement with the ESTM in \textbf{Figure 2.5C}, they find that only a charge introduced near the center of the indole moiety can have such an opposite effect on \(^1\text{L}_b\) and \(^1\text{L}_a\), while charges positioned elsewhere shift \(^1\text{L}_b\) and \(^1\text{L}_a\) in tandem.

\subsection*{2.4.11 Effect of charge magnitude, sign, distance, and plurality on the ESTMs}

The ESTMs in \textbf{Figures 2.3-2.5} were all obtained using a charge of +0.1\textit{e} on a surface that is two vdw radii from the atomic centers. In an attempt to test the generality of these ESTMs, and to better quantify them and understand how they are impacted by changes in the charge magnitude, sign, and position, we performed a number of additional tests on the first excited state of LF. One test involved changing the magnitude of the charge to +0.2\textit{e} instead of +0.1\textit{e}. A second test was to change the sign of the charge to -0.1\textit{e} instead of +0.1\textit{e}. A third set of tests involved moving the position of the charges from 2 vdw radii to 1 vdw, 3 vdw, and 4 vdw radii. Finally, we regenerated the ESTM in the presence of four permanent +0.1\textit{e} point charges.
2.4.12 Effect of charge magnitude on the ESTM

Doubling the point charge to $+0.2e$ results in an ESTM that looks identical to the one obtained using $+0.1e$ (compare Figures 2.10 A and B). However, the legends in these figures show that the range of spectral tuning due to the $+0.2e$ charge ($0.108$ eV) is almost double that of the $+0.1e$ charge ($0.055$ eV). Since the QM-MM electrostatic interaction is described using Coulomb’s law, it is expected that doubling the magnitude of a point charge would also double the absorption energy shift caused by that charge. Indeed, Figure 2.6A shows that this is the case; Figure 2.6A plots the spectral shift introduced by each of the $+0.2e$ point charges against the spectral shift introduced by the corresponding $+0.1e$ point charge at the same coordinate. The resulting plot is linear with a slope of 1.9928. The reason that the slope is not exactly 2 is because the QM wave function is recomputed in the presence of the point charge, which accounts for polarization of the QM wave function by the point charges as well as for small changes in the excited-state configuration. It would be expected that as the magnitude of the point charge increases, there will be further deviations from an ideal Coulomb law due to polarization effects.

**Figure 2.6** Plots showing the effect of changes in the charge magnitude, sign, position, and plurality on the spectral shift induced in the LF model chromophore. Each point in the plots corresponds to a point on the ESTM surface. Specifically, we investigate the effect of: A. changing
the charge magnitudes from +0.1e to +0.2e; B. changing the charge sign from +0.1e to -0.1e; C. Changing the charge positions from 2 vdW radii to 1 vdW radius (blue), 3 vdW radii (green), and 4 vdW radii (yellow). The inset shows a plot of the slopes from the plot against “hypothetical slopes” that are anticipated from a simple Coulomb model between two points varied by the same distances; D. recomputing the effect of each point charge in the presence of four other permanent point charges. In all plots, linear trendlines are shown in black and the associated regression equation and $R^2$ values are shown in the top left of each graph.

2.4.13 Effect of charge sign on the ESTM

Using a negative point charge (-0.1e) instead of positive one results in an inverted ESTM where blue regions and red regions are interconverted (compare Figures 2.10 A and C). That is the only qualitative change in the ESTM, that otherwise appears identical. As shown in Figure 2.6B, the plot associated with the sign change is again linear with a slope of -1.007, meaning that changing the sign of the charge simply leads to a spectral shift in the opposite direction with almost the same magnitude.

2.4.14 Effect of charge distance on the ESTM

The effect of placing the point charges at 1 vdW, 2 vdW, 3 vdW, and 4 vdW distance is tested. As in other tests, the ESTMs have qualitatively the same color distribution regardless of the distance of the point charges, and only the magnitude of the shift changes (see Figure 2.11). Note that this test is different than others because the points have been regenerated on the surfaces of different radii and therefore have different positions.

Predicting the magnitude of the shift due to changes in position is not as straightforward as in other cases. For a two-body problem, Coulomb’s law predicts that the effect should scale inversely with distance. This would suggest that moving the point charges to 1 vdW radius (half the distance)
would double the potency of that charge, while moving the charge to 3 and 4 vdW radii would scale their spectral shifting effect by 2/3 and 1/2, respectively. However, the point charges are not interacting with a single QM atom, but instead an entire molecule, and therefore the distances do not scale linearly with vdW radii. From Figure 2.6C, in the case of LF, we find that the spectral shift magnitudes change by a factor of 2.44, 0.52, and 0.32 for 1 vdW, 3 vdW, and 4 vdW radii (compared to the simple-minded factors of 2, 0.67, and 0.5 expected from Coulomb’s law considering interaction with a single point scaled to 1/2, 3/2, and 2 times the distance). Using Coulomb’s 1/r law can therefore provide a very crude estimate of the spectral shift due to more distant point charges, but to more accurately quantify the effect of distance on the spectral shift for any given chromophore would require generating the ESTMs at different distances and plotting the respective slopes for each distance against the “hypothetical” ones expected from Coulomb’s law, as done in the inset of Figure 2.6C for LF.

2.4.15 Effect of multiple charges

For ESTMs to prove useful for understanding the effect of a mutation in a protein, it is important to establish that the same ESTM can be used in the presence of other point charges. Therefore, we recomputed the ESTM in the presence of four permanent +0.1e point charges (two strongly red-shifting, one having a negligible effect on the spectral properties of LF, and one blue-shifting as shown in Figure 2.12 C). The total effect of all four charges is therefore red-shifting the absorption by -0.029 eV). Comparison of Figures 2.6A and B demonstrates that even in the presence of other point charges, the ESTM remains qualitatively almost identical. We quantify this effect in Figure 2.6D, where we find again a linear relation between the single-point charge model and the 0.1e+0.4e scattered point charge model, with a slope of 1.0336, which is close to the expected 1 of a perfectly additive model, and a y-intercept of -0.0288 consistent with the red-shift caused by
the 4 existing points. This demonstrates that the effect of the multiple point charges is additive, and the model is therefore also useful for understanding the electrostatic spectral tuning in proteins that do not strongly perturb the electronic structure or geometry of the chromophore.

Figures 2.10, 2.11, and 2.12 demonstrate that the ESTMs generated in this study are qualitatively useful to understand the effect of an environment on the absorption spectrum, regardless of charge magnitude, sign, distance, and the presence of additional existing point charges (as in the case of a protein). In addition, Figure 2.6 shows that the ESTMs are also semi-quantitative, and the effect of changing the point charge magnitude, sign, or introducing multiple point charges can be predicted. This is true as long as the QM wave function is not strongly perturbed by the point charges. The only factor not easily quantifiable is the effect of moving the point charge to different distances; simply scaling the spectral shift by the change in vdW distance can be used for a crude estimate, but more a more quantitative approach would require the calculation of ESTMs at multiple vdW radii.

ESTMs are limited to electrostatic interactions and do not account for other interactions, such as π-stacking, that may induce a change in the absorption spectrum of the chromophore. Moreover, ESTMs in this study employ a gas-phase optimized structure and therefore do not include the effect of steric tuning by the protein. While this is not a concern in rigid chromophores, in more flexible chromophores the ESTMs may be recomputed at different (e.g., QM/MM optimized) geometries to study the effect of geometry on the ESTM map.
Figure 2.7 Structures and Electrostatic spectral tuning maps (ESTMs) for model chromophores of biological photoreceptors. The systems shown are: A. Me-rPSB11; B. MeSPC; C. PCB; and D. LF. For LF, the top ESTM is for the first bright excited state (corresponding to the experimental peak at ~450 nm), and bottom ESTM is for the second bright excited state (corresponding to the
experimental peak at ~370 nm). The legends indicate the magnitude of the shifts relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses).

**Figure 2.8** Structures and ESTMs for deprotonated (top) and protonated (bottom) forms of p-hydroxybenzylidene-imidazolinone (pHBDI), a model chromophore of the green fluorescent protein. The legends indicate the magnitude of the shifts relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses).
Figure 2.9 Structures and ESTMs for the side chains of aromatic amino acids: A. phenylalanine; B. tyrosine; C. tryptophan $^1L_a$ state (top), and tryptophan $^1L_b$ state (bottom). The legends indicate the magnitude of the shift relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses).
Figure 2.10 ESTMs for lumiflavin generated using A. +0.1e point charges; B. +0.2e point charges; and C. -0.1e point charges. The legends indicate the magnitude of the shift relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses).
Figure 2.11 ESTMs for lumiflavin generated using +0.1e point charges at A. 1 vdW distance; B. 2 vdW distance; C. 3 vdW distance; and D. 4 vdW distance. The legends indicate the magnitude of the shift relative to the vertical excitation energy of the gas-phase reference in eV (and in nm in parentheses).
Figure 2.12 ESTMs for lumiflavin generated using A. the normal protocol with +0.1e point charges; and B. the same protocol but in the presence of four permanent +0.1e point charges. The legends indicate the magnitude of the shift relative to the vertical excitation energy of the gas-
phase reference in eV (and in nm in parentheses). Note that for panel B the reference has been shifted due to the already existing point charges. Panel C. shows the positions and shift magnitudes of the four chosen permanent point charges.

2.5 Conclusions

We describe an approach to generate “electrostatic spectral tuning maps,” or ESTMs, that visually represent how an electrostatic environment will affect the absorption of a chromophore. We generate maps for model chromophores of several biological photoreceptors, a model of the fluorophore of the green fluorescent protein, and model side chains of aromatic amino acids. The ESTMs discussed in this work are intuitive, consistent with existing spectral tuning computational and experimental studies, and provide new insight into the spectral tuning of biological chromophores, fluorophores, and aromatic amino acids.

These ESTMs are relatively easy to construct. They involve running many single-point calculations in the presence of point charges, a task that is embarrassingly parallel and can be performed on most open-source as well as commercially available QM software packages using any affordable excited-state QM method. The ease and speed in which ESTMs can be generated makes them useful precursors to full QM/MM studies that are more time-consuming and involved. We used these maps to study the effect of an electrostatic environment on the excitation energy of biological chromophores. However, in principle, the same kind of maps can be employed to study the effect of an electrostatic environment on other properties such as oscillator strengths, absolute energies (which can be related to stability, useful for understanding the relative stability of two different isomers or transition states), ionization energies, and electron affinities. ESTMs for the last two properties can provide insight into how redox properties of a system are tuned by a protein environment.
Finally, ESTMs are complimentary to maps that describe the electrostatic potential of a protein on a vdW surface, such as those generated using the Adaptive Poisson-Boltzmann Solver (APBS) / PDB2PQR software.\textsuperscript{14-16} Such maps have been generated for microbial rhodopsin, for instance, where the electrostatic potential of the protein is mapped onto the vdW surface of the rPSB chromophore.\textsuperscript{18} It may be possible to employ both kinds of maps to predict the effect of a given protein on a chromophore’s absorption spectrum.
2.6 References


81. Tomasello, G.; Olaso-Gonzalez, G.; Altoe, P.; Stenta, M.; Serrano-Andres, L.; Merchan, M.; Orlandi, G.; Bottoni, A.; Garavelli, M., Electrostatic Control of the Photoisomerization


3 ELECTRONIC SPECTRA OF FLAVIN IN DIFFERENT REDOX AND PROTONATION STATES: A COMPUTATIONAL PERSPECTIVE ON THE EFFECT OF THE ELECTROSTATIC ENVIRONMENT

Mohammad Pabel Kabir, Yoelvis Orozco-Gonzalez, and Samer Gozem


Copyright © the Owner Societies 2019

My contribution to this chapter was data curation, formal analysis, investigation, methodology, validation, visualization, writing-original draft.

3.1 Abstract

Flavins are versatile molecules due to their ability to exist in multiple redox and protonation states. At physiological conditions, they are usually encountered either as oxidized quinones, neutral semiquinones, anionic semiquinones, neutral hydroquinones, or anionic hydroquinones. We compute the electronic near-UV/vis spectra for flavin in each of these five states. Specifically, we compute vertical, adiabatic, and vibronic excitations for all excited states that have wavelengths longer than 300 nm. We employ the calculations to assign the peaks in the corresponding experimental UV/vis spectra from literature. We also compare the effect of polar and non-polar solvents on the spectra using a polarizable continuum model. Finally, we construct "electrostatic
spectral tuning maps” for prominent peaks in each of the five states. These maps qualitatively
describe how the flavin electronic spectra will be shifted by an anisotropic electrostatic
environment such as a protein. Understanding how flavin’s UV/vis absorption spectrum is
modulated by its environment can aid in experiments employing flavin as a probe of internal
electrostatics of a protein and in engineering new color variants of flavoproteins.

3.2 Key words
Flavin, Flavoprotein, UV/vis spectrum, Electronic Excitation, Vibrational excitation, Franck-
Condon factors

3.3 Abbreviations
FMN, Flavin mononucleotide; FAD, Flavin adenine dinucleotide; LOV, Light-oxygen-voltage;
BLUF, Blue-light sensing using FAD domain; ESTM, Electrostatic Spectral Tuning Map; DFT,
Density functional theory; QM/MM, Quantum mechanical/Molecular mechanical, TD-DFT, Time
dependent density functional theory; FC, Franck-Condon.

3.4 Introduction
Flavin-dependent proteins are found in all kingdoms of life, where they are responsible for many
biological processes and catalytic roles.\textsuperscript{1-4} Most flavoproteins are oxidoreductases, although they
may also play non-redox roles such as in transferases, lyases, isomerases, and ligases.\textsuperscript{3} Flavins
have also been implicated as the chromophores in a number of photoreceptors and light-responsive
enzymes including cryptochromes,\textsuperscript{5} light-oxygen-voltage (LOV) sensing domains,\textsuperscript{6} blue-light
sensing using FAD (BLUF) domains,\textsuperscript{7} and DNA photolyases.\textsuperscript{8, 9} Flavoproteins typically bind
flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD), both of which are derivatives
of riboflavin (vitamin B\textsubscript{2}).
The ubiquity of flavin cofactors in living organisms is in large part due to their versatility; they are multi-redox, photoactive, and may exist in different protonation states (in some flavin literature referred to as ionization states). With the exception of a few proteins that stabilize the superoxidized N(5)-oxide form of flavin\textsuperscript{10}, flavin is often encountered as a quinone, semiquinone, or hydroquinone (Fl, FlH\textsuperscript{-}, or FlH\textsubscript{2}, respectively. see Scheme 1). Each of these three redox states may also be deprotonated; the pK\textsubscript{a} values of Fl and FlH\textsubscript{2} are around 10.3 and 6.7 in aqueous solution, respectively,\textsuperscript{11-13} while FlH\textsuperscript{-} has a pK\textsubscript{a} of around 8.3 in aqueous solution\textsuperscript{14,15} and 6.3-7.3 in different proteins.\textsuperscript{16,17} Due to the relatively high pK\textsubscript{a} of Fl, we did not consider the anionic quinone in this study. Anionic semiquinone (FlH\textsuperscript{-}) and anionic hydroquinone (FlH\textsuperscript{2-}) may, however, be encountered at a physiological pH. Each of those five states, shown in Scheme 1, has a characteristic UV/vis absorption spectrum, a fact that is often employed to identify them spectroscopically.
Figure 3.1 Structure of flavins in different redox and protonation states. The atoms are labeled for the oxidized structure.

Flavin’s ability to exist in multiple redox/protonation states means that there is a potential for probing the microenvironment of a protein using several different states through simple redox/pH changes. However, while a few studies have discussed how the UV/vis absorption of Fl is affected by a protein or solvent environment,\textsuperscript{18,19} this is not the case for FlH\textsuperscript{−}, FlH\textsubscript{2}, Fl\textsuperscript{−}, and FlH\textsuperscript{−}. Knowledge of the excited states featured in the UV/vis absorption spectrum of flavin in each redox/protonation state and how they are affected by nearby charges/dipoles would allow the correlation of these spectral shifts to the presence of nearby charged/polar amino acids or hydrophobic regions in a protein. Moreover, a better understanding of how the absorption
wavelengths of flavins are modulated by a protein, particularly for Fl, can aid in the development of new flavoprotein color derivatives for applications such as bioimaging, biosensing, and optogenetics.

In this work, we simulate the near-UV/vis spectra of flavin in the five different redox/protonation states shown in Scheme 1 (Fl, FlH?, Fl?, FlH2, and FlH?). Specifically, we compute vertical excitations, adiabatic excitations, and vibronic excitations for all excited states with an absorption wavelength in the near-UV/vis range. We then model solvent effects on the spectral shifts and relative intensities in different solvents.

Recently, we showed that the effect of the surrounding electrostatic environment on the absorption wavelengths of a chromophore can be understood in an intuitive way using Electrostatic Spectral Tuning Maps (ESTMs). The maps are built by moving a point charge on the van der Waals surface of the chromophore and calculating the change in the chromophore’s excitation energy relative to the gas phase, thereby mapping on the surface how the absorption spectrum of the chromophore is affected by the point charge at each position. In this work, we also produce ESTMs for prominent peaks in the UV/vis absorption spectra of Fl, FlH?, Fl?, FlH2, and FlH? and discuss how such ESTMs can be used to understand the effect of a protein or solvent on the spectrum of each of these five states. We then discuss the results of our calculations in the context of experimental spectra and computational studies in literature.
3.5 Theoretical Background

Experimental UV/vis absorption spectra can be directly compared to excitation energies from electronic structure calculations in different ways. Typically, vertical excitation energies (i.e., the energy difference between the ground and excited state at the equilibrium ground state geometry of the system) are comparable to the experimentally observed maximum absorption wavelengths ($\lambda_{\text{max}}$) of the corresponding peaks in the UV/vis absorption spectrum. This similarity between computed vertical excitation energies and the experimental $\lambda_{\text{max}}$ is justified by the Franck-Condon (FC) principle. However, it has been shown in several studies, using both time-dependent density function theory (TD-DFT) and wave function methods, that computed vertical excitation energies in the case of the first excited state of Fl are considerably blue-shifted with respect to the experimental $\lambda_{\text{max}}$. 27-32 Theoretical models in literature that most accurately reproduce the experimental $\lambda_{\text{max}}$ for Fl account for adiabatic excitations and vibrational overlaps between the ground and excited state wave functions, i.e. FC factors, 28-30 or account in some other way for sampling of the vibrations of the chromophore. 33 Adiabatic excitation energies are energy differences between the excited and ground states at their respective equilibrium geometries, while vibronic excitations also account for vibrational energy levels (see Figure 3.2).

In this work, we report vertical and adiabatic energies as well as FC factors for all near-UV/vis bright excited states of Fl, FlH+, Fl-, FlH2, and FlH-. We use these properties to simulate the near-UV/vis absorption spectra of each species. Note that while the FC factors give an indication of the vibronic broadening of each peak, additional broadening effects are more difficult to take into account from first principles, so we determine the broadening factors empirically by comparison with the corresponding experimental UV/vis spectra. See the Computational Approach section and Supplementary Information (SI) document for more details.
Figure 3.2 Schematic figure showing the difference between vertical, adiabatic, and vibronic excitations. The vertical excitation (black arrow) is the electronic energy difference between the ground and excited state obtained at the ground state geometry. The adiabatic energy accounting for zero-point vibrational energy correction (green arrow) is the energy difference between the lowest vibrational states of the ground and excited electronic state (v=0 to v'=0). Vibronic excitation energies (green and orange arrows) are energy differences between all ground and excited vibrational states. We assume that the ground state is always in the lowest vibrational state (i.e., we compute 0 K spectra where higher ground vibrational states are not considered). The intensities of vibronic transitions are determined by the overlaps of the ground state wave function (red) and excited state wave functions (blue).

3.6 Computational Approach

The isoalloxazine ring acts as the chromophore responsible for flavin’s spectral properties at near-UV/vis wavelengths,\(^\text{32}\) so we employed a reduced model, lumiflavin, for all calculations in this work (see Scheme 1).
All ground state energies, geometries, and frequencies were computed using density functional theory (DFT) with the B3LYP functional\textsuperscript{34,36} and cc-pVTZ basis set.\textsuperscript{37} Similarly, all excited state energies, geometry optimizations, and frequency calculations were performed using TD-DFT\textsuperscript{38} with the B3LYP functional and cc-pVTZ basis set. These calculations were used to determine vertical and adiabatic excitation energies. All adiabatic energies reported in this work include the zero-point vibrational energy (ZPVE) correction.

FC factors were computed between the ground and excited states from their respective frequencies and normal modes. FC factors were computed analytically within the double-harmonic approximation at 0 K including Duschinsky rotations\textsuperscript{39} of the normal modes. The calculation of FC overlaps was sometimes complicated by small differences in conformation in the exited state compared to the ground state. Typically, this includes methyl-group rotations, although in the case of \textit{FlH}_2 and \textit{FlH}^- it also includes bending or pyramidalization of the isoalloxazine backbone.\textsuperscript{40-42} Such changes all occur along low-frequency modes but result in significantly lower FC overlaps. In such cases, several approximations were made during the calculation of FC overlaps. Specifically, for \textit{FlH}_2 and \textit{FlH}^-, this includes computing FC factors in a constrained planar model or reduced model where methyl groups are replaced by hydrogens. These approximations are discussed in detail in the SI. The FC factors are used in this work to determine the vibronic structure of each electronic transition and may not be quantitative due to these reasons. However, all adiabatic and zero-point vibrational energies were computed using full lumiflavin models that were properly optimized.

To compare our computed energies with experimental UV/vis spectra, vibronic excitation bands were broadened using gaussian functions. The full width at half-maximum (FWHM) were determined empirically. Details of how the spectra were simulated are provided in the SI. Note
that in the case of \textbf{FIH$_2$} and \textbf{FIH$^-$}, in part due to problems with FC factor calculation, it was necessary to use a large FWHM and renormalize the height of the peaks to get a good agreement with the experimental spectra.

All calculations were performed using a polarizable continuum model (PCM) to account for solvation,\textsuperscript{43,44} unless it is specified that the model is in the gas phase. Specifically, calculations on \textbf{Fl}, \textbf{FIH$_2$}, and \textbf{FIH$^-$} were performed in water because experimental data are available in aqueous solutions. However, \textbf{FIH$^-$} and \textbf{Fl$^-$} are unstable in aqueous solutions. Therefore, all calculations for the two semiquinone forms were performed in methyl propanoate, which has a dielectric constant of 6.078 and is an approximate representation of a typical dielectric environment of a protein pocket.\textsuperscript{45}

The effect of changing the solvent on the UV/vis spectra was also investigated. Since recomputing the adiabatic and FC factors in each solvent would be too time-consuming, only ground state geometries and vertical energies were recomputed in different solvents (water, cyclohexane, and gas phase) to obtain the solvent shift and changes in intensities. These wavelength and intensity changes were then applied to shift each absorption peak as a whole in the broadened spectra.

ESTMs were computed for the two excitations that have the strongest near-UV/vis absorption, with the exception of \textbf{FIH$^-$} where only one ESTM corresponding to the lowest excited state was computed. The ESTMs were computed using the steps outlined in ref. 26. Briefly, we located points on a surface 2 van der Waals radii from each atom of the molecule at a density of one point per Å$^2$ using the Python script \textit{pyvdwsurface}.\textsuperscript{46} We placed a +0.1e point charge at each position and, for each individual point, ran a TD-DFT B3LYP/cc-pVTZ energy calculation to compute the excitation energy of the chromophore in the presence of that point charge. The points were then
represented using a color reflecting the spectral shift they induce in the chromophore, and a color-coded ESTM was generated using these points.

The purpose of the ESTMs is to visually represent how a chromophore’s excitation energies are shifted by nearby point charges. Red regions on the ESTM surface mean that a positive charge in that region would induce a red-shift in the corresponding peak in the absorption spectrum, while blue regions represent points that induce a blue-shift. We have shown (ref. 26) that a negative charge would have exactly the opposite shifting effect, and that the magnitude of the shifts are directly proportional to the magnitude of the charge as long as the charge is not large enough to significantly perturb the electronic structure of the chromophore. The choice of placing the charges at 2 van der Waals radii is because this is the closest distance that an external atom can be from the chromophore before repulsive forces dominate, although we have discussed the effect of increasing the distance as well.26 The legends associated with these ESTMs also give an idea about the magnitude of the spectral shift. Therefore, the ESTM can be used qualitatively or semi-quantitatively to understand the effect of a charge of any sign, magnitude, and distance on the excitation energies of the chromophore.

DFT and TD-DFT geometry optimizations and frequency calculations were all performed using Gaussian.47 The FC factors were computed with ezSpectrum.48 ESTMs were computed from single-point calculations performed in Gaussian and plotted using the Mayavi scientific data visualizer.49
3.7 Results and Discussion

<table>
<thead>
<tr>
<th>State</th>
<th>TD State Order</th>
<th>Vertical Excitation Energy (eV)</th>
<th>Adiabatic Excitation Energy (eV)</th>
<th>Vertical Oscillator Strength (Intensity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fl</td>
<td>1</td>
<td>2.933</td>
<td>2.551</td>
<td>0.234</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.591</td>
<td>3.176</td>
<td>0.250</td>
</tr>
<tr>
<td>FlH−</td>
<td>1</td>
<td>2.132</td>
<td>1.808</td>
<td>0.127</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.792</td>
<td>2.402</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.875</td>
<td>2.489</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3.441</td>
<td>3.170</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>3.495</td>
<td>3.199</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3.860</td>
<td>3.715</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>4.204</td>
<td>3.992</td>
<td>0.091</td>
</tr>
<tr>
<td>Fl−</td>
<td>3</td>
<td>2.834</td>
<td>2.555</td>
<td>0.142</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3.304</td>
<td>3.034</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>3.467</td>
<td>3.155</td>
<td>0.297</td>
</tr>
<tr>
<td>FlH2</td>
<td>1</td>
<td>3.014</td>
<td>2.308</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.885</td>
<td>3.256</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.173</td>
<td>4.018</td>
<td>0.178</td>
</tr>
<tr>
<td>FlH−</td>
<td>1</td>
<td>3.119</td>
<td>2.456</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.590</td>
<td>3.306</td>
<td>0.133</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.342</td>
<td>3.888</td>
<td>0.237</td>
</tr>
</tbody>
</table>

Table 3.1 Computed vertical excitation energies and oscillator strengths (intensities) as well as adiabatic excitation energies for bright excited states in **Fl**, **FlH−**, **Fl−**, **FlH2**, and **FlH−**. The number of the excited state in the TD-B3LYP/cc-pVTZ calculation is labeled for easy referencing in the text.

The computed vertical and adiabatic energies for all the excited states are tabulated in Table 3.1. The adiabatic energies were computed for bright excited states with absorption wavelengths above 300 nm. Other states have zero or very low intensity, and their contributions to the spectra are ignored. **Fl** and **FlH−** both have first excited states with oscillator strength above 0.1. This may
help explain the strong fluorescence in Fl and FlH\textsuperscript{-}. Recently, it has been shown that FlH\textsuperscript{-} has relatively strong fluorescence in \textit{Pseudomonas aeruginosa} nitronate monooxygenase (PaNMO) that is around 10\% of the fluorescence intensity of free FMN in solution\textsuperscript{17} Fl\textsuperscript{-} has a first excited state with negligible intensity, which may help explain why only weak fluorescence has been reported in insect Type 1 cryptochrome\textsuperscript{50}. FlH\textsubscript{2} and FlH\textsuperscript{-} also have a first excited state with a small but non-zero intensity; Indeed, FlH\textsubscript{2}, and FlH\textsuperscript{-} are not fluorescent in solution and only display a relatively weak fluorescence in glass\textsuperscript{51}.

3.7.1 The Fl UV/vis spectrum and ESTMs.

The computed UV/vis spectra in different solvents, the experimental UV/vis spectrum in aqueous solution\textsuperscript{52} and ESTMs for Fl are shown in Figure 3.3. The absorption spectrum of Fl is characterized by two strongly absorbing states in the near-UV/vis region. The first peak, with $\lambda_{\text{max}}$\textsuperscript{}$\approx$450 nm, corresponds to state 1 in Table 3.1 while the second characteristic peak at $\lambda_{\text{max}}$\textsuperscript{}$\approx$370 nm corresponds to state 4 in Table 3.1. Notice that when Fl is placed in non-polar environments, the first excited state often has a more defined vibronic structure showing three vibronic bands instead of one broad band. Indeed, the computed vibronic excitations are also clustered such as to give three peaks if a smaller FWHM is used for broadening (see inset, corresponding to the first electronic excited state simulated with smaller FWHM).
Figure 3.3 Computed (solid lines) and experimental (dashed lines) UV/vis spectra of the oxidized flavin (Fl). The vibronic excitation energies computed in a water PCM model are shown as impulse lines in blue. The corresponding broadened peaks are also shown in blue. The solvent shift (computed from differences in vertical excitation energies) and changes in relative intensity associated with different solvents are also shown for gas (light red) and cyclohexane (bright red). The experimental UV/vis spectrum corresponds to aqueous FMN. ESTMs corresponding to the two lowest bright excited states are shown to the right. The legend indicates the magnitude of the shifts relative to the gas-phase reference excitation energy in eV (and in nm in parentheses). The inset shows the first excited state peak simulate with a FWHM of 0.1 eV to show the vibronic band structure in case of smaller broadening.

We computed the effect of changing the solvent to cyclohexane and to the gas phase. With decreasing solvent polarity, both peaks with $\lambda_{\text{max}}$ at 370 nm and 450 nm become blue-shifted. When we change the solvent from water to gas phase, the peak at 450 nm blue-shifts by around 15 nm while the peak at 370 nm blue-shifts around 30 nm, so the latter peak is more sensitive to the solvent polarity than the first. Moreover, the relative intensity of the two peaks changes
systematically with decreasing solvent polarity. Note that the spectra are normalized to the more intense peak, corresponding to state 1.

The ESTMs for Fl have been discussed recently in Ref. 26. We included the ESTMs here recomputed with the cc-pVTZ basis set for completeness and to allow a direct comparison with the ESTMs of other flavin states. While the solvent effects explain how a homogeneous dielectric environment affects the UV/vis absorption spectrum of Fl, the ESTMs describe how a point charge at any given position will affect the absorption energy of the chromophore. For more details on how to interpret ESTMs, see the last paragraph of the Computational Approach section. An ESTM is shown for each of the two electronic peaks in the UV/vis spectrum. The legends on the ESTMs indicate the magnitude of the spectral shift induced by a +0.1e point charge on the van der Waals surface. One notable difference between the two ESTMs is that the ESTM for state 1 is less intense in color than ESTM for state 4, indicating that the peak at 450 nm is less sensitive to its environment than the peak at 370 nm.

Notice that it is not necessarily straightforward to connect the solvent effect to the ESTM; the solvent surrounds the molecule completely and has different dipoles at different parts of the molecule depending on the solvent orientation, complicating such an interpretation. The ESTM is more useful for predicting how a change in the environment at a particular position (such as a protein mutation or conformational change) may influence the UV/vis spectrum of the chromophore. However, some inferences may still be made about the solvent effect from the ESTM; for instance, the region surrounding the hydrophilic part of the molecule is red for state 4, consistent with the fact that a polar solvent with positive poles pointed at the carbonyl oxygens and at N₃ will red-shift that state.
The ESTMs for Fl have been discussed recently in ref. 26. The ESTMs indicate that charges introduced near the benzene ring would shift the 370 nm peak without having an effect on the 450 nm peak, charges near the C_4 carbonyl group would shift both peaks in the same direction, while charges near the N_1 nitrogen would shift the peaks in opposite directions. This can be a useful guide to make inferences about the electrostatic microenvironment of a protein from the relative shifts induced in the two peaks.

### 3.7.2 Comparison with experimental data for Fl.

There is a strong agreement between the computed and experimental $\lambda_{\text{max}}$ values for Fl, as well as in the relative intensities of the two peaks, which accounts for the excellent agreement between the experimental and simulated UV/vis spectra when an empirical broadening is used. The predicted solvent effect is also in very good agreement with literature,\textsuperscript{53, 54} which also show that both visible peaks in Fl blue-shift while the peak at ~370 nm decreases in intensity in solvents of decreasing polarity. Finally, to support the conclusion from the ESTMs that the peak at ~370 nm is more sensitive to the environment than the peak at ~450 nm, we analyzed the experimental UV/vis spectra of 28 different natural variants, mutants, and engineered derivatives of LOV domains (see SI).\textsuperscript{6, 30, 55-72} Indeed, we find that for all 28 proteins, the $\lambda_{\text{max}}$ of state 1 only varies over a range of 11 nm (from 439 to 450 nm),\textsuperscript{59} while the $\lambda_{\text{max}}$ of state 4 varies over a range of 28 nm (from 352 to 380 nm).\textsuperscript{63, 67}

### 3.7.3 Comparison with other computational studies for Fl.

There are numerous computational studies on the UV/vis spectroscopy of Fl in literature. An incomplete list includes refs. 27-29, 31-33, 73-79. A number of computational studies have also modeled UV/vis spectra of flavin derivatives or analogues.\textsuperscript{29, 79-83} While the majority of these studies reported vertical excitations, some have reported adiabatic and FC factors for the first
excited state of Fl as well.\textsuperscript{28-30} We find, in agreement with these reports,\textsuperscript{28-30} that the TD-DFT vertical excitations for states 1 and 4 are ca. 25 nm blue-shifted with respect to the experimental λ\textsubscript{max}. Instead, a good match between the computed and experimental λ\textsubscript{max} is achieved using adiabatic excitation and taking into account the FC factors. Note that this agreement could be due to an error cancellation between the vertical excitation energy and geometric reorganization energy, as suggested by a recent benchmark study on vertical and adiabatic excitation energies in a large set of compounds.\textsuperscript{84} Therefore, vertical excitation energies computed with other methods may yield better agreement with the experimental λ\textsubscript{max}, although most reports in literature, including those that use wave-function methods,\textsuperscript{31,32,85} also show vertical excitation energies that are blue-shifted with respect to the experimental λ\textsubscript{max}.

A number of computational studies have investigated the effect of solvent polarity on the λ\textsubscript{max} of Fl for the two visible peaks and have shown that decreasing solvent polarity results in both peaks blue-shifting and a reduction in the relative oscillator strength of the peak at ~370 nm, in agreement with our results.\textsuperscript{27,31,86}

More recently, several studies have employed hybrid quantum mechanical/molecular mechanical (QM/MM) models to study the spectral properties of Fl in a protein environment. This includes QM/MM studies on LOV proteins and their derivatives,\textsuperscript{30,87-93} BLUF,\textsuperscript{94-96} and DNA photolyases.\textsuperscript{97,98} However, most of these studies do not discuss how the protein shifts the absorption wavelength of Fl.

With the goal to red-shift the fluorescence the Fl-binding fluorescent protein iLOV, Nemukhin and co-workers\textsuperscript{89-91} have discussed placing positively charged residues near N\textsubscript{5} of Fl. Their predictions are consistent with our ESTMs, which also indicate that positive charges near N\textsubscript{5} will
red-shift the absorption of $\text{Fl}$. However, no studies yet have attempted to red-shift the absorption or fluorescence of $\text{Fl}$ by introducing a negative charge near $N_1$, which might be expected to achieve the same effect according to the ESTM for the same state.

3.7.4 The $\text{FlH}^-$ UV/vis spectrum and ESTM.

The computed UV/vis spectra in different solvents, the experimental UV/vis spectrum in PaNMO,$^{17}$ and the ESTM for $\text{FlH}^-$ are shown in Figure 3.4. The experimental UV/vis absorption spectrum of $\text{FlH}^-$ features several peaks at wavelengths above 300 nm. This state is easy to identify spectroscopically because it has the most red-shifted absorption peak among all the flavin redox/protonation states discussed here; state 1 in Table 3.1 corresponds to the absorption centered at around 600 nm. The peak at 450-500 nm is predominantly due to state 3 but also has a small contribution from state 2. The multiple peaks from 300-400 nm have contributions from state 5, 6, 10, and 11. The shoulder at around 400 nm is predominantly from state 5 since the contribution of state 6 is small, while the tall peak at 300-350 nm mainly corresponds to state 10 but also has a contribution from state 11.
Figure 3.4 Computed (solid lines) and experimental (dashed lines) UV/vis spectra of the neutral semiquinone form of flavin (FlH·). The vibronic excitation energies computed in a methyl propanoate PCM model are shown as impulse lines in magenta. The corresponding broadened peaks are also shown in magenta. The solvent shift (computed from differences in vertical excitation energies) and changes in intensity associated with different solvents are also shown for gas (light red), cyclohexane (bright red), and water (blue). The experimental UV/vis spectrum corresponds to FADH· in PaNMO. The ESTM corresponding to the lowest bright excited state is shown to the right. The legend indicates the magnitude of the shifts relative to the gas-phase reference excitation energy in eV (and in nm in parentheses).

The choice of methyl propanoate is because its dielectric constant (6.078) is a crude approximation for the dielectric environment of a protein cavity.

Figure 3.4 shows the effect of recomputing the spectrum in different solvents. Although FlH· is unstable in all of these solvents, the computed solvent effects provide the effect of polarity on the FlH· spectrum in general. Unlike the case of Fl, the computed UV/vis bands of FlH· shift and
change intensity, sometimes drastically, in different solvents. For example, states 2 and 3 have a very weak absorbance in the gas phase but become a well-defined peak in polar environments. State 1 redshifts by around 13 nm while state 5 redshifts by around 18 nm when the environment changes from highly polar (water) to non-polar (gas) environment and the peak corresponding to state 5 moves away from the peak at 300-350 nm, causing changes in the shape of the peaks in the near-UV range.

Due to the number of states appearing in the near-UV/vis spectrum of \( \text{FlH}^- \) and the fact that several of them are close in energy, it is difficult to compute ESTMs for all the states. Therefore, we have computed the ESTM only for the first excited state, which is the most distinguishing feature of the \( \text{FlH}^- \) UV/vis spectrum. Interestingly, although the ESTMs for the first excited state of \( \text{Fl} \) and \( \text{FlH}^- \) are similar, polar solvents red-shift the first excited state of \( \text{Fl} \) and blue-shift the first excited state of \( \text{FlH}^- \). The reason for this difference could be that that protonation of N5 causes polar solvents to orient their poles differently in the two systems (specifically, they may orient their positive pole towards the deprotonated N5 in \( \text{Fl} \) and instead orient their negative pole towards the same N5 when it is protonated in \( \text{FlH}^- \)).

3.7.5 Comparison with experimental data for \( \text{FlH}^- \).

The agreement between the experimental and computed \( \lambda_{\text{max}} \), particularly for the peak at \( \sim 600 \) nm, is not as good as in the case of \( \text{Fl} \). In fact, in this case, the vertical excitation energy is in better agreement with the experimental \( \lambda_{\text{max}} \), likely due to a cancellation of error. Part of the reason could be that the calculation was performed in a methyl propanoate PCM solvent while the experimental data is obtained in a protein. A second reason could be the lower quality of TD-DFT in the case of \( \text{FlH}^- \), an open shell molecule. However, despite these shortcomings, the calculation reproduces the
spectral features and relative peak intensities of the experimental \textbf{FlH}∙ UV/vis spectrum in PaNMO very well.

While solvent data is not available for \textbf{FlH}∙ due to its short lifetime in solution, our calculations are highly consistent with experimental UV/vis spectra of N(3)-methyl-N(5)-ethyl-N(5)-monohydrolumiflavin semiquinone,\textsuperscript{99} an alkylated flavin semiquinone model that is stable in solution. Many of the same solvent effects discussed above have also been observed for the corresponding peaks in the alkylated semiquinone UV/vis spectrum, which has many of the same spectral features of neutral flavoprotein semiquinones.\textsuperscript{99-106}

3.7.6 Comparison with other computational studies for \textbf{FlH}∙.

To our knowledge, few studies have modeled the UV/vis spectra of \textbf{FlH}∙.\textsuperscript{17,85,98} Faraji et al. have computed the vertical excitation energy of the lowest excited state of \textbf{FlH}∙ in a QM/MM model of photolyase at the SOS-CIS(D) level of theory,\textsuperscript{98} while Ai et al. recently reported CASPT2//CASSCF vertical excitation energies of the lowest excited state (574 nm) as well as the most intense band (363 nm) in \textbf{FlH}∙.\textsuperscript{85} The CASPT2//CASSCF values are in good agreement with our TD-DFT vertical excitation energies (the corresponding TD-DFT computed values are 582 nm and 360 nm, respectively).

3.7.7 The Fl∙ UV/vis spectrum and ESTMs.

The computed UV/vis spectra in different solvents, the experimental UV/vis spectrum in PaNMO,\textsuperscript{17} and the ESTMs for TD states 3 and 6 are shown in Figure 3.5. The experimental UV/vis absorption spectrum of Fl∙ features two main peaks at $\lambda_{\text{max}} \approx 490$ nm and $\lambda_{\text{max}} \approx 370$ nm. These two peaks are due to TD states 3 and 6 in Table 3.1, respectively, although the peak at around 370 nm also has a contribution from TD state 5.
Figure 3.5 Computed (solid lines) and experimental (dashed lines) UV/vis spectra of the anionic semiquinone form of flavin (Fl\textsuperscript{−}). The vibronic excitation energies computed in a methyl propanoate PCM model are shown as impulse lines in magenta. The corresponding broadened peaks are also shown in magenta. The solvent shift (computed from differences in vertical excitation energies) and changes in intensity associated with different solvents are also shown for gas (light red), cyclohexane (bright red), and water (blue). The experimental UV/vis spectrum corresponds to FAD\textsuperscript{−} in PaNMO.\textsuperscript{17} ESTMs corresponding to the two lowest bright excited states are shown to the right. The legend indicates the magnitude of the shifts relative to the reference gas-phase excitation energy in eV (and in nm in parentheses).

The computed UV/vis spectrum for Fl\textsuperscript{−} in methyl propanoate was also presented recently in ref. 17. Since Fl\textsuperscript{−}, just like FlH\textsuperscript{−}, is unstable in solution, we also computed the spectrum in methyl propanoate to compare the results with experiments from PaNMO.

The effect of solvent on the absorption spectrum of Fl\textsuperscript{−} is relatively minimal, unlike FlH\textsuperscript{−}. Only the relative intensities of the peaks are affected, while changes in the solvent polarity cause
virtually no change in the energy of the two peaks. Indeed, both ESTMs show a limited sensitivity to the environment; the peak at ~370 nm shows a maximum sensitivity of only a few nm to nearby +0.1e point charges, while the second peak at ~490 nm is shifted by just 1.2 nm at most. Note that the ESTM for the peak at ~490 nm is computed for the excited state that contributes the most intensity to that peak.

3.7.8 Comparison with experimental data for Fl⁻.

The calculations reproduce all of the spectral features of Fl⁻ in PaNMO, including the relative intensities of the two peaks. The computed λmax values are close to the experimental ones, with the peak at λmax≈490 nm slightly blue-shifted with respect to experiment and the peak at λmax≈370 nm slightly red-shifted. The limited effect of solvent on the UV/vis spectrum of Fl⁻ is in good agreement with experimental studies on metal chelates of substituted Fl⁻, which also show very small solvatochromic shifts.107

3.7.9 Comparison with other computational studies for Fl⁻.

Similar to FlH⁺, few studies have reported computed UV/vis data for Fl⁻.17, 85 Ai et al. report a CASPT2//CASSCF vertical excitation energies of 427 nm for the most intense band that is considerably red-shifted with respect to the experimental band (λmax≈370 nm). The TD-DFT calculations in this work indicate that the peak at ~370 nm is instead a combination of two bright excited states with vertical excitation energies of 358 and 375 nm.

3.7.10 The FlH₂ UV/vis spectrum and ESTMs

The computed UV/vis spectra in different solvents, the experimental UV/vis spectrum in aqueous solution,52 and the ESTMs for FlH₂ are shown in Figure 3.6. The FlH₂ experimental spectrum features an ill-defined peak at the near-UV edge. Note that, unlike Fl, FlH⁺, and Fl⁻, the ground
state of $\text{FlH}_2$ is non-planar because it is bent along a so-called “butterfly” mode.$^{40-42}$ As a result, FC factors could not be computed within the double-harmonic approximation without making a number of additional approximations (see the SI). To reproduce the experimental UV/vis spectrum, it was necessary to renormalize each peak independently and use a different FWHM for each peak.

**Figure 3.6** Computed (solid lines) and experimental (dashed lines) UV/vis spectra of the neutral hydroquinone form of flavin ($\text{FlH}_2$). The vibronic excitation energies computed in a water PCM model are shown as impulse lines in blue. The corresponding broadened peaks are also shown in blue. The solvent shift (computed from differences in vertical excitation energies) and changes in intensity associated with different solvents are also shown for gas (light red) and cyclohexane (bright red). The experimental UV/vis spectrum corresponds to aqueous $\text{FADH}_2$. $^{52}$ ESTMs corresponding to the two bright excited states near the 300-400 nm range of the spectrum are shown to the right. The legend indicates the magnitude of the shifts relative to the gas-phase reference excitation energy in eV (and in nm in parentheses).
The solvent does not have a large effect on the FlH₂ UV/vis spectrum. State 2, which is slightly red-shifted in water compared to cyclohexane, is more sensitive to the solvent than state 3, which only shifts a little to the blue in polar environments.

The ESTMs in Figure 3.6 correspond to states 2 and 3. The two ESTMs are almost mirror images, which means that the two peaks should behave in opposite ways to a change in their environment. Indeed, the PCM calculations indicate that the solvent polarity has opposite effects on the two peaks. More generally, it is expected that a mutation or protein conformational change that introduces a positive charge near the hydrophilic part of flavin would blue-shift the shoulder at 350-400 nm while red-shifting the peak at 300 nm. The opposite will be true for the introduction of a positive charge near the hydrophobic part of the molecule.

3.7.11 Comparison with experimental data for FlH₂

It was more difficult to assign the experimental peaks to the TD-DFT calculation in the case of FlH₂ than for the other molecules. The spectral features could only be reproduced by using a relatively large FWHM for broadening, normalizing the peaks independently, and setting the intensity of state 1 to zero because it could not be easily assigned to a peak in the experimental spectrum. Note that states 2 and 3 have considerably higher computed oscillator strength than state 1 (see Table 3.1), but the reason this is not reflected in the vibronic impulse lines in Figure 3.6 is because the FC overlaps were small for states 2 and 3 due to difficulties in computing FC overlaps. Therefore, it is likely that state 1 is a low-intensity peak that contributes to the tail of the experimental spectrum. In fact, Ghisla et al. have shown that a peak with λₘₐₓ≈405 nm develops in the UV/vis spectrum of FlH₂ at low temperatures. This suggests that state 1 corresponds to that peak. Indeed, the vertical TD-DFT energy for state 1 is 411 nm, but the simulated λₘₐₓ is considerably red-shifted due to the large computed geometry reorganization energy; the adiabatic
energy for state 1 is computed at 537 nm. State 2 contributes to the shoulder at around 350-400 nm, while state 3 contributes to the sharp rise in absorption at around 300 nm.

3.7.12 Comparison with other computational studies for FlH₂

Calculations on FlH₂ have been performed with semi-empirical methods, TD-DFT, and CASPT2//CASSCF. The TD-DFT calculations by Choe et al. also employ B3LYP, although with a different basis set and solvation model than used in our calculations. Their study only reports vertical excitation energies, which are close (within 0.15 eV) to those in Table 3.1. The CASPT2//CASSCF report by Ai et al. only reports the vertical excitation energy corresponding to the peak at λ_max≈300 nm.

3.7.13 The FlH⁻ UV/vis spectrum and ESTMs

Computed UV/vis spectra in different solvents, experimental UV/vis spectra in aqueous solution and photolyase, and ESTMs for FlH⁻ are shown in Figure 3.7. Just like FlH₂, this system is non-planar and has an ill-defined spectrum with a shoulder appearing in the near-UV range. State 1 is a dark state that does not appear in the experimental spectrum or only contributes to the tail, and therefore we do not consider its contribution to the simulated spectra. The shoulder at around 350 - 400 nm and the peak at around 300 nm correspond to states 2 and 3, respectively, which have considerably higher oscillator strength than state 1 (see Table 3.1).
Figure 3.7 Computed (solid lines) and experimental (dashed lines) UV/vis spectra of the anionic hydroquinone form of flavin (FlH\(^-\)). The vibronic excitation energies computed in a water PCM model are shown as impulse lines in blue. The corresponding broadened peaks are also shown in blue. The solvent shift (computed from differences in vertical excitation energies) and changes in intensity associated with different solvents are also shown for gas (light red) and cyclohexane (bright red). The experimental UV/vis spectrum corresponds to aqueous FADH\(^-\) in aqueous solution\(^52\) (black) and in photolyase\(^109\) (grey). ESTMs corresponding to the two bright excited states near the 300-450 nm range of the spectrum are shown to the right. The legend indicates the magnitude of the shifts relative to the gas-phase reference excitation energy in eV (and in nm in parentheses).

Although FlH\(^-\) and FlH\(_2\) have similar absorption spectra, FlH\(^-\) displays much larger variations than FlH\(_2\) in different solvent polarities. In particular, the shoulder at 350-450 nm in aqueous solution gets red-shifted in non-polar solvents and develops into a well-defined peak. The ESTMs for FlH\(^-\) are also different from those of FlH\(_2\). The legends in the ESTMs reveal a significantly
higher sensitivity to the electrostatic environment, particularly for the peak at 350-450 nm, consistent with the larger computed solvent effect in $\text{FlH}^-$. The ESTMs for states 2 and 3 in $\text{FlH}^-$ look similar, unlike $\text{FlH}_2$ where they were mirror images. This means that changes in the surrounding environment are expected to shift the two peaks in the same direction, which is exactly what is observed in the PCM solvent calculations; both peaks blue-shift due to increasing polarity.

### 3.7.14 Comparison with experimental data for $\text{FlH}^-$

The computed $\lambda_{\text{max}}$ obtained by broadening the excitations of states 2 and 3 match reasonably well with the experimental $\lambda_{\text{max}}$ value. At the moment, it is not clear if state 1 contributes to the broadening of the shoulder that extends from around 350 nm to 500 nm, or if state 1 is truly a dark state that does not contribute to the experimental spectrum at all.

In **Figure 3.7**, we show the experimental UV/vis spectrum for $\text{FlH}^-$ in both aqueous solution and photolyase. The peak at 350-400 nm becomes red-shifted and more well-defined in photolyase compared to water. This can be understood due to the lower dielectric environment of the protein relative to water and is therefore strongly consistent with the computed solvent effects.\(^{109}\)

### 3.7.15 Comparison with other computational studies for $\text{FlH}^-$

Faraji et al.\(^{98}\) and Ai et al.\(^{85}\) have both reported on excitation energies for $\text{FlH}^-$. The former is a QM/MM study of photolyase and reports a SOS-CIS(D) vertical excitation energy in reasonably good agreement with the band at ~360 nm. The latter also reports the vertical excitation energy for the peak at ~360 nm, which they calculate at 388 nm. Our TD-DFT/cc-pVTZ vertical excitation for the corresponding peak in water is 345 nm.
3.7.16 Calculation of Franck-Condon factors

Franck-Condon (FC) factors are the overlap integrals between ground and excited state vibrational wave functions. In this work, FC factors were computed within the double-harmonic approximation and using Duschinsky rotations in ezSpectrum.\textsuperscript{48} The software aligns the ground and excited state geometries to maximize overlap, performs Duschinsky rotations to maximize the overlap between normal modes, and then computes the overlap integral between the vibrational wave functions. This is most successful when there are limited changes in the geometry between the ground state and excited state, such as cases where the geometric changes are restricted to bond length changes. However, large differences in angles or torsions may result in reduced FC overlaps. Those were the issues we encountered when computing FC factors between ground and excited states of flavin in some states. Most often, such issues arose when there were methyl-group rotations, but in the case of the reduced flavin species bending or pyramidalization of the isoalloxazine backbone also complicated the calculation of FC factors. While such changes all occur along low-frequency modes, they still result in a significantly lower FC overlap. To remedy this, there were a number of approximations in the calculation of FC overlaps of some states:

- Excited state geometries were optimized keeping the same methyl group orientations as in the ground state. Sometimes, these methyl conformations were local minima and not global minima on the excited state potential energy surface resulting in small imaginary frequencies associated with methyl group rotations. Those methyl group rotations were then assigned small positive frequencies (10-30 cm\textsuperscript{-1}) in the calculation of FC factors such that they had little to no contribution to vibronic broadening of the electronic transitions\textsuperscript{110} (an alternate approach used by Thiel and co-workers is to exclude such modes\textsuperscript{111, 112}).
In the $\text{Fl}$, $\text{FlH}^*$, and $\text{Fl}^-$ states, flavin remains planar in both the ground and $\pi \rightarrow \pi^*$ excited states, but this is not the case for $\text{FlH}_2$ and $\text{FlH}^-$. The hydroquinone ground state is bent at the central ring along a low-frequency (approximately 30-50 cm$^{-1}$) “butterfly bending” mode. However, the relevant $\pi \rightarrow \pi^*$ excited states in $\text{FlH}_2$ and $\text{FlH}^-$ are planar. This again resulted in low FC overlaps due to the difference in this bending motion. Therefore, to compute FC factors, the ground state hydroquinone was reoptimized in a planar conformation and any imaginary frequencies resulting from this constraint were assigned small positive frequencies (10-30 cm$^{-1}$) instead.

In the case of $\text{FlH}_2$, even the planar constraint did not yield substantially improved FC overlaps due to methyl rotations. Therefore, only for $\text{FlH}_2$, FC factors were computed in a reduced model with no methyl group substituents.

Note that the approximations above were only use for computing the FC factors, while all adiabatic and zero-point vibrational energies were computed using full flavin models that were properly optimized. The FC factors are used in this work to determine the vibronic structure of each electronic transition and may not be quantitative due to the reasons discussed above.
3.7.17 Simulation of experimental spectra using broadening and normalization

The simulation of experimental UV/vis spectra requires us to consider broadening of the computed vibronic excitation energies. In case of Fl, FlH+, and Fl− a full width at half maximum (FWHM) of 0.25 eV is used for all excited states. Relative peak intensities were computed from the relative oscillator strengths and FC overlaps of the corresponding excited states. Peak intensities for these three states were all normalized to the highest peak but were kept at the same relative intensities. In the case of FlH2 and FlH+, it was necessary to use different FWHM and renormalize each peak independently to get spectra that agree with experiment. Specifically, a FWHM of 0.80 eV is used for both peaks in the FlH2 spectrum while for FlH+ FWHM of 0.35 eV and 0.90 eV were used for the second and third excited states, respectively. In both FlH2 and FlH+, the first excited state is assumed to be dark. Therefore, the first, second, and third excited state intensities were scaled by factors of 0.00, 0.38, 1.00 and 0.00, 1.00, 0.50, respectively, for FlH2 and FlH+, respectively. This tuning of the FWHM and renormalizing that is needed to match experimental data in hydroquinones is likely due to the approximations made in the calculation of the FC factors. However, we note that broadening is expected to be larger in hydroquinones systems than in the quinone and semiquinones because of the low-frequency bending/pyramidalization mode differences in the ground and excited state.
Table 3.2 Experimental $\lambda_{\text{max}}$ values from literature for oxidized flavin

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Organism</th>
<th>Name</th>
<th>1st peak $\lambda_{\text{max}}$ (nm)</th>
<th>2nd peak $\lambda_{\text{max}}$ (nm)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N/A</td>
<td>Aqueous FMN</td>
<td>445</td>
<td>373</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>Aqueous FAD</td>
<td>450</td>
<td>375</td>
<td>113</td>
</tr>
<tr>
<td>4HIA</td>
<td><em>Rhodobacter sphaeroides</em></td>
<td>RsLOV</td>
<td>447</td>
<td>380</td>
<td>63</td>
</tr>
<tr>
<td>4EES</td>
<td><em>Arabidopsis thaliana</em></td>
<td>iLOV</td>
<td>447</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td><em>Arabidopsis thaliana</em></td>
<td>iLOV-Q489K</td>
<td>440</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>6GPU</td>
<td><em>Arabidopsis thaliana</em></td>
<td>miniSOG</td>
<td>448</td>
<td></td>
<td>61</td>
</tr>
<tr>
<td>4EEU</td>
<td><em>Arabidopsis thaliana</em></td>
<td>phiLOV2.1</td>
<td>450</td>
<td></td>
<td>55, 56, 59</td>
</tr>
<tr>
<td>2PR5</td>
<td><em>Bacillus subtilis</em></td>
<td>EcFbFP</td>
<td>448</td>
<td></td>
<td>21, 55, 57, 59, 62, 71</td>
</tr>
<tr>
<td>4KUK</td>
<td><em>Dinoroseobacter shibae</em></td>
<td>DsLOV</td>
<td>449</td>
<td></td>
<td>64</td>
</tr>
<tr>
<td></td>
<td><em>Chlamydomonas reinhardtii</em></td>
<td>CreiLOV</td>
<td>450</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>1N9L</td>
<td><em>Chlamydomonas reinhardtii</em></td>
<td>LOV1</td>
<td>447.0</td>
<td></td>
<td>66</td>
</tr>
<tr>
<td></td>
<td><em>Chlamydomonas reinhardtii</em></td>
<td>LOV1 F41Y</td>
<td>445.0</td>
<td>352.0</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td><em>Chlamydomonas reinhardtii</em></td>
<td>LOV2</td>
<td>445.5</td>
<td></td>
<td>66</td>
</tr>
<tr>
<td>3UE6</td>
<td><em>Vaucheria frigida</em></td>
<td>VafLOV</td>
<td>450</td>
<td></td>
<td>58, 68</td>
</tr>
<tr>
<td>5J3W</td>
<td><em>Pseudomonas putida</em></td>
<td>Pp1FbFP</td>
<td>450</td>
<td>376</td>
<td>55, 59, 65</td>
</tr>
<tr>
<td></td>
<td><em>Pseudomonas putida</em></td>
<td>Pp2FbFP</td>
<td>449</td>
<td></td>
<td>55, 59, 62, 65</td>
</tr>
</tbody>
</table>
3.8 Conclusion

In this work, we report vertical, adiabatic, and vibronic electronic excitations for five redox/protonation states of flavin. The excitation energies were used to simulate the UV/vis spectra by applying gaussian functions to broaden the computed vibronic excitations. We find that the simulated UV/vis spectra are in excellent agreement with experimental spectra in the case of Fl.

We simulate the UV/vis spectra for FlH∙ and Fl∙¯ in methyl propanoate, an approximate model of a protein dielectric, since experimental spectra are only available in a protein and not in solution. For both FlH∙ and Fl∙¯, we were able to reproduce the spectral features and relative intensities of...
all peaks in the near-UV/vis spectra reasonably well. In the case of FlH$_2$ and FlH$^-$, however, the agreement between TD-DFT computed and experimental excitation spectra is not as good as in Fl, particularly for FlH$_2$. This is in part due to difficulties with the calculation of FC factors due to the non-planarity of these two molecules in the ground state.

We simulated solvent effects on the absorption spectra and constructed “electrostatic spectral tuning maps” for the five redox/protonation states of flavin. Our calculations show, in agreement with earlier computational studies, that the UV/vis spectrum of Fl is not very sensitive to the solvent or protein environment, although the $\lambda_{\text{max}}$ and intensity of the peak at ~370 nm is more sensitive than the first excited state. In the case of the first excited state, however, our ESTMs suggest the specific electrostatic interactions near N$_1$ or C$_4$/N$_5$ may be used to shift the $\lambda_{\text{max}}$ of the first excited state as well. The UV/vis spectrum of FlH$^-$ is considerably more sensitive to the electrostatic environment, such that decreasing the polarity is associated with a decrease of the intensity of the peak at ~450-500 nm and significant shifts in the relative positions/intensities of the peaks near 300-350 nm. On the other hand, Fl$^-$ is far less sensitive to its electrostatic environment. Finally, the opposite trend is observed for the neutral and anionic hydroquinones; while FlH$_2$ is not very sensitive to the environment, FlH$^-$ is significantly more sensitive, particularly in the visible range from around 350-450 nm. This can also be seen experimentally by comparing, for instance, the UV/vis spectrum of FlH$^-$ in aqueous solution with that of photolyase.

Flavin spectra are good probes of internal electrostatics in proteins. For example, it is already known that the UV/vis spectrum of oxidized flavin can be used to distinguishing between hydrophobic and hydrophilic environments. However, flavin’s ability to exist in multiple redox/protonation states means that there is potential for probing the microenvironment of a protein using several different states through simple redox/pH changes, providing substantially
more data. The success of this approach relies heavily on understanding how the absorption properties of each of flavin’s redox and protonation state are modulated by the protein environment. The aim of this paper is to aid in this understanding by computing the spectra in each redox/protonation state and computing the solvent effect and ESTMs. Such an understanding is also a first step towards studying the redox properties and photophysics of these states and how they are influenced by nearby electrostatics.

3.9 References


46. RT, M. *pyvdw_surface*, 2014.


nitronate monooxygenase from Pseudomonas aeruginosa PAO1 establishes NMO class I and II. *Journal of Biological Chemistry* 2014, 289 (34), 23764-23775.


THE EFFECT OF HYDROGEN-BONDING ON FLAVIN’S INFRARED ABSORPTION SPECTRUM

Mohammad Pabel Kabir, Yoelvis Orozco-Gonzalez, Gary Hastings and Samer Gozem


Copyright © Elsevier B.V. 2021

My contribution to this chapter was data curation, formal analysis, investigation, methodology, validation, visualization, writing-original draft.

4.1 Abstract

Cluster and continuum solvation computational models are employed to model the effect of hydrogen bonding interactions on the vibrational modes of lumiflavin. Calculated spectra were compared to experimental Fourier-transform infrared (FTIR) spectra in the diagnostic 1450 – 1800 cm\(^{-1}\) range, where intense \(\nu_{C=O}, \nu_{C=N}, \nu_{C_2=O}, \text{ and } \nu_{C_4=O}\) stretching modes of flavin’s isalloxazine ring are found. The computations indicate that \(\nu_{C=O}\) and \(\nu_{C=N}\) mode frequencies are relatively insensitive to intermolecular interactions while the \(\nu_{C_2=O}\) and \(\nu_{C_4=O}\) modes are sensitive to direct (and also indirect for \(\nu_{C_2=O}\)) hydrogen-bonding interactions. Although flavin is neutral, we find
that use of basis sets without the diffuse functions provide incorrect relative frequencies and intensities. The 6-31+G* basis set is found to be adequate for this system, and there is limited benefit to considering larger basis sets. Calculated vibrational mode frequencies agree with experimentally determined frequencies in solution when cluster models with multiple water molecules are used. Accurate simulation of relative FTIR band intensities, on the other hand, requires a continuum (or possibly quantum mechanical/molecular mechanical) model that accounts for long-range electrostatic effects. Finally, an experimental peak at ca. 1624 cm\(^{-1}\) that is typically assigned to the \(\nu_{\text{C}=O}\) vibrational stretching mode has a complicated shape that suggests multiple underlying contributions. Indeed, our calculations show that this band has contributions from both the C\(_6\)-C\(_7\) and C\(_2\)=O stretching vibrations.

4.2 Key words
Flavin, FMN, FAD, hydrogen bonding, infrared spectroscopy, FTIR.

4.3 Introduction
Flavin-dependent proteins are found in all kingdoms of life, where they are responsible for a wide range of biological processes.\(^1\)-\(^4\) Flavoproteins typically bind either flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD), both of which are derivatives of riboflavin (vitamin B\(_2\)). Flavin acts as a redox agent in many enzymatic reactions,\(^5\),\(^6\) although there are also multiple instances where flavin catalyzes reactions with no redox change.\(^7\),\(^8\) Flavin has also been implicated as the active chromophore in a number of photoreceptors and light-responsive enzymes.\(^9\)-\(^13\)

The physical and chemical changes accompanying catalytic and photochemical processes in flavoproteins and the resulting changes to the microenvironment such as hydrogen-bonding interactions could be probed using Fourier-transform infrared (FTIR) spectroscopy. However, since proteins are macromolecules with multiple functional groups, their congested FTIR spectra
make it difficult to isolate specific interactions. Flavins are multi-redox, photoactive compounds that can exist in multiple protonation states, making it possible to probe changes in their FTIR spectra as a result of some chemical or physical change using FTIR difference spectroscopy (DS). FTIR DS techniques are sensitive to small changes in intra- and inter-molecular interactions that may occur during any process and are therefore excellent tools to probe flavin redox chemistry or photochemistry.

**Figure 4.1** The tricyclic structure common to FMN, FAD, riboflavin, and lumiflavin.

The interpretation of experimental FTIR DS in proteins is not always straightforward, in part because vibrational frequencies and intensities are affected by electrostatic and hydrogen-bonding interactions within the protein. On the other hand, it is precisely these changes that make redox molecules like flavin useful probes of their protein microenvironment. Interpretation of experimental FTIR DS requires knowledge of how electrostatic and hydrogen-bonding interactions modify flavin’s FTIR spectrum. This knowledge can be obtained from computational modeling, which is one of the motivations for the work reported here.

One approach to interpreting how specific interactions modulate the FTIR spectra of a protein-bound molecule (e.g., flavin in a flavoprotein) is with hybrid quantum mechanical / molecular
Several QM/MM methods and tools have been developed to make it possible to simulate FTIR DS of protein-bound molecules more efficiently and to help assign peaks in experimental spectra to specific molecular modes. There are also some recent instances where such QM/MM methods were employed to simulate the FTIR spectra of flavins in different flavoproteins. However, while such calculations are specific to the protein being modeled, it is also desirable to look at a straightforward model system to try to understand how nearby interactions at different positions influence flavin’s vibrational spectra.

The goal of this study is to answer the following two questions:

1. Can explicit solvent or continuum solvation models reproduce the relative frequencies and intensities of prominent peaks in flavin’s IR spectrum in an aqueous solution?

2. How do specific non-bonding interactions with nearby polar molecules affect the vibrational frequencies and intensities of flavin’s isoalloxazine ring?

In this manuscript we will focus on the four most prominent vibrational modes of flavin’s isoalloxazine ring, which all appear in the 1450 – 1800 cm$^{-1}$ region for flavin in solution. Several experimental FTIR spectra for FMN or FAD in solution are shown in Figure 4.8. The four bands are observed near 1548 cm$^{-1}$ ($v_{C=C}$), 1580 cm$^{-1}$ ($v_{C=N}$), 1640 cm$^{-1}$ ($v_{C_2=O}$), and 1700 cm$^{-1}$ ($v_{C_4=O}$) for samples in D$_2$O. All experimental spectra in Figure 4.8 display bands within 5 cm$^{-1}$ of these indicated values, with the exact frequency depending on whether experiments employ FMN or FAD or whether the samples are in D$_2$O or H$_2$O (the $v_{C_2=O}$ and $v_{C_4=O}$ bands are upshifted in H$_2$O). The corresponding normal modes are shown in Figure 4.9 (the assignments in the parentheses describe the dominant stretching mode). Since we focus on vibrational modes of the
isoalloxazine ring of flavin, we employ lumiflavin as a model system (see Figure 4.1 for structures).

4.4 Methodology

The main goal of this study is to determine how intermolecular interactions between flavin and other polar molecules influence flavin’s vibrational mode frequencies. As a starting point for disentangling the effect of such interactions, we will model the FTIR spectrum of lumiflavin interacting with one water molecule placed at different positions around the isoalloxazine ring. To find such potential interactions, we use the approach outlined in Figure 4.2.

Figure 4.2 Approach used to find potential lumiflavin-water interactions. We first optimized lumiflavin in the gas phase (panel 1). We then use a surface that is 3.4 Å from all atoms of the molecule (panel 2) and locate points on the surface at a density of 1 point per 5 Å² (panel 3). Note that 3.4 Å is slightly longer than the length of a typical hydrogen bond distance between heavy atoms. We place an oxygen atom at each point and add two hydrogen atoms to each oxygen to get a total of 106 water positions (panel 4). For lumiflavin in the presence of each one of these waters, we run an unconstrained geometry optimization and frequency calculation (a total of 106 geometry optimization and frequency calculations were performed, each with lumiflavin and one single water molecule). Of the 106 calculations, 92 converged to a structure where the water is hydrogen-bonded or non-covalently bound to the lumiflavin.
The approach in Figure 4.2 is related to a previously reported Electrostatic Spectral Tuning Map (ESTM) approach,\textsuperscript{38,39} and employs the open-source python library pyvdwsurface.\textsuperscript{40} In this work, the approach is used to search for local geometry minima that represent different interactions between lumiflavin and water molecules. While 92 structures were optimized, many of these structures were similar, indicating there are only a few local minima for isoalloxazine-water interactions. Specifically, eight positions were identified where water had a tendency to go. Waters placed in the hydrophilic side of lumiflavin ended up in one of the five positions shown in orange in Figure 4.3, while waters placed near the hydrophobic side of lumiflavin ended up in one of the three positions shown in blue in Figure 4.3.

In addition to those eight structures that each included a single water molecule, we optimized several cluster models with multiple water molecules. Specifically, we chose a structure with two water molecules both hydrogen-bonded to one carbonyl (green structures in Figure 4.3) and three cluster models with more than two water molecules (magenta structures in Figure 4.3). The cluster models include a model with all waters near the hydrophilic side (6W; a sixth water was added near N\textsubscript{3} to stabilize the two nearby water molecules) as well as two models with waters both near both the hydrophilic and hydrophobic sides of lumiflavin (8W and 9W).
Figure 4.3 Lumiflavin + water structures employed in this study. Structures in orange and blue regions were found using the approach outlined in Figure 4.2. Structures in the green region include two water molecules, while structures in the magenta region include six to nine water molecules. The labels, which will be used throughout this work, are indicated next to each structure.

All geometry optimizations and frequency calculations employed the B3LYP density functional theory and 6-31+G* basis set, unless indicated otherwise. B3LYP remains one of the most widely used functionals for the calculation of ground-state frequencies and has been found to be consistently reliable in many molecules when used with a constant scaling factor. Water molecules were treated at the QM level in all cases. In addition to the models shown in Figure 4.3, lumiflavin’s geometry and frequencies were computed using a polarizable continuum model.
(PCM) using the integral equation formalism (IEF) accounting for a water solvent. Gas-phase models serve as a reference for the effect of solvation or specific hydrogen-bonding interactions. For all models, only positive frequencies were calculated.

FTIR experiments are normally conducted for samples in D$_2$O (Figure 4.8) to avoid bands due to the very intense water bending vibrations that can overwhelm flavin bands of interest. Therefore, our frequency calculations included an isotope effect by replacing all hydrogen atoms likely to undergo isotope exchange with deuterium atoms. Specifically, the flavin N$_3$ hydrogen and all water hydrogens were replaced with deuterium for the frequency calculations. All calculations were undertaken using Gaussian 16 software. A comparison between frequency calculations with and without deuteration are reported in Figure 4.10, and reproduce the experimentally observed differences between flavin spectra in H$_2$O and D$_2$O.

To quantify the strength of hydrogen bonding interactions in the different models shown in Figure 4.3, local mode force constants $k_a$ were computed using LMODEA. $k_a$ is a powerful metric for representing bond strength between two atoms described by an internal bond length coordinate ($q$). It has also been demonstrated on multiple occasions that local mode theory can be used for determining the strength of non-covalent interactions such as hydrogen bonding.

Using LMODEA, $k_a$ was computed for hydrogen---acceptor (H---A) non-covalent interactions. $k_a$ can be related to a more chemically intuitive parameter, the bond order ($n$), through a power relationship, as long as the bond order is defined with respect to some reference or boundary. Therefore, we employ the following equation derived by Freindorf et. al. to compute the H---A “bond orders” for flavin-water interactions for each of the models in Figure 4.3.

$$n = (0.532) \, k_a^{0.278} \quad (1)$$
Freindorf et. al.\textsuperscript{56} determined this power relationship between \( n \) and \( k_a \) through a comprehensive computational study of many hydrogen bonding pairs. Specifically, for H---A non-covalent interactions, the bond order was calibrated using 0 for no hydrogen bonding interaction and 0.5 for the strongest ([F---H---F]\textsuperscript{−}) hydrogen bond where the hydrogen is equally shared between two fluorine centers. For bonded interactions, \( n \) is 1 for the F–H monomer (full single covalent bond). Using these reference values, the weakest hydrogen bond reported by Freindorf et. al. was between ammonia and difluorine (\( n=0.134 \)).\textsuperscript{56} These values will serve as calibration points for the discussion of hydrogen bonding interactions in this work.

### 4.5 Results and Discussion

Table 4.1 displays \( k_a \) and \( n \) values for all of the hydrogen bonding interactions in the 9W model. \( n \) is computed from \( k_a \) using equation (1). As expected, the strongest interactions (largest \( k_a \) and \( n \)) are between water molecules and carbonyl oxygens. The values reported in Table 4.1 are consistent with typical hydrogen bond interaction strengths reported in Ref. 56. The N\textsubscript{3}–H---OH\textsubscript{2} interaction is also considerably strong, likely due to the strong polarity of the N\textsubscript{3}-H bond caused by aromaticity of N\textsubscript{3}. As similar strength has been observed in hydrogen bonding interactions involving imidazole.\textsuperscript{56}

Table 4.1 Computed bond distance (\( q \)), local mode force constant (\( k_a \)), and bond order (\( n \)) for various hydrogen-bonding interactions in 9W.

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Bond</th>
<th>( q ) (Å)</th>
<th>( k_a ) (mdyn/ Å)</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>N\textsubscript{1}---HOH</td>
<td>N\textsubscript{1}---H</td>
<td>2.093</td>
<td>0.122</td>
<td>0.30</td>
</tr>
<tr>
<td>N\textsubscript{3}–H---OH\textsubscript{2}</td>
<td>H---O</td>
<td>1.853</td>
<td>0.231</td>
<td>0.35</td>
</tr>
<tr>
<td>C\textsubscript{2}=O---HOH</td>
<td>O---H</td>
<td>1.874</td>
<td>0.199</td>
<td>0.34</td>
</tr>
</tbody>
</table>
The weakest non-covalent interactions in Table 4.1, as expected, are C–H---water interactions, which do not constitute hydrogen bonds. The bond orders of these interactions, however, are of similar magnitude (or even strong) than interactions between ammonia and non-polar difluorine reported by Freindorf et al. (n=0.134). This is likely due to the small dipole in C–H bonds, which creates a just strong enough dipole interaction with the water to keep it from dissociating during geometry optimizations.

To determine whether hydrogen bonding interaction strengths are consistent in the other models shown in Figure 4.3, n was computed for each hydrogen bonding interaction in all the models. The results are included in Table 4.5. The calculations show that n is consistent for each type of interaction in the different models, with the exception of hydrogen bonding interactions near N1; since 6W, 8W, and 9W include two water molecules in the vicinity of N1 and C2=O, the waters are well oriented in those systems to form a strong hydrogen bond interaction with each of N1 and C2=O. However, in the case of the N1 and C2=O models with one water molecule (shown in orange in Figure 4.3), there is no water network to support two individual hydrogen bonds so instead the

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Bond Order</th>
<th>Bond Length</th>
<th>Bond Order</th>
<th>Bond Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2=O---HOH’</td>
<td>O---H</td>
<td>1.713</td>
<td>0.334</td>
<td>0.39</td>
</tr>
<tr>
<td>C4=O---HOH</td>
<td>O---H</td>
<td>1.868</td>
<td>0.209</td>
<td>0.34</td>
</tr>
<tr>
<td>N5---HOH</td>
<td>N5---H</td>
<td>2.069</td>
<td>0.134</td>
<td>0.30</td>
</tr>
<tr>
<td>C9–H---OH2</td>
<td>H---O</td>
<td>2.388</td>
<td>0.048</td>
<td>0.23</td>
</tr>
<tr>
<td>C7–Methyl–H ---OH2</td>
<td>H---O</td>
<td>2.649</td>
<td>0.023</td>
<td>0.19</td>
</tr>
<tr>
<td>C8–Methyl–H ---OH2</td>
<td>H---O</td>
<td>2.705</td>
<td>0.033</td>
<td>0.21</td>
</tr>
<tr>
<td>C7–Methyl–H ---OH2’</td>
<td>H---O</td>
<td>2.584</td>
<td>0.018</td>
<td>0.17</td>
</tr>
<tr>
<td>C8–Methyl–H ---OH2’</td>
<td>H---O</td>
<td>2.594</td>
<td>0.006</td>
<td>0.13</td>
</tr>
</tbody>
</table>
single water interacts less strongly with both the N₁ and C₂=O of flavin, yielding smaller bond orders for those interactions (Table 4.5).

**Figure 4.4** Experimental\(^{12}\) (bottom panel) and computed (all other panels, each panel representing a different model) IR spectra in the 1750 – 1450 cm\(^{-1}\) region. Four bands dominate the spectra and are assigned to \(\nu_{C=O}\) (blue), \(\nu_{C=N}\) (green), \(\nu_{C₂=O}\) (grey), and \(\nu_{C₄=O}\) (yellow). Colored dashed lines are used to help compare calculated and experimental spectra. The background is colored in a way that corresponds to the color scheme in **Figure 4.3** to indicate the model used. In part A, calculated frequencies were scaled by 0.964, while in part B the frequencies were scaled so that the \(\nu_{C=O}\) calculated frequency is equal to 1548 cm\(^{-1}\). The scale factors are indicated on the right side of panel B.

**Figure 4.4** shows experimental (bottom panel, for FMN) and calculated (all other panels, for lumiflavin) IR spectra in the 1750 – 1450 cm\(^{-1}\) region. The PCM model in **Figure 4.4** includes only the lumiflavin and PCM solvent, while all other models include explicit QM water molecules.
without continuum solvation. All computed frequencies were broadened using gaussians with an 8 cm$^{-1}$ full-width at half-maximum (FWHM) to simulate the broadening observed experimentally for the intense $\nu_{C=\text{C}}$ band near 1548 cm$^{-1}$. Initially, a constant scaling factor of 0.964 was applied to all computed frequencies (Figure 4.4A). While those calculations reproduce the main features of the experiment, the calculated bands in most models are consistently downshifted relative to the corresponding experimental bands. Therefore, the calculations were (re)scaled such that the calculated $\nu_{C=\text{C}}$ peak frequency is found at 1548 cm$^{-1}$, matching the experimentally determined frequency (Figure 4.4B).

In all models using explicit QM water molecules, a scaling factor between 0.972 – 0.977 is needed to match the $\nu_{C=\text{C}}$ peak frequency. However, the PCM model required a scaling factor of 0.987, considerably larger than the recommended 0.964 scaling factor. Of greater concern is that the PCM spectrum displays an extra, relatively intense peak near 1520 cm$^{-1}$ (Figure 4.4B) that is not present in the experimental spectrum (or any of the other calculated spectra). This additional PCM calculated peak is predominantly due to a coupled $\text{C}_6=\text{C}_7/\text{C}_{10a}=\text{N}_1$ stretching vibration (see Figure 4.1 for atom number scheme and see Figure 4.11 left for the normal mode).

In the 6W-9W cluster models, there is a second vibration with a relatively strong intensity that is only slightly offset from the $\nu_{C=\text{O}}$ peak at around 1624 cm$^{-1}$. This mode is due to a combination of $\text{C}_6-\text{C}_7$ and $\nu_{C=\text{O}}$ stretching modes and is not easily resolved as an identifiable peak in our broadened spectra. This may explain why, in the experimental spectra, the $\nu_{C=\text{O}}$ peak near 1640 cm$^{-1}$ appears broader than the $\nu_{C=\text{O}}$ peak near 1700 cm$^{-1}$. A shoulder is resolved at around 1640 cm$^{-1}$ in some experiments, especially for FAD spectra but also in the FMN spectra of Iuliano et
al.\textsuperscript{32} and El-Khoury et al.\textsuperscript{33} (in D\textsubscript{2}O, see Figure 4.8). The contribution of such a mode to this peak has also been discussed in another computational study.\textsuperscript{60}

In principle, since normal modes are delocalized and include coupled vibrations from multiple bonds and angles, contributions of various internal coordinates to a specific normal mode (such as the one at 1624 cm\textsuperscript{-1}) could be better understood using a decomposition of normal modes into local modes.\textsuperscript{61,62} However, in molecules such as lumiflavin, the choice of a set non-redundant set of internal coordinates required to perform such a decomposition is not obvious.

\textbf{Figure 4.5 A.} The frequencies of the $\nu_{C=C}$ (blue), $\nu_{C=N}$ (green), $\nu_{C_2=O}$ (grey), and $\nu_{C_4=O}$ (yellow) modes computed with different models. The experimental reference\textsuperscript{32} is indicated with a dashed line, and the corresponding frequency on the right vertical axis. A single scaling factor of 0.964 is used. The background is colored coded as in Figure 4.3 and Figure 4.4 to indicate the model used.

\textbf{B.} The same data but with the frequencies scaled so that the calculated $\nu_{C=C}$ frequency matches the experiment.

To more easily compare calculated and experimental frequencies for the different molecular models, \textbf{Figure 4.5A} shows a plot of the calculated frequencies of each of the four modes ($\nu_{C=C}$, $\nu_{C=N}$, $\nu_{C_2=O}$, and $\nu_{C_4=O}$) for the different models, using a constant scaling factor of 0.964. With
this scaling factor, none of the models give calculated frequencies that agree well with experiments. However, all models show a similar calculated frequency for the $\nu_{C=\pi}$ mode (which is underestimated relative to the experiment based on the chosen scale factor). This indicates that the $\nu_{C=\pi}$ modes are not sensitive to any of the flavin-water intermolecular interactions. **Figure 4.5B** plots the data using a tailored scaling factor (the same scaling factors shown in **Figure 4.4B**). Using this frequency scaling, the PCM calculated spectrum is more in line with the experimental data (except the additional peak discussed above). An important observation, which may be useful in calculations for flavin in proteins, is that the calculated frequency differences between the $\nu_{C=\pi}$ and $\nu_{C=N}$ modes are similar for all the molecular models (**Figure 4.5B**). That is, the computed $\nu_{C=\pi}$ and $\nu_{C=N}$ mode frequencies are insensitive to the molecular model, and hence are independent of hydrogen bonding interactions.

On the other hand, the $\nu_{C_2=O}$ and $\nu_{C_4=O}$ mode frequencies are very sensitive to the details of the interactions with water molecules (**Figure 4.5**). In the calculations, the $\nu_{C_2=O}$ and $\nu_{C_4=O}$ mode frequencies downshifts 69 cm$^{-1}$ and 57 cm$^{-1}$, respectively, relative to the gas phase when hydrogen bonded to two water molecules (see data with green background in **Figure 4.5B** compared to the gas phase). Additional waters (i.e., in the 6W, 8W, and 9W cluster models; magenta background in **Figure 4.5B**) downshift the $\nu_{C_2=O}$ frequency further but upshift the $\nu_{C_4=O}$ frequency relative to the model where it is the only hydrogen bonded group. Overall, it appears that the $\nu_{C_2=O}$ frequency is more sensitive to H-bonding (even if that H-bonding is to other nearby atoms) than the $\nu_{C_4=O}$ frequency, which is only downshifted by direct H-bonding. When only the C$_4$=O is H-bonded but not C$_2$=O, the $\nu_{C_4=O}$ frequency is lower than that of $\nu_{C_2=O}$ (see C$_4$=O and 2(C$_4$=O) in **Figure 4.5B**).
In summary, the calculations shown in Figure 4.5B indicate that the $\nu_{C=\cdot C}$ and $\nu_{C=\cdot N}$ frequencies are not sensitive to hydrogen-bonding interactions, while the $\nu_{C=\cdot O}$ frequency displays the highest sensitivity to any H-bonding near flavin’s hydrophilic ring. The $\nu_{C=\cdot O}$ frequency is only shifted by direct hydrogen-bonding. These observations are consistent with the flavin spectra in H$_2$O and D$_2$O (see Figure 4.11); using a deuterated solvent has the largest effect on the $\nu_{C=\cdot O}$ peak, the next largest effect on the $\nu_{C=\cdot O}$, and no detectable effect on the $\nu_{C=\cdot C}$ and $\nu_{C=\cdot N}$ frequencies.

In addition to reproducing the vibrational frequencies, we also consider whether the calculations can reproduce the experimentally observed relative intensities of the IR bands. Experimental relative intensities were established by considering the relative areas under the FTIR absorption bands (using the relative heights is misleading since the peaks are not broadened to the same extent). This relative area is established by fitting each absorption band to a gaussian function and then integrating to obtain the area under each fit. The calculated relative intensities are obtained by dividing the computed mode intensities. The computed and experimental relative intensities are compared in Figure 4.6.
Figure 4.6 The relative intensities of the $\nu_{C=C}$ and $\nu_{C=N}$ peaks (blue) and $\nu_{C=O}$ and $\nu_{C=O}$ peaks (green) computed with different models. The experimental references are indicated with a dashed line with corresponding relative intensity shown on the right (obtained by finding the relative areas under the peaks from the experimental FTIR spectrum in ref 34).}

Experimentally, the $\nu_{C=C}$ band is more than 4.15 times more intense than the $\nu_{C=N}$ band. Only the PCM model reproduces this experimental relative intensity ratio, while the gas-phase and atomistic models all indicate $\nu_{C=C}$ and $\nu_{C=N}$ bands of nearly equal intensity (blue data points in Figure 4.6). Since the $\nu_{C=C}$ and $\nu_{C=N}$ relative intensities can only be reproduced using a dielectric continuum, this suggests that long-range electrostatic interactions contribute to changing the relative intensities. This hypothesis is supported by QM/MM calculations\textsuperscript{60} that reproduced the large difference in intensities of the $\nu_{C=C}$ and $\nu_{C=N}$ modes only when using a solvent QM/MM model but not in the gas phase. In conclusion, it is likely that the different intensities of the $\nu_{C=C}$ and $\nu_{C=N}$ modes of flavin are not due to a single specific flavin-water interaction, but rather are due to a dielectric effect of the collective solvent molecules.
While the $\nu_{C=C}/\nu_{C=N}$ relative intensity calculated using the PCM approach is in line with the experiment, the PCM calculated $\nu_{C_2=O}/\nu_{C_4=O}$ relative intensity is not (Figure 4.6). Many of the other molecular models, however, do give calculated $\nu_{C_2=O}/\nu_{C_4=O}$ relative intensities that align well with the experiment (Figure 4.6).

Finally, we have studied the effect of changing the basis set on the results of the calculations. We pick three representative models that reproduce aspects of the experimental data well: PCM, 6W, and 9W. A comparison of how the calculated mode frequencies depend on the basis set is outlined in Figure 4.7A. The frequencies were scaled such that the computed $\nu_{C=C}$ frequency matches the experiment (as in Figure 4.4B) for each basis set. How the relative mode intensity ratios vary as a function of basis set is outlined in Figure 4.7B.

**Figure 4.7 A.** The frequencies of the $\nu_{C=C}$ (blue), $\nu_{C=N}$ (green), $\nu_{C_2=O}$ (grey), and $\nu_{C_4=O}$ (yellow) modes computed with different basis sets with the PCM (circles), 6W atomistic (triangles), and 9W atomistic (squares) models. The experimental reference is indicated with a dashed line and the corresponding frequency on the right. Frequencies were scaled such that the C=C computed frequency matches the experiment. **B.** The relative intensities of the $\nu_{C=C}$ and $\nu_{C=N}$ peaks (blue) and $\nu_{C_2=O}$ and $\nu_{C_4=O}$ peaks (green) computed with different basis sets with the PCM (circles), 6W atomistic (triangles), and 9W atomistic (squares) models. The experimental reference, indicated
with a dashed line and the corresponding relative intensity on the right, is obtained by integrating to find the relative areas under the peaks from the experimental FTIR spectrum from ref. 32.

In the absence of diffuse functions (i.e., with the 6-31G* basis set), the $\nu_{C_2=O}$ and $\nu_{C_4=O}$ frequencies in the PCM and 6W models agree poorly with the experiment (Figure 4.7A). The calculations converge with the 6-31+G* basis set, and there appears to be limited benefit of considering calculations (of the type performed here) with a larger basis set, especially for the 6W and 9W models. Diffuse functions are likely necessary to correctly describe the noncovalent interactions between the flavin and water. We note that several previous studies have used either the 6-31G or 6-31G* basis set to model the FTIR spectrum of flavin, which may explain why these models could not reproduce well the relative frequencies of those prominent vibrational modes.

To better understand the origin of the basis-set dependence observed in Figure 4.7, we computed bond orders from $\kappa$s computed using different basis sets (Table 4.2). We find that computed bond orders are consistent across all the basis sets (within 0.01-0.05), with a notable exception; The N1--HOH interaction is weaker in the case of the 6-31G and 6-31G* basis sets relative to larger basis set calculations. Due to missing diffuse functions, this interaction is weaker in the small basis calculations, which may be the origin of the error observed for those basis sets in Figure 4.7. For larger basis sets, on the other hand, it appears that having diffuse and polarization functions on hydrogen atoms does not considerably change the description of hydrogen bonding interactions, since 6-31+G* bond orders in Table 4.2 are largely consistent with results obtained with larger basis sets.
**Table 4.2** Computed bond orders (n) for all hydrogen-bonding interactions in 9W computed with different basis sets.

<table>
<thead>
<tr>
<th>Interaction</th>
<th>6-31G</th>
<th>6-31G*</th>
<th>6-31+G*</th>
<th>6-311+G*</th>
<th>6-311+G**</th>
<th>6-311++G**</th>
</tr>
</thead>
<tbody>
<tr>
<td>N₁---HOH</td>
<td>0.16</td>
<td>0.18</td>
<td>0.30</td>
<td>0.30</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>N₃–H---OH₂</td>
<td>0.36</td>
<td>0.34</td>
<td>0.35</td>
<td>0.35</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>C₂=O---HOH</td>
<td>0.35</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>C₂=O---HOH’</td>
<td>0.41</td>
<td>0.40</td>
<td>0.39</td>
<td>0.40</td>
<td>0.38</td>
<td>0.38</td>
</tr>
<tr>
<td>C₄=O---HOH</td>
<td>0.36</td>
<td>0.35</td>
<td>0.34</td>
<td>0.34</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>N₅---HOH</td>
<td>0.31</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>C₉-H---OH₂</td>
<td>0.28</td>
<td>0.27</td>
<td>0.23</td>
<td>0.26</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>C₇–Methyl–H--OH₂</td>
<td>0.21</td>
<td>0.21</td>
<td>0.19</td>
<td>0.21</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>C₈–Methyl–H--OH₂</td>
<td>0.19</td>
<td>0.19</td>
<td>0.21</td>
<td>0.21</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>C₇–Methyl–H--OH₂’</td>
<td>0.21</td>
<td>0.22</td>
<td>0.17</td>
<td>0.21</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>C₈–Methyl–H--OH₂’</td>
<td>0.21</td>
<td>0.20</td>
<td>0.13</td>
<td>0.21</td>
<td>0.20</td>
<td>0.19</td>
</tr>
</tbody>
</table>

The relative peak intensities for the different models (and PCM in particular) show a much larger variation with the change in basis set (**Figure 4.7B**). Specifically, the excellent agreement of the relative intensities in the case of the PCM model with 6-31+G* appears fortuitous, since a further increase in basis set size leads to an increase of the relative intensities relative to the experiment. Increasing the basis set size gradually increases the \( \nu_{\text{C} = \text{C}} / \nu_{\text{C} = \text{N}} \) relative intensities in the explicit water QM models, but not enough to get close to the experimental relative intensities. The \( \nu_{\text{C}_2 = \text{O}} / \nu_{\text{C}_4 = \text{O}} \) relative intensities are captured very well by the explicit water QM models, on the other hand. This suggests that while the long-range electrostatic effect of the solvent has an effect on the \( \nu_{\text{C}_2 = \text{C}} / \nu_{\text{C}_4 = \text{C}} \) relative intensities, the \( \nu_{\text{C}_2 = \text{O}} / \nu_{\text{C}_4 = \text{O}} \) relative intensities are less sensitive to these long-range electrostatics.
Figure 4.8 Experimental FTIR absorption spectra for FMN or FAD in H$_2$O or D$_2$O. The spectra from references 1-7, as labeled on the right for each spectrum.
**Figure 4.9** The normal mode vectors corresponding to the four prominent FTIR peaks discussed in this study.

**Figure 4.10** Computed (for lumiflavin) and experimental\(^2,7\) (for FMN) FTIR absorption spectra in both deuterated (top four panels) and undeuterated (bottom four panels) solvent. In the deuterated PCM model, only the ionizable N\(_3\) hydrogen is deuterated, while in the cluster (6W and 9W) deuterated models the explicit waters are also deuterated.
Figure 4.11 Left. The normal mode vectors for the additional intense peak that appears in PCM calculations (but that is not intense in other atomistic calculations or the experiment). Right. The normal mode vectors appearing at around 1624 cm$^{-1}$ in water cluster models. This peak is close in frequency to the C$_2$=O vibrational mode, which may explain why the experimental C$_2$=O vibrational peak is broader than the C$_4$=O peak, even appearing to have a shoulder in some D$_2$O experimental spectra.

Table 4.3 Computed frequencies (in cm$^{-1}$) in the gas-phase and solvent models, scaled by 0.964.

<table>
<thead>
<tr>
<th>Model</th>
<th>C=C</th>
<th>C=N</th>
<th>C$_2$=O</th>
<th>C$_4$=O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas</td>
<td>1534</td>
<td>1571</td>
<td>1704</td>
<td>1720</td>
</tr>
<tr>
<td>PCM</td>
<td>1513</td>
<td>1550</td>
<td>1606</td>
<td>1661</td>
</tr>
<tr>
<td>N1</td>
<td>1534</td>
<td>1569</td>
<td>1681</td>
<td>1720</td>
</tr>
<tr>
<td>N3</td>
<td>1533</td>
<td>1570</td>
<td>1652</td>
<td>1717</td>
</tr>
<tr>
<td>C2=O</td>
<td>1535</td>
<td>1570</td>
<td>1666</td>
<td>1719</td>
</tr>
<tr>
<td>C4=O</td>
<td>1535</td>
<td>1571</td>
<td>1704</td>
<td>1672</td>
</tr>
<tr>
<td>N5</td>
<td>1535</td>
<td>1572</td>
<td>1691</td>
<td>1712</td>
</tr>
<tr>
<td>C7C8-t</td>
<td>1533</td>
<td>1569</td>
<td>1689</td>
<td>1717</td>
</tr>
<tr>
<td>C7C8-b</td>
<td>1533</td>
<td>1569</td>
<td>1690</td>
<td>1717</td>
</tr>
<tr>
<td>C9</td>
<td>1533</td>
<td>1569</td>
<td>1689</td>
<td>1716</td>
</tr>
<tr>
<td>2(C2=O)</td>
<td>1535</td>
<td>1567</td>
<td>1637</td>
<td>1719</td>
</tr>
<tr>
<td>2(C4=O)</td>
<td>1535</td>
<td>1572</td>
<td>1705</td>
<td>1665</td>
</tr>
<tr>
<td>6W</td>
<td>1530</td>
<td>1563</td>
<td>1618</td>
<td>1675</td>
</tr>
<tr>
<td>8W</td>
<td>1529</td>
<td>1558</td>
<td>1613</td>
<td>1672</td>
</tr>
<tr>
<td>9W</td>
<td>1528</td>
<td>1556</td>
<td>1611</td>
<td>1670</td>
</tr>
</tbody>
</table>
Table 4.4 Computed frequencies (in cm\(^{-1}\)) using gas-phase and different solvent models, scaled by such that the C=C vibrational mode matches the experimental value (1548 cm\(^{-1}\)).

<table>
<thead>
<tr>
<th>Model</th>
<th>C=C</th>
<th>C=N</th>
<th>C2=O</th>
<th>C4=O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas</td>
<td>1548</td>
<td>1585</td>
<td>1720</td>
<td>1736</td>
</tr>
<tr>
<td>PCM</td>
<td>1548</td>
<td>1586</td>
<td>1643</td>
<td>1699</td>
</tr>
<tr>
<td>N1</td>
<td>1548</td>
<td>1583</td>
<td>1696</td>
<td>1736</td>
</tr>
<tr>
<td>N3</td>
<td>1548</td>
<td>1585</td>
<td>1668</td>
<td>1734</td>
</tr>
<tr>
<td>C2=O</td>
<td>1548</td>
<td>1583</td>
<td>1680</td>
<td>1733</td>
</tr>
<tr>
<td>C4=O</td>
<td>1548</td>
<td>1584</td>
<td>1718</td>
<td>1686</td>
</tr>
<tr>
<td>N5</td>
<td>1548</td>
<td>1585</td>
<td>1705</td>
<td>1726</td>
</tr>
<tr>
<td>C7C8-t</td>
<td>1548</td>
<td>1584</td>
<td>1706</td>
<td>1734</td>
</tr>
<tr>
<td>C7C8-b</td>
<td>1548</td>
<td>1584</td>
<td>1707</td>
<td>1734</td>
</tr>
<tr>
<td>C9</td>
<td>1548</td>
<td>1584</td>
<td>1706</td>
<td>1733</td>
</tr>
<tr>
<td>2(C2=O)</td>
<td>1548</td>
<td>1581</td>
<td>1651</td>
<td>1733</td>
</tr>
<tr>
<td>2(C4=O)</td>
<td>1548</td>
<td>1585</td>
<td>1719</td>
<td>1679</td>
</tr>
<tr>
<td>6W</td>
<td>1548</td>
<td>1581</td>
<td>1637</td>
<td>1694</td>
</tr>
<tr>
<td>8W</td>
<td>1548</td>
<td>1577</td>
<td>1632</td>
<td>1692</td>
</tr>
<tr>
<td>9W</td>
<td>1548</td>
<td>1576</td>
<td>1632</td>
<td>1692</td>
</tr>
</tbody>
</table>

Table 4.5 Computed bond orders (n) for non-covalent hydrogen bonding interactions for each of the models in Figure 4.3

<table>
<thead>
<tr>
<th>Interaction</th>
<th>N1</th>
<th>C=O</th>
<th>N5</th>
<th>C=C</th>
<th>C=O</th>
<th>C2=O</th>
<th>C4=O</th>
<th>2C=O</th>
<th>2C=O</th>
<th>6W</th>
<th>8W</th>
<th>9W</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1--HOH</td>
<td>0.25</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N3--H---OH2</td>
<td>0.22</td>
<td>0.22</td>
<td>0.34</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2=O--HOH</td>
<td>0.33</td>
<td>0.32</td>
<td>0.33</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2=O--HOH'</td>
<td>0.30</td>
<td>0.30</td>
<td>0.40</td>
<td>0.39</td>
<td>0.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4=O--HOH</td>
<td>0.31</td>
<td>0.31</td>
<td>0.34</td>
<td>0.34</td>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N5--HOH</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9--OH2</td>
<td>0.25</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7-Methyl--OH2</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8-Methyl--OH2</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7-Methyl--OH2'</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8-Methyl--OH2'</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.6 Conclusion

The reaction mechanisms of many flavoproteins are still poorly understood, and FTIR DS is an important tool for obtaining detailed mechanistic information such as changes in the flavin’s intermolecular interactions with nearby amino acids. However, interpreting FTIR DS experiments would require some understanding of how hydrogen bonding, electrostatics, and other intermolecular interactions affect normal mode vibrational frequencies and intensities of molecular groups of the flavin isoalloxazine ring. To start to address this issue, we have modeled the effect of hydrogen bonding interactions on four prominent vibrational modes of a flavin model system. We find that the $\nu_{C=C}$ and $\nu_{C=N}$ mode frequencies are not sensitive to any hydrogen-bonding interactions, but their relative intensities are affected by the dielectric environment of the solvent. On the other hand, while the $\nu_{C_2=O}$ and $\nu_{C_4=O}$ mode frequencies are strongly downshifted by direct (both $\nu_{C_2=O}$ and $\nu_{C_4=O}$) and indirect (only $\nu_{C_2=O}$) hydrogen-bonding interactions, their relative intensities are less sensitive to the dielectric environment.

To reproduce the relative frequencies of the four modes in flavin, a continuum solvation model or a cluster model with multiple water molecules is needed (requiring a basis set with diffuse functions), although the continuum solvation model predicts an additional relatively intense peak that does not appear in the experiments. To reproduce the relative intensities of the FTIR bands, on the other hand, it is necessary to use either a continuum model (which is not quantitative) or QM/MM model that includes many water molecules.

Finally, we note the utility of the simple approach in Figure 4.2, which is related to the ESTM approach used in this context for the first time to explore possible flavin-water interaction geometries in an unbiased way. This automated approach may prove useful instead of the manual addition of water molecules at a few selected positions.
4.7 References


34. Zhao, R., Photochemistry, photophysics and spectroscopy of redox states of flavins relevant to photoactive flavoproteins. *2012*.


40. RT, M. *pyvdwsurface*, 2014.


5 AN ALTERNATIVE STRATEGY FOR SPECTRAL TUNING OF FLAVIN-BINDING FLUORESCENT PROTEINS

Mohammad Pabel Kabir, Daniel Ouedraogo, Yoelvis Orozco-Gonzalez, Giovanni Gadda, and Samer Gozem


Copyright © The American Chemical Society 2023

My contribution to this chapter was data curation, formal analysis, investigation, methodology, validation, visualization, writing-original draft.

5.1 Abstract

iLOV is an engineered flavin-binding fluorescent protein (FbFP) with applications for *in vivo* cellular imaging. To expand the range of applications of FbFPs for multicolor imaging and FRET-based biosensing, it is desirable to understand how to modify their absorption and emission wavelengths (i.e., through spectral tuning). There is particular interest in developing FbFPs that absorb and emit light at longer wavelengths, which has proven challenging thus far. Existing
spectral tuning strategies that do not involve chemical modification of the flavin cofactor have focused on placing positively charged amino acids near flavin’s C4a and N5 atoms. Guided by previously reported electrostatic spectral tuning maps (ESTMs) of the flavin cofactor and by quantum mechanical/molecular mechanical (QM/MM) calculations reported in this work, we suggest an alternative strategy: placing a negatively charged amino acid near flavin’s N1 atom. We predict that a single-point mutant, iLOV-Q430E, has a slightly red-shifted absorption and fluorescence maximum wavelength relative to iLOV. To validate our theoretical prediction, we experimentally expressed and purified iLOV-Q430E and measured its spectral properties. We found that the Q430E mutation in iLOV results in a slight change in absorption and a 4-8 nm redshift in the fluorescence relative to iLOV, in good agreement with the computational prediction. Molecular dynamics simulations showed that the carboxylate side chain of the glutamate in iLOV-Q430E points away from the flavin cofactor, which leads to a future expectation that further red-shifting may be achieved by bringing the side chain closer to the cofactor.

5.2 Introduction

Fluorescent proteins (FPs) have been used as tags for biosensing and bioimaging applications for over two decades in molecular virology and medicine.\textsuperscript{1-8} The most widely used FPs are derived from the green fluorescent protein (GFP), and their spectral properties have been extensively studied experimentally and computationally.\textsuperscript{8-16} However, GFPs have some limitations; they require molecular oxygen and produce hydrogen peroxide during chromophore maturation.\textsuperscript{17-19} They are also ineffective genetic tags in small viruses that cannot handle the genetic load (GFP’s molecular weight is around 22 kD). For such cases, flavin-binding fluorescent proteins (FbFPs),\textsuperscript{20}
like those derived from light, oxygen, and voltage sensing (LOV) domains, are an attractive alternative because of their smaller size (10 kD). LOV domains non-covalently bind flavin mononucleotide (FMN), which is readily available \textit{in vivo} and does not require any chemical maturation reaction.\textsuperscript{21} iLOV is a recently engineered FbFP derived from the DNA shuffling of phototropin LOV1 and LOV2 domains.\textsuperscript{21, 22} Unlike wild-type LOV domains, iLOV does not contain a cysteine residue near the FMN cofactor and is consequently unable to form the cysteinylation adduct in the excited state that initiates the LOV domain photocycle.

The absorption and fluorescence wavelengths of maximal absorbance of iLOV are around 448 nm and 500 nm, respectively. iLOV mutants with different colors could be used for multicolor bioimaging and FRET-based biosensing, with several demonstrations already in the literature.\textsuperscript{27, 31-35} However, FbFPs are notoriously tricky to tune spectrally without chemical modification of the chromophore. Despite the many mutants expressed, no experimental studies have achieved a larger than 10 nm blueshift in fluorescence emission relative to the original iLOV.\textsuperscript{36} In contrast, attempts to redshift the absorption and emission of FbFPs have met several challenges, as detailed below.

There are two main strategies currently used for spectral tuning of FbFPs. The first approach is a chemical modification of the fluorophore to modify its electronic structure and, therefore, its spectroscopic properties.\textsuperscript{37-41} However, chemically modifying FMN involves synthesizing and loading the chromophore in the protein. This is difficult to achieve \textit{in vivo}, where natural flavin derivatives are instead readily available. The second, more convenient approach would be modulating the natural FMN fluorophore's electronic energies by modifying the surrounding protein (i.e., by mutagenesis).
Several computational and experimental studies in recent years focused on the spectral tuning of iLOV through protein point mutations, with recent attempts primarily focused on attempting to redshift the absorption. Khrenova and coworkers first recognized that placing a positive charge near flavin’s N5 and C4a atoms would redshift iLOV’s absorption and emission wavelength (see Figure 5.1 for flavin atom labels). They proposed a Q489K single-point mutation, reasoning that the positively charged amino group (Lys) near flavin’s N5 would stabilize its excited state π-electron system more than in the ground state, resulting in a redshift. QM/MM calculations supported their hypothesis. However, Davari and coworkers computationally and experimentally showed that the Q489K lysine side chain flips away from the N5 and C4a atoms of the chromophore, resulting in a blueshift in the absorption and emission instead of a redshift. In a follow-up QM/MM study, Khrenova et al. proposed additional mutations to stabilize the lysine side chain close to the N5 and C4a atoms of the chromophore. Recently, Wehler and coworkers experimentally attempted to prepare these mutants but could not prepare a functional red-shifted FbFP, as they found that a double-point mutant (iLOV-L470T/Q489K) gives ~2 nm blueshift and a triple-point mutation (iLOV-V392K/F410V/A426S) lost the ability to bind the chromophore due to the V392K mutation. Overall, while the strategy of placing a positive charge in the vicinity of flavin’s N5 and C4a atoms was theoretically shown to work, most of the amino acids on that side of the protein turned out to be conformationally unstable or essential for chromophore binding. However, recently, Röllen and coworkers prepared a red-shifted iLOV with a double point mutation (iLOV-V392T-Q489K) and got no shift in absorption and 6 nm redshift in emission. Red-shifted FbFPs were also derived from the thermostable protein CagFbFP, the maximum redshift managed to get 3 nm in absorption and 7 nm in emission for CagFbFP-Q148K/I52T.
Our group recently reported electrostatic spectral tuning maps (ESTMs)\textsuperscript{47, 48} and flavin-solvent hydrogen bonding interactions.\textsuperscript{49} Those serve as a starting point to find suitable mutations for spectral tuning. Here, based on these ESTMs, we suggest an alternative mutagenesis approach to redshift the absorption of iLOV; instead of focusing on placing a positive charge near C4a or N5, we introduce a negatively charged amino acid in the vicinity of flavin’s N1 atom through a Q430E single-point mutant. We first test the effect of this mutation using hybrid QM/MM calculations employing the average solvent electrostatic configuration (ASEC) free energy gradient (FEG) approach. We then express iLOV-Q430E in the lab and measure its absorption, excitation, and fluorescence spectra to verify if they are red-shifted relative to iLOV.

![Figure 5.1](image)

**Figure 5.1** The isalloxazine ring of FMN and atom number labels. R=CH\textsubscript{3} for lumiflavin and R=ribose-5′-phosphate for FMN.

5.3 Methods

5.3.1 Computational approach

The approach to generate ESTMs has already been documented elsewhere,\textsuperscript{47, 48} and will be briefly summarized in the Results and Discussion section. Here, we focus on the details of the ASEC-FEG calculations. The ASEC-FEG method builds on the average solvent electrostatic
configuration (ASEC) approach developed by Canuto and coworkers\textsuperscript{50} and approximates the FEG approach from Okuyama-Yoshida et al.,\textsuperscript{51} which is rooted in the free energy perturbation theory.\textsuperscript{52} The ASEC-FEG approach was first extended to proteins by Orozco-Gonzalez et al. for rhodopsins.\textsuperscript{53} We recently developed ASEC-FEG for flavoproteins.\textsuperscript{54,55} With ASEC, the quantum chemical calculations are performed in the field of a time-averaged electrostatic potential environment of the protein and solution (collectively referred to as a “solvent” in the ASEC acronym). Effectively, the protein and solution are represented as a “superposition” of structures obtained from MD simulations. This approach leaves the representation of rigid atoms intact while flexible atoms are replaced by a cloud of charges over the space sampled during the dynamics. Flexible atoms also have a broader and shallower Lennard Jones potential (see ref. \textsuperscript{54} for more details). The optimization of the QM system within the ASEC configuration is done self-consistently. ASEC is well suited for averaging the effect of long-range charge interactions that are difficult to capture with more traditional QM/MM methods.\textsuperscript{54}

The ASEC-FEG protocol is a series of scripts building on an existing QM/MM interface between the OpenMolcas\textsuperscript{56} and Tinker\textsuperscript{57} software packages.\textsuperscript{58,59} The protocol guides users through the model construction starting from the PDB file of the protein and culminating in the generation of an ASEC QM/MM model (see Figure 5.2). The protocol calls on several other software: PropKa 3.1\textsuperscript{60}, Dowser\textsuperscript{61}, SCWRL4.0\textsuperscript{62}, and Gromacs.\textsuperscript{63} Gromacs is used to add hydrogen atoms to the PDB, solvate the protein, and run MD simulations to equilibrate the system and sample the protein around the cofactor. The OpenMolcas\textsuperscript{56}/Tinker\textsuperscript{57} interface uses an additive QM/MM scheme that includes Lennard Jones and electrostatic interactions through the ElectroStatic Potential Fitted (ESPF) approach.\textsuperscript{64} The automation of this protocol, done in the same vein as efforts to automate the construction of QM/MM models for rhodopsins by Olivucci and coworkers,\textsuperscript{65-67} mitigates
problems with reproducibility of QM/MM calculations and allows the systematic investigation of closely related proteins using a consistent approach.

**Figure 5.2** Steps in the automated ASEC-FEG protocol for flavoproteins. Ovals with solid borders indicate required user input, while ovals with dashed borders indicate optional user input.

The initial coordinates of iLOV were taken from the X-ray structure PDB 4EES (resolution: 1.8 Å). Parameters for FMN were initially retrieved from the AMBER parameter database maintained by the University of Manchester. The Q430E mutation was introduced in the protein structure and modeled using SCWRL4. Dowser was used to remove non-bonded water molecules from the crystallographic structure. The total charge of the systems was neutralized by adding solution counterions. This model was used as a starting point for MD simulations and the generation of the ASEC environment. The MD calculations for iLOV and iLOV-Q430E were
performed with periodic boundary conditions in a 7.0 nm x 7.0 nm x 7.0 nm cubic solvent box. Geometry minimization and MD simulations were carried out using GROMACS. The AMBER99SB and TIP3P force fields were used for protein and water, respectively. During each step of the ASEC-FEG cycle, the MD calculations were performed in three phases: the system was first gradually heated from 0 to 300 K at 1 atm pressure over 300 ps. This was followed by 4700 ps of equilibration and 5000 ps of production simulations carried out with the NPT ensemble under standard ambient temperature and pressure. The ASEC configuration of the protein was formed by sampling 100 configurations at 50 ps time interval from the production part of the MD. The ASEC GROMACS file is then converted to Tinker format for QM/MM calculations.

For QM/MM calculations, the protein was divided into two subsystems (Figure 5.3): (i) the QM region, comprising the lumiflavin (structure shown in Figure 5.1), and (ii) the MM region, which includes all other atoms in the simulation (the ribose-5’-phosphate group, the protein, the solvent, and solution ions). The frontier between the QM and the MM parts is treated using the hydrogen link atom (LA, Figure 5.3). The charges for the MM atoms near the LA are set to zero and distributed over other MM atoms to avoid over polarizing the QM wavefunction. The QM subsystem is then optimized in the presence of a frozen ASEC MM environment with electrostatic embedding. Using the updated geometry and updated ESPF charges of the QM subsystem, another MD calculation is run for 5 ns to generate a new ASEC configuration. This process is repeated for several steps until the computed excitation energies stay consistent for four consecutive steps (i.e., within 0.02 eV nm of the four-step moving average). We then took the average excitation energies from those four steps and used them to compute the wavelength shift.
Figure 5.3 FMN inside the binding pocket of iLOV from PDB 4EES. The dashed lines indicate the distance in Angstroms between the Q489 and Q430 glutamine side chain nitrogen atoms and the closest FMN nitrogen atom. The red circle indicates the hydrogen Link Atom (LA), which separates the QM subsystem (the lumiflavin) and the MM subsystem (the ribose-5'-phosphate group, protein, and solvent). The figure was prepared using PyMol. The geometry optimization of the QM subsystem was performed using the complete-active-space self-consistent field (CASSCF) level of theory and the ANO-L-VDZP basis set. We tested the effect of the active space on the first excited state (π-π*) excitation energy of flavin using gas-phase benchmark calculations shown in Figure 5.10; we found that there is a limited benefit to increasing the active space beyond 10 electrons and 10 orbitals (five π, five π*). The π_N and π_CO -orbitals has an insignificant effect on the excitation energies. Therefore, QM/MM geometry optimizations were performed using CASSCF (10,10). State averaging was not used for ground state optimizations, while 2-root state averaging was used for excited-state optimizations. Excitation energies were computed using the complete active space second-order perturbation
theory (CASPT2) with the ANO-L-VDZP basis set. CASPT2 calculations were performed using the Cholesky decomposition\textsuperscript{73} and applying an imaginary level shift\textsuperscript{74} of 0.2. An IPEA shift,\textsuperscript{75} sometimes used for flavins for the purpose of error cancellation,\textsuperscript{76, 77} was not used here. It was recently shown that the IPEA shift is not needed in cases where the dynamical electron correlation is adequately accounted for in CASPT2 calculations.\textsuperscript{78} We also note that vertical excitation energy calculations typically underestimate the absorption wavelength of flavins compared to the experimental wavelength of maximal absorption; the calculation of the vibronic progression from Franck-Condon factors is needed for better quantitative agreement between theory and experimental spectra.\textsuperscript{48, 79-81}

### 5.3.2 Bacterial strains and plasmids

The iLOV and iLOV-Q430E were obtained as synthetic genes from GenScript (Inc., Piscataway, NJ, USA) and were flanked with a 5’- NdeI and a 3’- XhoI restriction endonuclease recognition sites in a pET20b(+) plasmid. The pET plasmid harboring the iLOV and iLOV-Q430E genes contains an N-terminal His\textsubscript{6}-tag fused to the target proteins to facilitate heterologous expression in \textit{E. coli} and affinity chromatography purification. The genes were transformed into \textit{E. coli} strain DH5\textalpha{} and Rosetta(DE3)pLysS competent cells for storage and expression, respectively. The resulting plasmids were verified by sequencing (Psmagen, Inc., Rockville, MD, USA), and permanent stocks of the cells were prepared and stored at -80 °C.

### 5.3.3 Protein expression and purification

Permanently frozen stocks of \textit{E. coli} cells Rosetta(DE3)pLysS harboring iLOV or iLOV-Q430E genes were used to inoculate 100 mL of Luria-Bertani broth medium containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol, and cultures were grown at 37 °C overnight to be used as a preculture. A 10 mL portion of preculture was used to inoculate 1.0 L of Luria-Bertani broth.
medium containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol. When the cultures reached optical densities of ~0.6 at 600 nm, the temperature was lowered to 18 °C, and Isopropyl-thio-galactoside (IPTG) was added to a final concentration of 0.1 mM. After 18 h, the cells were harvested by centrifugation at 5000g for 20 min at 4°C. All purification steps were carried out at 4 °C. The wet cell paste was suspended in 0.1 mM PMSF, 0.2 mg/mL lysozyme, 10 % glycerol, and 50 mM pH 8.0 phosphate buffer solution, containing 300 mM NaCl, 10 mM Imidazole, and 10 % glycerol in a ratio of 1 g of wet cell paste to 4 mL of lysis buffer. The suspended cells were then allowed to incubate with stirring for 30 min on ice with 5 μg/mL RNase and 5 μg/mL DNase in the presence of 10 mM MgCl₂. The resulting slurry was sonicated for 60 cycles of 20 seconds with the pulse on and 10 seconds with the pulse off for 20 min. The cell debris was removed by centrifugation at 10000 x g for 20 min. The supernatant was loaded onto a 5 mL Ni-NTA column (GE Healthcare), equilibrated with 50 mM pH 8.0 phosphate buffer solution, 300 mM NaCl, 10 mM Imidazole, and 10 % glycerol. The proteins were purified with gradient elution from 10 to 250 mM imidazole in 50 mM pH 8.0 phosphate buffer solution, 300 mM NaCl, and 10 % glycerol buffer. The eluted fractions containing the iLOV protein were dialyzed against four changes of 10 mM pH 8.0 phosphate buffer solution, 10 mM NaCl, and 10 % glycerol. After the dialysis, the proteins were centrifuged at 10000 x g for 20 min to remove any precipitated protein. The iLOV and the iLOV-Q430E were then stored at -20 °C.

5.3.4 Flavin reconstitution

After column chromatography, the iLOV-Q430E variant protein was devoid of bound FMN cofactor. The variant protein was incubated with excess free FMN to load the FMN cofactor to the protein. The incubation was carried out at 4 °C overnight. The excess FMN was removed
with Amicon Ultra Centrifugal filters. The free FMN was extracted from the FMN-dependent *pseudomonas aeruginosa* nitronate monooxygenase (PaNMO) variant, H183F.

### 5.3.5 UV-Visible Absorption and Fluorescence Spectroscopy

The UV-visible absorption spectra of iLOV and iLOV-Q430E were recorded with an Agilent Technologies model HP 8453 PC diode-array spectrophotometer equipped with a thermostated water bath. The proteins were prepared fresh by gel filtration through PD-10 desalting columns (General Electric, Fairfield, CT) just before being used. The extinction coefficients of the enzyme-bound FMN to the iLOV protein were determined in 20 mM pH 7.0 phosphate buffer solution after incubation of the protein with 4 M urea at 40 °C for 1 h, based upon an $\varepsilon_{450}$ value of 12.2 mM$^{-1}$ cm$^{-1}$ for free FMN and the method published by Whitby et al.$^{82}$ The fluorescence emission spectra of the iLOV and iLOV-Q430E variant protein were recorded in 20 mM pH 8.0 phosphate buffer solution at 15 °C, with a Shimadzu model RF-5301 PC spectrofluorometer using a 1 cm path length quartz cuvette. All fluorescence spectra were corrected by subtracting the corresponding blanks to account for Rayleigh and Raman scattering. The samples at a concentration of 10 μM protein-bound flavin were excited at the low-energy peak of the UV-visible absorption spectrum, and emission scans were determined from 475 to 600 nm.

### 5.4 Results and Discussion
Figure 5.4 A. An ESTM map reproduced from ref. 48. The map indicates the change in the vertical excitation energy between the ground state ($S_0$) and the first singlet excited state ($S_1$) introduced by a $+0.1$-probe charge placed at the van der Waals surface of lumiflavin in its $S_0$ equilibrium geometry. The map suggests mutations that redshift the excitation energy (see arrows and labels). The legend indicates the magnitude of the excitation energy shifts relative to the gas-phase reference excitation energy in eV (and in nm in parentheses). B. The same ESTM map was computed at the $S_1$-optimized lumiflavin geometry, corresponding to the fluorescent minimum. The legend indicates the magnitude of the emission energy shifts relative to the gas-phase reference emission energy in eV (and in nm in parentheses).

5.4.1 ESTMs and charge analysis

ESTMs for flavin were reported recently.47, 48 These maps are intuitive visual tools that indicate how external positive or negative charges in the vicinity of a molecule influence its absorption or emission spectra. Briefly, ESTMs are constructed by moving a point charge on the van der Waals surface of the molecule and calculating the change in excitation and emission energies. The ESTM for the first singlet excited state ($S_1$) of flavin, which is experimentally at $\sim$448 nm in iLOV, is shown in Figure 5.4A. Here, we also recomputed the ESTM at the excited-state optimized flavin geometry to map how the fluorescence energy, which is experimentally at $\sim$500 nm in iLOV, is
modified by nearby point charges (Figure 5.4B). Both ESTMs were computed using time-dependent-density functional theory (TD-DFT) with the B3LYP functional\textsuperscript{83} and cc-pVTZ basis set.\textsuperscript{84}

The red and blue colored regions of the ESTMs in Figure 5.4 indicate that a spectral shift can be achieved if there is an electrostatic potential change in those regions. Both Figures 5.4A and 5.4B have similar features: a red region near C4a and N5 atoms, a blue region near the N1 atom, and a white region near the xylene portion of flavin. This indicates that (a) a positive charge near the C4a or N5 atoms would redshift the absorption/emission, (b) a positive charge near N1 would blueshift the absorption/emission, and (c) charges introduced near the xylene portion of the flavin would have a negligible effect on the spectral properties for this state. A negative probe charge would have the exact opposite effect.\textsuperscript{47} These calculations are largely consistent with transition dipole moment measurements in flavins.\textsuperscript{85, 86} While the excitation energy ESTM (Figure 5.4A) and emission ESTM (Figure 5.4B) appear very similar, the magnitude of the shift reported in the plot legends reveals a subtle difference: the emission energy (in eV) is more sensitive to the presence of negative charges near the N1 flavin atom than the corresponding absorption energy.

The ESTMs in Figure 5.4 were used to generate strategies for spectral tuning lumiflavin. The maps indicate that a positive charge near the C4a or N5 flavin atoms would lead to a red-shifted absorption/emission; this has been the strategy proposed by Khrenova et al. and the ensuing computational and experimental work.\textsuperscript{23, 41-45} A second strategy, not yet explored in FbFPs, would be to introduce a negatively charged amino acid or negative part of a dipole in the vicinity of the N1 atom of flavin. Here, we pursue this strategy with Q430E.
**Figure 5.5 Center:** LoProp charge population analysis (with hydrogen atom charges summed onto the heavy atoms they are connected to) for lumiflavin in its ground (S\(_0\)) and first singlet excited (S\(_1\)) state. Red circles indicate negative charge density, and blue circles indicate positive charge density. The area of the circles is directly proportional to the charge on the corresponding atom. The S\(_1\) and S\(_0\) states have slightly different charge distributions that are difficult to discern without close inspection. Therefore, the difference in the atomic charges (S\(_1\)–S\(_0\)) is also shown in the middle. In the S\(_1\)–S\(_0\) difference plot, the areas of the circles are proportional to the magnitude of the charge difference between S\(_0\) and S\(_1\), with red circles indicating reduced charge (higher electron density) on the atom after excitation from S\(_0\) to S\(_1\) and blue circles indicating increased charge (lower electron density). **Left:** A scheme illustrating the effect of placing a positively charged lysine side chain close to the flavin N5/C4a, which would result in a redshift (see text for details). **Right:** A scheme illustrating the effect of placing a negatively charged glutamate side chain close to the flavin N1, which should also result in a redshift (see text for details).
To better understand the electronic structure changes underlying the electrostatic spectral tuning properties of flavin, we computed atomic charges from both the ground and excited-state wave functions of a lumiflavin gas-phase model (Figure 5.5). The charges were obtained using LoProp population analysis from CASSCF(10,10)/ANO-L-VDZP wave functions. The LoProp approach provides physically meaningful localized properties and mitigates issues like basis set dependence sometimes encountered with other population analysis methods.87

Exciting flavin from the $S_0$ to the $S_1$ state changes the electron distribution along the $\pi$-conjugated isoalloxazine ring. This effect is subtle but can be best visualized by plotting the change in the charges at each atomic center from $S_0$ to $S_1$ (Figure 5.5 center). Specifically, there is an increase in electron density at the C4a and N5 flavin atoms and a decrease in electron density at the N1 flavin atom and several atoms in the xylene portion of the flavin. There is also a slight decrease in electron density at the C2=O carbonyl. The charge redistribution is perfectly consistent with the ESTMs in Figure 5.4; the increased charge density at the C4a/N5 flavin atoms means there is potential for spectral tuning by placing a positive charge nearby. Such a positive charge, e.g., a protonated lysine side chain, would stabilize the excited $S_1$ state slightly more than the ground $S_0$ state, leading to a redshift in the excitation energy (Figure 5.5 left). Conversely, there are several atoms where the electron density decreases upon excitation to $S_1$. In most cases, those atoms are shielded from external charges by methyl groups or hydrogen atoms, which explains the less intense color of the ESTM map near the C7, C9, C9a, and N10 atoms despite the decrease in the electron density on those atoms. However, the N1 atom is exposed, allowing charged amino acids to approach and creating an opportunity for spectral tuning at that site. The decreased electron density on the N1 flavin atom means that a negative charge nearby, e.g., a glutamate side chain,
would destabilize the ground state more than the excited state, decreasing the $S_0-S_1$ energy gap and resulting in a red-shifted absorption (Figure 5.5 right).

The approach shown on the right of Figure 5.5 (placing a negatively charged amino acid near flavin’s N1 atom) seems less desirable than the approach on the left (placing a positively charged amino acid near flavin’s C4a/N5 atoms) from a bioengineering standpoint, since it relies on an unfavorable interaction between flavin and a negatively charged residue. However, given the limited success with spectral tuning at the C4a/N5 site, we attempted spectral tuning with a negatively charged residue with the Q430E single-point mutation.

5.4.2 QM/MM simulations of iLOV and iLOV-Q430E

The crystal structure of iLOV indicates that Q430 is 3.4 Å away from the N1 atom of FMN (Figure 5.3). Therefore, we chose to replace Q430 with isosteric glutamic acid. Since glutamic acid has a $pK_a$ of 4.07 in solution and Q430 has polar residues nearby, we anticipated that the mutated glutamic acid Q430E would be deprotonated and introduce a negative charge near the flavin’s N1 atom without causing a significant structural change in the protein. To test this hypothesis, we performed QM/MM geometry optimizations followed by excited-state energy calculations for both iLOV and iLOV-Q430E using the ASEC-FEG method, as outlined in the Methods section.
**Figure 5.6** This figure shows three representatives QM/MM optimized snapshots to show the distance between the flavin N1 and the Q430 (iLOV) or E430 (iLOV-Q430E) side chain. The labeled distances are reported averages from one of the production runs of the MD simulations. Specifically, we label the distance between the flavin N1 and the Q430 side chain nitrogen in iLOV (3.5 Å), the E430 side chain carboxylate carbon in the ground-state optimized system (E430$_{\text{abs}}$, 8.9 Å), and the E430 side chain carboxylate carbon in the excited-state optimized system (E430$_{\text{fl}}$, 8.4 Å). The figure was prepared using PyMol.$^{72}$

In snapshots obtained from molecular dynamics (MD) simulations of iLOV, the average computed distance of the Q430 nitrogen from the flavin N1 atom (3.5 Å, **Figure 5.6**), is in good agreement with the crystal structure (3.4 Å, **Figure 5.3**). In contrast, MD simulations of iLOV-Q430E revealed that the E430 glutamate side chain flips away from flavin and maintains an average distance of 8.9 Å from the flavin N1 atom (E430$_{\text{abs}}$ in **Figure 5.6**). This conformational change is likely driven by a lack of hydrogen bonding with neighboring amino acids, which causes E430 to point outwards towards the surface of the protein.

Next, we proceeded to calculate the vertical excitation energy. The iLOV and iLOV-Q430E calculations reached self-consistency quickly after just one step of the ASEC-FEG cycle. After that, the following four steps, each involving MD to regenerate the ASEC environment, re-optimizing the flavin chromophore, and re-computing the vertical excitation energy, yielded similar results within 1 nm of each other. These calculations indicated that the vertical excitation energy of iLOV-Q430E is 4 nm red-shifted compared to iLOV (**Table 5.2**).

In a recent joint computational and experimental study,$^{54}$ we found that solution ions may affect the outcome of ASEC-FEG calculations, especially when there are charged residues inside the
active site of a flavoprotein. Therefore, we repeated the calculations for both iLOV and iLOV-Q430E after adding 4 pairs of Na+ and Cl- solution ions, approximately 1 NaCl per 2775 water solvent molecules, equivalent to 20 mM of salt used in the experiments in this work. These calculations showed that the Q430E mutation has almost no effect on the absorption spectrum of iLOV, causing a shift of less than 0.5 nm (Table 5.3).

The MD and ASEC-FEG simulations show that the E430 does not remain in the same position as Q430 and causes just a slight 0-4 nm redshift in the vertical excitation energy relative to iLOV, which was initially discouraging. However, given the difficulty in red-shifting iLOV even by a few nm, we decided to proceed with expressing iLOV-Q430E and iLOV experimentally to compare their absorption and emission properties.

### 5.4.3 Experimental absorption and emission spectra of iLOV and iLOV-Q430E

iLOV and iLOV-Q430E were expressed and purified successfully. The absorption and excitation/emission spectra were measured and are shown in Figure 5.7 and Figure 5.8, respectively. The data are also tabulated in Table 5.1. We found that iLOV-Q430E gave a modest 1-2 nm redshift in the first excited state absorption wavelength, consistent with our computational prediction of a weak redshift. However, the fluorescence wavelength is shifted to red by 4-8 nm compared to the emission in iLOV. This is comparable to the shift recently achieved by Röllen and coworkers, where they have prepared double point mutation for iLOV and thermostable dimer protein CagFbFP and found a 6 nm redshift in emission but no shift in absorption for iLOV-V392T-Q489K while for CagFbFP-Q148K-I52T they got 3 nm in absorption and 7 nm in emission.23
Figure 5.7 UV-visible absorption spectra of iLOV (panel A, blue), iLOV-Q430E (panel B, red), and free FMN (Panel C, black). The spectra were recorded in 20 mM pH 7.0 phosphate buffer solution and 15 °C.

Figure 5.8 The excitation and emission spectra of iLOV (panel A, blue), iLOV-Q430E (panel B, red), and free FMN (panel C, black). The spectra were recorded in 20 mM pH 7.0 phosphate buffer and 15 °C.
Figure 5.9 A. Difference excitation spectra of iLOV and iLOV-Q430E minus that of free FMN.

B. Difference emission spectra of the iLOV and iLOV-Q430E minus that of free FMN. The reference line in panels A and B represents the ΔEX (free FMN-Free FMN) and ΔEM (free FMN-Free FMN), respectively.

Table 5.1 Experimental absorption and fluorescence properties of iLOV proteins.

<table>
<thead>
<tr>
<th></th>
<th>iLOV</th>
<th>iLOV-Q430E</th>
<th>Free FMN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximal absorbance, nm</td>
<td>370, 448</td>
<td>366, 450</td>
<td>372, 445</td>
</tr>
<tr>
<td>Extinction coefficient, mM⁻¹cm⁻¹</td>
<td>11.9, 15.6</td>
<td>11.8, 12.8</td>
<td>10.0, 12.5</td>
</tr>
<tr>
<td>Fluorescence emission, nm</td>
<td>500 ± 1, 526 ± 1 c</td>
<td>508 ± 1, 530 ± 1</td>
<td>530 ± 1</td>
</tr>
<tr>
<td>Fluorescence intensity</td>
<td>113 ± 1, 86 ± 1 c</td>
<td>78 ± 1, 74 ± 1</td>
<td>61 ± 1</td>
</tr>
</tbody>
</table>

UV-visible and fluorescence data were recorded in 20 mM pH 8.0 phosphate buffer solution and 15 °C.

b Excited at the low energy peak of the flavin.

c Standard errors refer to the average of three independent measurements.
Next, we investigated the effect of the Q430E mutation on the fluorescence brightness by measuring the relative absorbance and fluorescence intensity of iLOV and iLOV-Q430E at the same concentration (Figure 5.9). We found that the absorbance and fluorescence intensities in iLOV-Q430E are reduced to about 50% compared to iLOV.

The experimental spectra indicate that the fluorescence wavelength is more sensitive to the Q430E point mutation than the absorption. This may be partly explained by comparing the excitation and emission ESTMs in Figure 5.4A and Figure 5.4B, respectively. The emission ESTM indicates a higher sensitivity of the emission energy compared to the absorption. However, to check if we can reproduce this effect in the QM/MM calculations, we repeated the ASEC-FEG calculations for both iLOV and iLOV-Q430E using the charges and gradient of the $S_1$ excited state of flavin instead of the $S_0$ ground state. We computed the $S_1 - S_0$ vertical emission energy at the excited state $S_1$ geometry. Note that, due to the use of excited-state charges for flavin during the MD calculations, the protein adapts to the excited-state charge distribution. The 5 ns MD calculations have a similar timescale as a typical fluorescence lifetime, giving the protein a reasonable time to rearrange around the excited state configuration. In this case, it took slightly longer to achieve self-consistency of the ASEC-FEG calculations, so the first three steps were discarded, and the excitation energy was averaged over the next four ASEC-FEG steps. The ASEC-FEG calculations indicate that iLOV-Q430E has a 9 nm red-shifted vertical emission wavelength compared to iLOV (Table 5.4). Inspecting the MD simulations revealed that, on average, the E430 moves closer to the flavin in the excited state compared to the ground state. The average distance between flavin’s N1 atom and the C-atom of the side chain carboxylate ion is 8.4 Å for the excited state, compared to 8.9 Å in the ground state geometry (E430$^{\text{abs}}$ and E430$^{\text{fl}}$ in Figure 5.6). Therefore, the more significant Stokes shift in iLOV-Q430E compared to iLOV can be attributed to two factors; one is
electronic since the ESTM already shows a higher sensitivity of flavin’s $S_0$–$S_1$ energy difference to charges near the N1 atom after flavin relaxes on its $S_1$ potential energy surface to the fluorescent minimum (Figure 5.4). The second effect comes from the protein; the deprotonated E430 is less repelled by the flavin excited state than the ground state and moves slightly closer, on average, to the flavin chromophore after it is excited to $S_1$.

**Table 5.2** CASPT2//CASSCF/ANO-L-VDZP computed ground ($S_0$) and first singlet excited ($S_1$) state energies. Each column refers to a step of the ASEC-FEG protocol. In addition to absolute energies (in Hartree), the $S_0$–$S_1$ vertical excitation energies (in eV and nm) are also shown, as well as the shift (in nm) caused by the Q430E mutation.

<table>
<thead>
<tr>
<th></th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Step 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>iLOV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_0$ energy</td>
<td>-870.7066891</td>
<td>-870.7208093</td>
<td>-870.7258861</td>
<td>-870.7170955</td>
<td></td>
</tr>
<tr>
<td>(Hartree)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_1$ energy</td>
<td>-870.5992511</td>
<td>-870.6136004</td>
<td>-870.6183912</td>
<td>-870.6093119</td>
<td></td>
</tr>
<tr>
<td>(Hartree)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta E_{S_1:S_0}$ (eV and nm)</td>
<td>2.924 eV (424 nm)</td>
<td>2.917 eV (425 nm)</td>
<td>2.925 eV (424 nm)</td>
<td>2.933 eV (423 nm)</td>
<td>2.925 eV (424 nm)</td>
</tr>
<tr>
<td><strong>iLOV-Q430E</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_0$ energy</td>
<td>-870.6976046</td>
<td>-870.6984876</td>
<td>-870.6965667</td>
<td>-870.6927712</td>
<td></td>
</tr>
<tr>
<td>(Hartree)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_1$ energy</td>
<td>-870.5911651</td>
<td>-870.5919402</td>
<td>-870.590109</td>
<td>-870.5862978</td>
<td></td>
</tr>
<tr>
<td>(Hartree)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta E_{S_1:S_0}$ (eV and nm)</td>
<td>2.896 eV (428 nm)</td>
<td>2.899 eV (428 nm)</td>
<td>2.897 eV (428 nm)</td>
<td>2.897 eV (428 nm)</td>
<td>2.897 eV (428 nm)</td>
</tr>
<tr>
<td><strong>Shift caused</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>by the Q430E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>mutation</strong></td>
<td>$\Delta E_{S_1:S_0}$ (nm)</td>
<td></td>
<td></td>
<td></td>
<td>4 nm</td>
</tr>
</tbody>
</table>
**Table 5.3** CASPT2//CASSCF/ANO-L-VDZP computed ground ($S_0$) and first singlet excited ($S_1$) state energies in the presence of 4 additional Na$^+$ and Cl$^-$ ions. Each column refers to a step of the ASEC-FEG protocol. In addition to absolute energies (in Hartree), the $S_0$-$S_1$ vertical excitation energies (in eV and nm) are also shown, as well as the shift (in nm) caused by the Q430E mutation.

<table>
<thead>
<tr>
<th></th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Step 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>iLOV</td>
<td>$S_0$ energy (Hartree)</td>
<td>-870.7303668</td>
<td>-870.7269008</td>
<td>-870.7253756</td>
<td>-870.708917</td>
</tr>
<tr>
<td></td>
<td>$S_1$ energy (Hartree)</td>
<td>-870.623182</td>
<td>-870.6190768</td>
<td>-870.6190198</td>
<td>-870.602247</td>
</tr>
<tr>
<td></td>
<td>$\Delta E_{S_1:S_0}$ (eV and nm)</td>
<td>2.917 eV (425 nm)</td>
<td>2.934 eV (423 nm)</td>
<td>2.894 eV (428 nm)</td>
<td>2.903 eV (427 nm)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.912 eV (426 nm)</td>
</tr>
<tr>
<td>iLOV-Q430E</td>
<td>$S_0$ energy (Hartree)</td>
<td>-870.7099589</td>
<td>-870.7178432</td>
<td>-870.719662</td>
<td>-870.6988243</td>
</tr>
<tr>
<td></td>
<td>$S_1$ energy (Hartree)</td>
<td>-870.60197</td>
<td>-870.6106224</td>
<td>-870.605017</td>
<td>-870.5911968</td>
</tr>
<tr>
<td></td>
<td>$\Delta E_{S_1:S_0}$ (eV and nm)</td>
<td>2.905 eV (427 nm)</td>
<td>2.918 eV (425 nm)</td>
<td>2.910 eV (426 nm)</td>
<td>2.929 eV (423 nm)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.915 eV (425 nm)</td>
</tr>
<tr>
<td></td>
<td>Shift caused by the Q430E mutation $\Delta \Delta E_{S_1:S_0}$ (nm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.4** CASPT2//CASSCF/ANO-L-VDZP computed ground ($S_0$) and first singlet excited ($S_1$) state energies at the $S_1$ minimum geometry. Each column refers to a step of the ASEC-FEG protocol. Note that the calculations needed 3 steps to reach self-consistency, so results from Steps 1-3 were discarded. In addition to absolute energies (in Hartree), the $S_0$-$S_1$ vertical excitation energies (in eV and nm) are also shown, as well as the shift (in nm) caused by the Q430E mutation.

<table>
<thead>
<tr>
<th></th>
<th>Step 4</th>
<th>Step 5</th>
<th>Step 6</th>
<th>Step 7</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>iLOV</td>
<td>$S_0$ energy (Hartree)</td>
<td>-870.7135153</td>
<td>-870.7162409</td>
<td>-870.7212094</td>
<td>-870.7294337</td>
</tr>
<tr>
<td></td>
<td>$S_1$ energy (Hartree)</td>
<td>-870.6324525</td>
<td>-870.6344888</td>
<td>-870.6396116</td>
<td>-870.6477739</td>
</tr>
<tr>
<td></td>
<td>$\Delta E_{S_1:S_0}$ (eV and nm)</td>
<td>2.206 eV (562 nm)</td>
<td>2.225 eV (557 nm)</td>
<td>2.220 eV (558 nm)</td>
<td>2.222 eV (558 nm)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.218 eV (559 nm)</td>
</tr>
<tr>
<td>iLOV-Q430E</td>
<td>$S_0$ energy (Hartree)</td>
<td>$S_1$ energy (Hartree)</td>
<td>$\Delta E_{S_1-S_0}$ (eV)</td>
<td>$\Delta E_{S_1-S_0}$ (nm)</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-870.6925643</td>
<td>-870.6124873</td>
<td>2.179 eV</td>
<td>2.185 eV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-870.6866233</td>
<td>-870.6063316</td>
<td>2.186 eV</td>
<td>2.185 eV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-870.7039868</td>
<td>-870.6236473</td>
<td>2.187 eV</td>
<td>2.186 eV</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-870.6853863</td>
<td>-870.6050131</td>
<td>2.184 eV</td>
<td>2.184 eV</td>
<td></td>
</tr>
</tbody>
</table>

Shift caused by the Q430E mutation $\Delta E_{S_1-S_0}$ (nm)

9 nm

**Figure 5.10** A plot of first excited state excitation energy vs active space for B3LYP and CASPT2 methods. The excitation energies were calculated from MP2-optimized geometry with OpenMolcas and Gaussian software.

### 5.5 Conclusion

Starting from simple visual guides (ESTMs), we proposed a red-shifting mutant of a flavin-binding fluorescent protein, iLOV. This prediction was further tested using QM/MM ASEC-FEG calculations and, ultimately, through mutagenesis and spectroscopy experiments that confirmed that the intended redshift did occur. The strategy used here, placing a negatively charged residue near flavin’s N1 atom, is an alternative to the more widely attempted and studied approach of
putting a positive charge near flavin’s C4a and N5 atoms. We note that the two strategies are not mutually exclusive and may be combined to potentially achieve a further redshift of FbFPs. The calculations also indicate that further redshift may be possible by introducing a negatively charged side chain closer to the flavin N1 than in iLOV-Q430E. This could be achieved by engineering double or triple mutants that stabilize the negatively charged E430 near flavin’s N1 atom, as for Q489K near C4a/N5 atoms.

This work also demonstrates how computational tools and experiments can synergistically achieve a certain desired protein engineering goal. There have been multiple attempts to redshift the absorption spectrum of iLOV over the past decade. Early screening experimental studies generated tens of mutations but did not achieve the desired redshift without aid from rational design. Computational studies subsequently provided valuable insight into a strategy for how to redshift the absorption wavelength of iLOV; however, computations may miss nuances associated with point mutations that experiments can reveal. It was only through an iterative computational and experimental process that a redshift was ultimately achieved in iLOV and CagFbFP.23, 46 In this study, we first employed simple computational tools like ESTMs as “hypothesis generators.” We then constructed more realistic QM/MM ASEC-FEG calculations to model the proposed system and study its dynamics and spectral properties. We finally carried out the experiments to verify the results of the calculations. Conversely, the experiments often bring up new observations and questions for the calculations to answer, as was the case here for the more significant Stokes shift observed in iLOV-Q430E compared to iLOV.
5.6 References


33. Shcherbakova, D. M.; Shemetov, A. A.; Kaberniuk, A. A.; Verkhusha, V. V., Natural photoreceptors as a source of fluorescent proteins, biosensors, and optogenetic tools. *Annual review of biochemistry* 2015, 84, 519.


5.7 References


6 EXCITED-STATE BENCHMARK STUDIES FOR FLAVIN COFACTOR IN DIFFERENT REDOX AND PROTONATION STATES

6.1 Abstract

Flavoproteins play vital roles in various biological processes, including in light sensing and response. The flavin cofactor acts as the chromophore in light response. The photochemical reactions of flavin vary depending on the type of flavoprotein. To gain a deeper understanding of the ground and excited state chemistry of flavoproteins, various spectroscopic experiments and quantum mechanical calculations have been conducted and documented in the literature. However, accurately simulating both ground and excited state relative energies using efficient quantum mechanical methods is a challenging task. To address this issue, a systematic benchmark study was conducted using different density functional and multireference methods, taking into account various active spaces and the number of roots in state averaging.

The results of this study revealed that at least a [10,10] active space and three states in state averaging are necessary for $\pi$-$\pi^*$ excitation energy calculations for oxidized flavin. In addition, potential energy surfaces (PESs) scans were performed by linear interpolation of coordinates from the $S_0$ optimized structure to the $S_1$ optimized structure. A similar scan was generated that connects the $S_1$ and $S_2$ optimized structures. The MS-CASPT2 and EOM-EE-CCSD methods were found to provide similar results and were in good agreement with each other for the low-lying excited states of flavin. The findings of this study provide a better understanding of the photochemical reactions and excited state chemistry of flavoproteins and pave the way for more accurate simulations of the complex biological systems.
6.2 Introduction

Computational methods have been a powerful tool in understanding the mechanism of biological chromophores and photoreceptor proteins.\textsuperscript{1} However, there are far fewer studies on flavoprotein photoreceptors compared to photoreceptors such as Rhodopsins. There are a number of quantum mechanical (QM) studies on both the ground state and excited state of flavin, where time-dependent density functional theory (TD-DFT) is widely used to understand photophysical and photochemical properties of flavin in both the gas phase and solvents\textsuperscript{2-6}. However, the accuracy of DFT calculations is not guaranteed, especially for the description of excited-state crossings. Although DFT methods account for dynamical correlation through the functional choices, traditional TD-DFT methods do not describe non-dynamical correlation because of the use of a single Slater determinant (single reference). Non-dynamical correlation is vital for atoms or molecules with nearly degenerate orbitals. Thus, multiconfigurational wave function methods often need to be used to study photophysical and photochemical processes to account for both dynamical and non-dynamical electronic correlation.

One of the most commonly used multiconfigurational methods is the complete active-space self-consistent field (CASSCF) method\textsuperscript{7}. In CASSCF, the configuration space comprises the electrons and orbitals included in the active space. A full configuration interaction is generated within this space, and the orbitals are variationally optimized simultaneously with the configuration coefficients.

Often, while CASSCF provides shapes of the potential energy surfaces qualitatively and adequately describes their interactions, energies from CASSCF are not quantitively correct due to missing dynamical electron correlation in the orbitals outside of the active space. To improve the accuracy of energetics from CASSCF calculations, CASSCF is used as a zero-order wave function
for post-CASSCF methods that account for this dynamical electron correlation. A popular approach is the second-order perturbation theory using CASSCF as a reference wavefunction (e.g., CASPT2).\textsuperscript{8,9} In this work, we have used different variants of CASPT2 such as SS-CASPT2, MS-CASPT2, XMS-CASPT2, or XDW-CASPT2. We also perform calculations using the multiconfiguration pair-density functional theory method (MC-PDFT), which also uses CASSCF as a reference wavefunction\textsuperscript{10,11}. However, the factorial increase in computational cost with active space size places a limit on how large these active spaces can be used for these methods. For example, a system with 16 electrons in 16 orbitals for the singlet spin state produces 35 million electronic configurations.\textsuperscript{12} Therefore, considering the expense and the accuracy of the result, choosing the right active space is very important for calculating any atom or molecule. Another recent and popular method to find the appropriate active space is the density matrix renormalization group (DMRG) CASSCF approach.\textsuperscript{13,14} Orbital entanglement calculations with DMRG have been used here for the automatic selection of active orbital spaces.

In this work, we have calculated orbitals entanglements for flavin's five redox and protonation states from DMRG-CASSCF wave functions. Using the highly entangled orbitals as a starting point, we have carried out a benchmark study with different density functional and multireference methods, active spaces, and the number of states in state averaging. We also have performed potential energy surfaces (PESs) scan along the S\textsubscript{0} minima to S\textsubscript{1} minima and S\textsubscript{1} minima to S\textsubscript{2} minima to check the validity of the methods over a wide range of geometries.

6.3 Computational details

The ground state geometries of flavin were optimized in the gas phase with second-order Moller-Plesset perturbation theory (MP2) and cc-PVTZ basis sets. The excited states of flavin (S\textsubscript{1} and S\textsubscript{2}) were optimized with TD-DFT method with the B3LYP/cc-pVTZ functional and basis set.
Cs symmetry has been used when relevant. Linearly interpolated geometries are generated for the construction of PESs connecting minima of different states. The interpolation and extrapolation are done from the S0 minimum to the S1 minimum and from the S1 minimum to the S2 minimum. SA-CASSCF, SS-CASPT2, MS-CASPT2, XMS-CASPT2, XDW-CASPT2, and MC-PDFT calculations were performed with the ANO-L-VDZP basis sets. A [14,12] active space is used consisting of two non-bonding orbitals that have electron density on N-atom and five sets of \( \pi \) and \( \pi^* \) orbitals. A level shift was applied to all CASPT2 calculations to avoid intruder states. In addition, single-point calculations along the PESs were performed with B3LYP/cc-pVTZ, CAM-B3LYP/cc-pVTZ, \( \omega \)-b97X-D, EOM-EE-CCSD/cc-pVDZ and SOS-CISD/cc-pVDZ levels of theory.

MP2 and DFT calculations were carried out using the Gaussian16 program package.\(^{15}\) Multi-reference calculations were carried out with the OpenMolcas program package (version 21.02).\(^{16}\) The orbital entanglement DMRG calculation was carried out using OpenMolcas interfaced to the QCMaquis DMRG program.\(^{17}\) The orbital entanglement information of the DMRG calculations was analyzed with the autoCAS program.\(^{18}\) EOM-EE-CCSD and SOS-CISD calculations were performed using the QCHEM package.\(^{19}\)

### 6.4 Results and Discussion

We carried out vertical excitation calculations for all the five redox and protonation states of flavin, including oxidized flavin (Fl), neutral semiquinone radical (FlH\(^+\)), anionic semiquinone radical (Fl\(^-\)), hydroquinone (FlH\(_2\)), and anionic hydroquinone (FlH\(^-\)).

The MP2-optimized geometries revealed that Fl, FlH\(^+\), and Fl\(^-\) were planar while FlH\(_2\) and FlH\(^-\) bent along a "butterfly" motion. Subsequently, we generated orbital entanglement diagrams using DMRG-CASSCF to aid with active space selection. Finally, we calculated the ground-state
and first excited-state energies for all redox and protonation states of flavin using TD-B3LYP and different multireference methods with varying active spaces and the number of roots in the state averaging.

6.4.1 Oxidized flavin (Fl)

The entanglement diagram of oxidized flavin (Fl) in a \([38,36]\) active space, calculated from CASSCF-DMRG wave functions, is shown in Figure 6.1. The active orbitals are arranged in a large circle, with each orbital represented by a small grey circle. The radius of the grey circle associated with each orbital is proportional to its single-orbital entropy, and the thickness of the lines connecting the circles is proportional to the mutual information element for each pair of orbitals. The diagram reveals that the 15 orbitals enclosed in the red box (8 occupied, 7 unoccupied), corresponds to a \([16,15]\) active space, have higher single-orbital entropies and mutual information compared to other orbitals. These 15 orbitals consist of one \(\pi_n\) orbital (orbital #2), seven bonding \(\pi\) orbitals (orbital #7, 10, 11, 15, 17, 18, and 19), and seven anti-bonding \(\pi^*\) orbitals (orbital #20, 21, 22, 23, 24, 25 and 27). Additionally, these 15 orbitals entangle or exchange mutual information with more than one orbital, e.g., orbital #17 entangling with orbitals #7, 23, and 24, while orbital #18 mainly entangles with orbital #22. The orbitals in the red box are particularly important for \(\pi-\pi^*\) excitation energy calculations. This is because the \(\pi-\pi^*\) transition involves the excitation of an electron from a \(\pi\) orbital to a \(\pi^*\) orbital, and thus a thorough consideration of the electronic structure of these orbitals is a necessary starting point for accurate calculations of \(\pi-\pi^*\) transitions. The strong entanglement and static correlation observed in the orbitals of oxidized flavin suggest that including these orbitals in the active space is essential for obtaining accurate excitation energies.
Large single-orbital entropy and strong entanglement with multiple orbitals are characteristic of static correlation, while weak entanglement among many orbitals or strong entanglement between two orbitals only indicate dynamic correlation. One important implication of the strong entanglement and static correlation observed in the orbitals of oxidized flavin is that traditional single-reference methods may not be sufficient for accurate calculations of electronic properties. This is because static correlation arises when more than one electronic configuration is energetically competitive, and traditional methods that rely on a single reference wave function are unable to accurately describe such scenarios.

**Figure 6.1** Orbital entanglement diagram of oxidized flavin.

The accuracy and the expense of the CASSCF method depend on the proper choice of orbitals in active space as well as the number of states used in the state averaging. The 15 orbitals selected
from DMRG calculation correspond to [16,15] active space and can be computationally expensive for iterative calculations. Thus, a balance must be struck between accuracy and computational cost to achieve the best results. To this end, a benchmark study was conducted using the highly entangled 15 orbitals selected from DMRG calculation with various multireference methods, active spaces, and numbers of roots in the state averaging. The active spaces were created without compromising accuracy by excluding nonessential orbitals for the first excited state or the specific process of interest. First, the [16,15] active space π-bonding and anti-bonding orbitals were generated and listed with their occupancies in Figure 6.2. The excitation energy to the first excited state was calculated using post CASSCF methods. Next, the highest occupancy orbital (π₀-orbital) was removed to create the next active space of [14,14] orbitals, and the energies were calculated with all the methods. Similarly, active spaces of [12,12], [10,10], and so on were generated, and their ground state and excited state energies were calculated. Additionally, the first excited state energies were computed by averaging over two to seven states. The energies obtained using different methods, active spaces, and numbers of states in the state averaging were compared to TD-B3LYP/cc-pVTZ energies and presented in Figure 6.3.

Figure 6.2 (a) π -bonding and (b) π -anti-bonding orbitals for oxidized flavin.
At first glance, it may seem that a minimal active space, such as [2,2] or [4,4], would suffice for obtaining reasonable excitation energies in comparison to medium-sized active spaces (as shown in Figure 6.3). For instance, the [2,2] active space yields more accurate results than the [6,6] active space. However, this does not necessarily mean that the former active spaces would give correct potential energy surfaces, which will be elaborated later.

There are large energy variations observed as the active space changes from [2,2] to [10,10]. This suggests that up to [10,10], the active space remains unbalanced, and each added orbital pair significantly contributes to the electron correlation. However, the excitation energy remains relatively constant as more orbitals are added beyond [10,10]. The active spaces [12,12], [14,14], and [16,15] yield similar excitation energies to the [10,10] active space but are more computationally expensive. Therefore, the [10,10] active space is a suitable choice for $\pi$-$\pi^*$ excitation energy calculations. The required orbitals in the active space are labeled as $\pi_1$, $\pi_2$, $\pi_3$, $\pi_4$, $\pi_5$, $\pi_1^*$, $\pi_2^*$, $\pi_3^*$, $\pi_4^*$, and $\pi_5^*$ in Figure 6.2.

Additionally, the excitation energies to the first excited state using the [10,10] active space do not vary much with respect to the different number of states included in the state averaging. However, when two states are included in the state averaging, they have slightly higher energies than when more states are included. Therefore, for $\pi$-$\pi^*$ excitation energy calculations, at least three roots should be used in state averaging. It appears that SA-CASSCF and MS-CASPT2 calculations are relatively stable when averaging over a larger number of excited states, as there are no significant variations in the excitation energy observed when averaging over additional excited states.
Figure 6.3 SA-CASSCF, SS-CASPT2, MS-CASPT2, XMS-CASPT2, XDW-CASPT2, MC-PDFT, and TD-B3LYP excitation energies of oxidized quinone with different active spaces and the number of states included in the state averaging. (a) two states included in the state averaging, (b) three states included in the state averaging, (c) four states included in the state averaging, (d) five states included in the state averaging, (e) six states included in the state averaging, and (f) seven states included in the state averaging.

6.4.2 Other redox states

The entanglement diagrams for the orbitals of the neutral semiquinone radical, anionic semiquinone radical, neutral hydroquinone, and anionic hydroquinone are displayed in Figure 6.10. The diagrams for the neutral and anionic semiquinone radicals were constructed similarly from CASSCF-DMRG wave function calculations in [39,36] active space. In contrast, the diagrams for neutral and anionic hydroquinone were created in [40,36] active space. The calculations show that the eight bonding and six anti-bonding orbitals are crucial for both neutral
and anionic semiquinone radicals compared to other orbitals. On the other hand, nine bonding and six anti-bonding orbitals are significant for both neutral and anionic hydroquinone. The important orbitals for $\pi-\pi^*$ excitation calculations are enclosed in the red box (shown in Figure 6.11).

We performed excitation energy calculations for the four redox and protonation states of flavin using a similar approach to the oxidized flavin in Figure 6.3. To create different active spaces, we first ran calculations with the [15,14] active space orbitals for both neutral and anionic semiquinone radicals and [18,15] active space orbitals for neutral and anionic hydroquinone. The active space orbitals were then ordered according to their occupancy, and removed from the active space based on their occupancies. The energies obtained with different methods, active spaces, and numbers of states included in the state averaging were compared with B3LYP/cc-pVTZ energies and presented in Figures 6.13, 6.14, 6.15, and 6.16.

The first excited state excitation energies for neutral and anionic semiquinone radicals, neutral and anionic hydroquinone are lowest with [3,3], [3,3], [2,2], and [2,2] active space, respectively (Figure 6.4). However, the energies increase as orbitals are added to the active space for all four cases. At some point, the energies become constant: for neutral semiquinone, the energies become constant after [7,7] active space, for anionic semiquinone radical after [5,5] active space, neutral hydroquinone after [8,8] active space, and anionic hydroquinone after [6,6] active space. This indicates that we need at least [7,7], [5,5], [8,8], and [6,6] active spaces for neutral and anionic semiquinone.
Figure 6.4 CASSCF, SS-CASPT2, MS-CASPT2 & XMS-CASPT2, XDW-CASPT2, MC-PDFT, and TD-B3LYP excitation energy of (a) neutral semiquinone radical, (b) anionic semiquinone radical, (c) neutral hydroquinone, and (d) anionic hydroquinone with different active spaces and three number of states in the state averaging.

6.4.3 Potential Energy Surface (PES) Scan:

To gain insight into the photophysics of a molecule, it is necessary to understand the energetics of low-lying excited states. For the flavin molecule, we analyzed the energetics of the lowest two singlet and triplet excited states along specific modes. These modes were mapped by performing linear interpolations along two distinct pathways: from the S₀ minimum to the S₁ minimum, and from the S₁ minimum to the S₂ minimum. Through interpolation and extrapolation,
we generated twenty-one geometries for each pathway and calculated the single point excitation energies for each geometry to generate a potential energy surface (PES) for each path (Figure 6.5).

**Figure 6.5** TD-DFT/cc-pVTZ PESs of flavin: (a) from $S_0$ minimum to the $S_1$ minimum. (b) from $S_1$ minimum to the $S_2$ minimum.

**Figure 6.5** displays the PESs of flavin obtained using TD-DFT B3LYP/cc-pVTZ along both pathways. The first single excited state ($S_1$) is optically bright and primarily $\pi-\pi^*$ in nature. The second singlet excited state is dark and $n-\pi^*$ in nature, with the electron density of non-bonding orbitals located on the N-atoms. Furthermore, the PESs for both pathways of flavin using MS-CASPT2 are illustrated in **Figure 6.6**, depicting the four lower-lying excited states of flavin. The first excited state is a triplet state ($T_1$) followed by $S_1$. The third excited state is another triplet ($T_2$) that is $n-\pi^*$ in nature, and the fourth excited state is $S_2$, which displays similar characteristics to those observed through TD-DFT.
Excited states of flavin undergo intersystem crossing (ISC) and internal conversion (IC). ISC occurs between S₁ and T₂ states as well as between S₂ and T₁ states, while IC occurs between S₁ and S₂ states as well as between T₁ and T₂ states. To compare the curvature of the low-lying excited states, PESs have been constructed using various methods including SA-CASSCF, SS-CASPT2, XMS-CASPT2, DW-CASPT2, MC-PDFT, SOS-CISD, and EOM-EE-CCSD, along with MS-CASPT2. Figure 6.6 presents a comparison of the curvature of excited states of flavin from S₀ minimum to S₁ minimum at different levels of theory, as well as for various TD-DFT functionals.
Figure 6.7 Comparison of shape of excited states of linearly interpolated geometries between S\textsubscript{0} minimum to the S\textsubscript{1} minimum of flavin at different level of theories such as SA-CASSCF, CASPT2, MC-PDFT, DFT, EOM-EE-CCSD, SOS-CISD: (a) S\textsubscript{0} (b) S\textsubscript{1} (\(\text{1}^\pi\pi^*\)) (c) S\textsubscript{2} (\(\text{1}^n\pi^*\)) (d) T\textsubscript{1} (\(\text{3}^\pi\pi^*\)) (e) T\textsubscript{2} (\(\text{3}^n\pi^*\)).

Similarly, Figure 6.8 represents the comparison of curvature of low-lying excited states of flavin molecule from S\textsubscript{1} minimum to the S\textsubscript{2} minimum at different methods.
Figure 6.8 Comparison of shape of excited states of linearly interpolated geometries between $S_1$ minimum to the $S_2$ minimum of flavin at different level of theories such as SA-CASSCF, CASPT2, MC-PDFT, DFT, EOM-EE-CCSD, SOS-CISD: (a) $S_0$ (b) $S_1$ ($^1\pi\pi^*$) (c) $S_2$ ($^1n\pi^*$) (d) $T_1$ ($^3\pi\pi^*$) (e) $T_2$ ($^3n\pi^*$).

From Figure 6.7 and Figure 6.8, we find the choice of method affects several aspects of the computations including 1) the ground state geometry, 2) the curvature of the ground state PES along the mapped path, 3) the relative energetics of the excited singlet and triplet states, and 4) the curvature of the excited states. Due to this effect of different methods on the curvature of excited states, the position of crossing between $S_1$ and $T_2$ changes from one method to another. To clarify these, Figure 6.9 plots both the $S_1$ ($^1\pi\pi^*$) and $T_2$ ($^3n\pi^*$) states in addition to the ground state, in one plot.
**Figure 6.9** Effect of different methods on ISC between $S_1 (^1\pi\pi^*)$ and $T_2 (^3n\pi^*)$ states of flavin.

From Fig. 6.9, we again see quite a lot of variation between the position of the $S_1 (^1\pi\pi^*) / T_2 (^3n\pi^*)$ crossing between different methods. In general TD-DFT give crossing points that are close to the $S_1 (^1\pi\pi^*)$ minimum geometry, while multi-reference and EOM-EE-CCSD methods indicate that the $S_1 (^1\pi\pi^*) / T_2 (^3n\pi^*)$ lies in the middle between the two minima. As a result, multi-reference methods (or single-reference methods that can capture multi-configurational character, like EOM-CCSD and SOS-CISD) will give higher barriers to intersystem crossing from $S_1 (^1\pi\pi^*)$ to $T_2 (^3n\pi^*)$ compared to TD-DFT methods, which have a very small barrier separating the $S_1 (^1\pi\pi^*)$ minimum and the $S_1 (^1\pi\pi^*) / T_2 (^3n\pi^*)$ crossing (regardless whether range-separated functionals are used or not).

We also calculated the NPE (nonparallelity errors, see Table 6.1) values along the both paths $S_0$-$S_1$ and $S_1$-$S_2$ for different methods relative to the DW-CASPT2 curves. The NPEs quantify how
different the shapes of the curves are. However, NPE does not say anything about the relative positions of the two curves.

The NPE values for MS-CASPT2 is zero which indicates that it gives identical curvature compared to DW-CASPT2. Interestingly, the NPEs vary a lot from state to state. For example, while SS-CASPT2 is mostly consistent with DW-CASPT2, large NPEs are observed for the \( S_1 \) state compared to DW-CASPT2. The EOM-EE-CCSD method provides NPEs that are comparable to other multi-reference methods, with the largest NPEs observed for the ground and \( T_1 \) state along the \( S_1 \) to \( S_2 \) path. On average, the NPEs for the TD-DFT methods are the largest among the methods tested, excluding CASSCF.

**Table 6.1** Nonparallelity errors (NPEs) in eV for the ground and excited states along with \( S_0 \)-\( S_1 \) and \( S_1 \)-\( S_2 \) path.

<table>
<thead>
<tr>
<th>Methods</th>
<th>( S_0 )</th>
<th>( S_1 )</th>
<th>( S_2 )</th>
<th>( T_1 )</th>
<th>( T_2 )</th>
<th>( S_0 )</th>
<th>( S_1 )</th>
<th>( S_2 )</th>
<th>( T_1 )</th>
<th>( T_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASSCF</td>
<td>0.54</td>
<td>0.21</td>
<td>0.45</td>
<td>0.39</td>
<td>0.10</td>
<td>0.50</td>
<td>0.08</td>
<td>0.45</td>
<td>0.53</td>
<td>0.13</td>
</tr>
<tr>
<td>SS-CASPT2</td>
<td>0.08</td>
<td>0.42</td>
<td>0.06</td>
<td>0.01</td>
<td>0.15</td>
<td>0.04</td>
<td>0.33</td>
<td>0.11</td>
<td>0.04</td>
<td>0.16</td>
</tr>
<tr>
<td>MS-CASPT2</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>XMS-CASPT2</td>
<td>0.02</td>
<td>0.19</td>
<td>0.16</td>
<td>0.08</td>
<td>0.05</td>
<td>0.00</td>
<td>0.20</td>
<td>0.15</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td>MC-PDFT</td>
<td>0.12</td>
<td>0.28</td>
<td>0.13</td>
<td>0.02</td>
<td>0.16</td>
<td>0.23</td>
<td>0.13</td>
<td>0.25</td>
<td>0.18</td>
<td>0.28</td>
</tr>
<tr>
<td>B3LYP</td>
<td>0.23</td>
<td>0.10</td>
<td>0.34</td>
<td>0.37</td>
<td>0.16</td>
<td>0.35</td>
<td>0.25</td>
<td>0.10</td>
<td>0.42</td>
<td>0.13</td>
</tr>
<tr>
<td>CAM-B3LYP</td>
<td>0.37</td>
<td>0.18</td>
<td>0.20</td>
<td>0.41</td>
<td>0.29</td>
<td>0.43</td>
<td>0.30</td>
<td>0.12</td>
<td>0.52</td>
<td>0.22</td>
</tr>
<tr>
<td>( \omega )-B97X-D</td>
<td>0.35</td>
<td>0.17</td>
<td>0.19</td>
<td>0.40</td>
<td>0.28</td>
<td>0.38</td>
<td>0.25</td>
<td>0.11</td>
<td>0.46</td>
<td>0.20</td>
</tr>
<tr>
<td>EOM-EE-CCSD</td>
<td>0.26</td>
<td>0.11</td>
<td>0.05</td>
<td>0.29</td>
<td>0.15</td>
<td>0.15</td>
<td>0.08</td>
<td>0.23</td>
<td>0.19</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Figure 6.10 Orbital entanglement diagram of (a) neutral semiquinone (b) anionic semiquinone, (c) neutral hydroquinone, and (d) anionic hydroquinone.
Figure 6.11 $\pi$-bonding and $\pi$-anti-bonding orbitals of the (a) neutral semiquinone radical, (b) anionic semiquinone radical, (c) neutral hydroquinone, and (d) anionic hydroquinone.
Figure 6.12 CASSCF, SS-CASPT2, MS-CASPT2, XMS-CASPT2, XDW-CASPT2, MC-PDFT, and TD-B3LYP excitation energies of neutral semiquinone radical with different active spaces and the number of states included in the state averaging. (a) two states included in the state averaging, (b) three states included in the state averaging, (c) four states included in the state averaging, (d) five states included in the state averaging, (e) six states included in the state averaging, and (f) seven states included in the state averaging.
Figure 6.13 CASSCF, SS-CASPT2, MS-CASPT2, XMS-CASPT2, XDW-CASPT2, MC-PDFT, and TD-B3LYP excitation energies of anionic semiquinone radical with different active spaces and the number of states included in the state averaging. (a) two states included in the state averaging, (b) three states included in the state averaging, (c) four states included in the state averaging, (d) five states included in the state averaging, (e) six states included in the state averaging, and (f) seven states included in the state averaging.
Figure 6.14 CASSCF, SS-CASPT2, MS-CASPT2, XMS-CASPT2, XDW-CASPT2, MC-PDFT, and TD-B3LYP excitation energies of neutral hydroquinone with different active spaces and the number of states included in the state averaging. (a) two states included in the state averaging, (b) three states included in the state averaging, (c) four states included in the state averaging, (d) five states included in the state averaging, (e) six states included in the state averaging, and (f) seven states included in the state averaging.
Figure 6.15 CASSCF, SS-CASPT2, MS-CASPT2, XMS-CASPT2, XDW-CASPT2, MC-PDFT, and TD-B3LYP excitation energies of anionic hydroquinone with different active spaces and the number of states included in the state averaging. (a) two states included in the state averaging, (b) three states included in the state averaging, (c) four states included in the state averaging, (d) five states included in the state averaging, (e) six states included in the state averaging, and (f) seven states included in the state averaging.
6.5 Conclusion

We have generated orbital entanglement diagrams for all five of flavin's redox and protonation states using CASSCF-DMRG wave functions. The diagrams help identify the orbitals that are important for excited state energy calculation and necessary for the multiconfigurational quantum mechanical method. Using the highly entangled orbitals as a starting point, we carried out a benchmark study with different density functional and multireference methods, active spaces, and the number of states in state averaging. We found that the first excited state energy of oxidized flavin significantly decreases with active space [10,10] compared to [8,8] active space, and after that, the energies stay relatively constant with adding more orbitals to the active space. The quality of the result does not improve with the [12,12], [14,14], and [16,15] active space compared to [10,10] active space. However, the simulation time exponentially increases. Therefore, [10,10] active space would be sufficient for \(\pi-\pi^*\) excitation energy calculations in terms of quality and efficiency. Similarly, the benchmark study for other redox states shows that we need at least [7,7], [5,5], [8,8], and [6,6] active spaces for neutral semiquinone radical, anionic semiquinone radical, neutral hydroquinone, and anionic hydroquinone, respectively, to calculate \(\pi-\pi^*\) excitation energy.

We also found that the first excited state energy for all redox/protonation states does not vary much with respect to the different number of states included in the state averaging. However, the two states included in the state averaging have a bit higher energy compared to the higher state used in the state averaging. Therefore, we decided to use at least three roots in the state averaging for \(\pi-\pi^*\) excitation energy calculations.

We also have computed potential energy surface (PESs) scans connecting the \(S_0\) minimum to \(S_1\) minimum and \(S_1\) minimum to \(S_2\) minimum to check the validity of the methods over a wide range of geometries. While different states give different energetics and shapes of the excited-state
PESs, we find that, in general, methods that capture multi-configurational character (i.e., multi-reference perturbation methods, EOM-CC, or SOS-CISD) capture the relative energetics and crossings between the different singlet and triplet states more accurately than TD-DFT, regardless of whether long-range corrected functionals are used or not. This is also reflected in non-parallelity errors that are also larger for TD-DFT compared to DW-CASPT2.
6.6 References


7 MODULATION OF LOW-LYING EXCITED STATE ENERGETICS IN THE FLAVIN COFACTOR

7.1 Abstract

The effect of hydrogen bonding interactions and specific geometric distortions on the ground and low-lying excited energetics of lumiflavin were investigated by means of TD-B3LYP quantum chemical calculations. We found that the hydrogen bonding with N1, N3, and C2=O atoms leads to a larger stabilization of the ground state (S0) compared to the first excited state (S1), which increases the S0-S1 energy gap. In contrast, hydrogen bonding with the N5 and C4=O atoms leads to a larger stabilization of S1 compared to S0, which decreases the S0-S1 energy gap. The hydrogen bonding with C9, C7C8, and C7C8” atoms have a negligible stabilizing effect on the excited state relative to the ground state and slightly decreases the S0-S1 excitation energy. We also looked at the effect of some specific geometric distortions on the S0-S1 energy gap. An out-of-plane twisting distortion slightly decreases the S0-S1 energy gap, while a distortion along a bond-length alternation (BLA) coordinate has a much more pronounced effect on both the S0-S1 and S0-S2 (i.e., the second singlet excited state) energy gap.

7.2 Introduction

The photophysical and photochemical properties of flavoproteins mainly depend on the flavin cofactor, which behaves differently in different protein environments upon absorption of blue light (~447nm). For instance, in the Chlamydomonas reinhardtii LOV domain, the flavin cofactor excited to the singlet excited state rapidly decays to a long-lived triplet state via an intersystem crossing (ISC).1, 2 About 20% of the triplet state flavin gives phosphorescence, while the rest reacts with the conserved cysteine residue in the active site to form a flavin-cysteinyl-thiol photo-adduct. Interestingly, adduct formation can also occur from the singlet excited state.3
Engineered flavin binding proteins, such as iLOV and miniSOG, do not have a cysteine residue in the vicinity of the flavin cofactor. As a result, these proteins lose their ability to form photo-adducts at the excited states and gain fluorescence properties or longer triplet state lifetimes. Central to all those photophysical processes are the actual states involved, i.e., the various low-lying singlet and triplet states and their energetics. Understanding how protein interactions modulate the energetics of those states can be the key to explaining the changing photophysics of the flavin cofactor from one flavoprotein receptor to another.

Some important work in this direction has already been carried out. Marian and co-workers compared the energetics and ISC of low-lying excited states of flavin in the gas, aqueous solution, and in YtvA LOV of *Bacillus subtilis*. They reported that in the gas phase, there is a low-lying triplet $3n\pi^*$ state comparable in energy to the first singlet $1\pi\pi^*$ excited state. In contrast, the $3n\pi^*$ state gets destabilized in solution and becomes inaccessible from the $1\pi\pi^*$ state. In YtvA LOV, the energetics resemble those of aqueous flavin, although ISC may also occur via $3n\pi^*$ if that state becomes energetically accessible (e.g., through variations in the protein conformations). The nearby cysteine side chain also plays a direct role in the ISC via a heavy-atom effect.

However, we are missing important details about how specific interactions between flavin, and the surrounding protein affect the relative energetics of the $1n\pi^*$, $1\pi\pi^*$, $3n\pi^*$, and $3\pi\pi^*$ states in different LOV domains. In this work, we systematically investigated the effect of hydrogen bonding at different places on the energetics of low-lying excited states of flavin. We also looked at how certain out-of-plane twisting motion and bond-length alternation coordinates affect the energetics of these excited states. This study can provide insight into how the specific interactions between flavin and the surrounding protein can modulate the energetics of the various excited states in different flavoprotein receptors.
Figure 7.1 The isoalloxazine ring of FMN and atomic labels. R=CH$_3$ for lumiflavin and R= ribose-5’-phosphate for FMN.

7.3 Computational details

In this study, lumiflavin, a reduced flavin model, was utilized for all calculations. Density functional theory (DFT) with the B3LYP functional and cc-pVTZ basis set was employed for ground state energy, geometry optimization, and frequency calculations.$^{12-14}$ Time-dependent DFT (TD-DFT)$^{16}$ with the same B3LYP functional and cc-pVTZ basis set$^{15}$ was used for all excited state energy calculations, geometry optimizations, and frequency calculations, unless stated otherwise. Vertical excitation energies were obtained as the energy difference between the ground and excited states, while adiabatic excitation energies were calculated as the energy difference between the excited and ground states including the zero-point vibrational energy (ZPVE) correction.

All DFT and TD-DFT geometry optimizations and frequency calculations were performed using Gaussian 16 software.$^{17}$ Twisted geometries were generated by distorting the flavin molecule along a low-frequency twisting mode (mode 1). These calculations are essential for understanding the impact of specific interactions on the relative energetics of the excited states of flavin, which can provide insight into the changing photophysics of the flavin cofactor in different protein environments.
7.4 Methodology

We employed a reduced flavin model, lumiflavin, to investigate the effect of hydrogen bonding on the low-lying excited state energies. To simulate the hydrogen bonding between the flavin cofactor and the surrounding amino acids, we used water molecules as a model system. We placed 106 water molecules at the Van der Waals distance of the flavin and optimized the system in the presence of each water molecule using the B3LYP functional and cc-pVTZ basis sets. From these calculations, we identified eight positions where water molecules are most likely to interact with the flavin cofactor, as shown in Figure 7.2. These are the same structures reported in Chapter 4, and in reference 18.

![Figure 7.2 Lumiflavin with water structures was employed in this study.](image)

We computed the vertical excitation energies for each of the eight structures in Figure 7.2. The results for the lowest two singlet excited states and five triplet states are presented in Table 7.2. To generate the geometries of flavin distorted along a twisting coordinate, we introduced an out-
of-plane twisting of the flavin along the lowest vibrational normal mode of lumiflavin computed with the B3LYP/cc-pVTZ method. The resulting geometries are shown in Figure 7.3(a).

In addition, we interpolated between the ground ($S_0$) and first singlet excited state ($S_1$) minimum geometries along the Franck-Condon mode. This mode follows a bond-length alternation (BLA) coordinate along the $N_1 - C_{10a} - C_{4a} - N_5$ backbone atoms of flavin. The resulting geometries are shown in Figure 7.3(b). We then calculated the vertical excitation energies of the low-lying excited states for the final structures generated by the interpolation, as shown at the bottom of Table 7.2.

Our results provide insights into the hydrogen bonding effect on the low-lying excited state energies of lumiflavin. The location of water molecules around the flavin cofactor was found to be critical for the interaction, and our findings provide a basis for understanding the role of hydrogen bonding in the functioning of flavin-containing enzymes.

**Figure 7.3** Interpolated geometry along a) twisting out of the plane b) bond length alternation (BLA)
7.5 Result and Discussion

We investigated the effect of hydrogen bonding between water molecules and lumiflavin on the low-lying excited state energies of the flavin cofactor. Of the eight structures shown in Figure 7.2., three have water molecules interacting with the hydrophobic region of flavin, while the remaining five structures have water molecules hydrogen bonded with the hydrophilic carbonyl region of flavin. Our calculations showed that the bond lengths of some backbone atoms in the optimized structures change significantly when water molecules are present at the N$_1$, N$_3$, N$_5$, C$_2$=O, and C$_4$=O positions, which is indicative of strong hydrogen bonding. Specifically, the bond length of N$_1$-C$_{10a}$, C$_2$-O$_{15}$, C$_4$-O$_{14}$, C$_2$-O$_{15}$, and C$_2$-N$_3$ increases and the bond length of N$_1$-C$_2$, C$_2$-N$_3$, C$_4$-C$_{4a}$, C$_2$-N$_3$, and C$_2$-C$_4$ decreases for water position at N$_1$, N$_3$, N$_5$, C$_2$=O and C$_4$=O, respectively.

To understand the excited state energies of lumiflavin in the presence of external water and for distorted geometries, we examined the relevant molecular orbitals (shown in Figure 7.4). The HOMO (π$_1$) showed electron density mainly at the N$_1$ atom and the benzene part of the isoalloxazine ring, while the LUMO (π$_1$*) had increased electron density on the C$_4$-C$_{4a}$ and C$_4$-C$_{10a}$ bonds. Two more π-bonding orbitals (π$_2$ and π$_3$) had electron density on the C$_8$-C$_9$ bond and N$_3$ atom, respectively, and the four lone pair molecular orbitals had increased electron density on oxygen and nitrogen atoms.
Table 7.1 Bond lengths of lumiflavin, in Angstroms, in the presence of water at different positions (indicated at the top of the table). The atomic labels are shown in Figure 7.1.

<table>
<thead>
<tr>
<th>Bond</th>
<th>gas</th>
<th>N1</th>
<th>N3</th>
<th>N5</th>
<th>C7C8</th>
<th>C7C8*</th>
<th>C2=O</th>
<th>C2=O</th>
<th>C8</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2-N3</td>
<td>1.410</td>
<td>1.403</td>
<td>1.399</td>
<td>1.415</td>
<td>1.412</td>
<td>1.411</td>
<td>1.403</td>
<td>1.413</td>
<td>1.411</td>
</tr>
<tr>
<td>C2-O15</td>
<td>1.212</td>
<td>1.214</td>
<td>1.223</td>
<td>1.210</td>
<td>1.211</td>
<td>1.211</td>
<td>1.214</td>
<td>1.210</td>
<td>1.212</td>
</tr>
<tr>
<td>C3-C4</td>
<td>1.375</td>
<td>1.379</td>
<td>1.377</td>
<td>1.371</td>
<td>1.376</td>
<td>1.376</td>
<td>1.379</td>
<td>1.366</td>
<td>1.376</td>
</tr>
<tr>
<td>C4-C4a</td>
<td>1.498</td>
<td>1.498</td>
<td>1.499</td>
<td>1.493</td>
<td>1.498</td>
<td>1.498</td>
<td>1.498</td>
<td>1.495</td>
<td>1.497</td>
</tr>
<tr>
<td>C4a-C10a</td>
<td>1.457</td>
<td>1.456</td>
<td>1.456</td>
<td>1.454</td>
<td>1.458</td>
<td>1.458</td>
<td>1.456</td>
<td>1.456</td>
<td>1.457</td>
</tr>
<tr>
<td>C4a-N5</td>
<td>1.293</td>
<td>1.292</td>
<td>1.292</td>
<td>1.291</td>
<td>1.292</td>
<td>1.292</td>
<td>1.292</td>
<td>1.292</td>
<td>1.293</td>
</tr>
<tr>
<td>C5a-C6</td>
<td>1.404</td>
<td>1.405</td>
<td>1.404</td>
<td>1.405</td>
<td>1.404</td>
<td>1.404</td>
<td>1.405</td>
<td>1.404</td>
<td>1.404</td>
</tr>
<tr>
<td>C7-C11</td>
<td>1.504</td>
<td>1.505</td>
<td>1.504</td>
<td>1.505</td>
<td>1.505</td>
<td>1.505</td>
<td>1.505</td>
<td>1.505</td>
<td>1.505</td>
</tr>
<tr>
<td>C7-C8</td>
<td>1.418</td>
<td>1.419</td>
<td>1.419</td>
<td>1.419</td>
<td>1.420</td>
<td>1.419</td>
<td>1.419</td>
<td>1.419</td>
<td>1.419</td>
</tr>
<tr>
<td>C8-C12</td>
<td>1.503</td>
<td>1.503</td>
<td>1.504</td>
<td>1.503</td>
<td>1.502</td>
<td>1.502</td>
<td>1.503</td>
<td>1.503</td>
<td>1.504</td>
</tr>
<tr>
<td>C8-C9</td>
<td>1.387</td>
<td>1.387</td>
<td>1.387</td>
<td>1.388</td>
<td>1.387</td>
<td>1.387</td>
<td>1.387</td>
<td>1.387</td>
<td>1.388</td>
</tr>
<tr>
<td>C9-C9a</td>
<td>1.399</td>
<td>1.400</td>
<td>1.400</td>
<td>1.398</td>
<td>1.399</td>
<td>1.399</td>
<td>1.400</td>
<td>1.399</td>
<td>1.399</td>
</tr>
<tr>
<td>C9a-C5a</td>
<td>1.413</td>
<td>1.413</td>
<td>1.413</td>
<td>1.414</td>
<td>1.413</td>
<td>1.413</td>
<td>1.413</td>
<td>1.413</td>
<td>1.413</td>
</tr>
<tr>
<td>N1-C10a</td>
<td>1.301</td>
<td>1.304</td>
<td>1.302</td>
<td>1.302</td>
<td>1.301</td>
<td>1.302</td>
<td>1.304</td>
<td>1.299</td>
<td>1.303</td>
</tr>
<tr>
<td>N1-C2</td>
<td>1.378</td>
<td>1.376</td>
<td>1.374</td>
<td>1.377</td>
<td>1.377</td>
<td>1.376</td>
<td>1.376</td>
<td>1.380</td>
<td>1.376</td>
</tr>
<tr>
<td>N10-C13</td>
<td>1.465</td>
<td>1.469</td>
<td>1.465</td>
<td>1.466</td>
<td>1.464</td>
<td>1.464</td>
<td>1.469</td>
<td>1.466</td>
<td>1.465</td>
</tr>
<tr>
<td>N5-C5a</td>
<td>1.364</td>
<td>1.363</td>
<td>1.364</td>
<td>1.364</td>
<td>1.364</td>
<td>1.364</td>
<td>1.363</td>
<td>1.364</td>
<td>1.364</td>
</tr>
</tbody>
</table>

Our TD-DFT calculations revealed that the first (S₁) and fourth singlet (S₄) excited states of flavin are optically bright and π- π* in nature. The first singlet excited state is a purely HOMO (π₁) to LUMO (π₁*) transition, while the fourth singlet excited state is a π₂ - π₁* transition. The second singlet excited state is n- π* in nature and optically dark, where the electron density of the non-bonding orbital is situated on N-atoms. The first three triplet state transitions are π- π* in nature, whereas the fourth and fifth triplet state transitions are nₕ₁ - π₁* and n₤₁ - π₁* in nature, respectively.
7.5.1 The effect of hydrogen bonding interactions

The effect of hydrogen bonding interactions on the energy gap between the first singlet excited state and the ground state in gas phase was investigated. The transition from HOMO to LUMO ($\pi_1 - \pi_1^*$) was found to have an energy gap of 3.03 eV. The HOMO ($\pi_1$) orbital had higher electron density at the N$_1$ atom, while the LUMO ($\pi_1^*$) had electron density on C$_4$-C$_{4a}$ and C$_4$-C$_{10a}$ bonds. The hydrogen bonding at different positions differentially stabilized these two states and affected the $S_0$-$S_1$ energy gap. Hydrogen bonding at N$_1$, N$_3$, and C$_2$-O stabilized the ground state more than the excited state and increased the energy gap to 3.05 eV, 3.06 eV, and 3.05 eV, respectively. On
the other hand, hydrogen bonding at N₅ and C₄=O positions stabilized the excited state (π₁*) more than the ground state and decreased the energy gap to 2.96 eV and 2.99 eV, respectively. Hydrogen bonding at C₉, C₇C₈, and C₇C₈” had a negligible stabilizing effect on the excited state and slightly decreased the excitation energy.

These observations were found to be consistent with the results of electrostatic spectral tuning maps (ESTMs) presented in Chapters 2 and 3 (reference 19 and 20). In the case of hydrophilic hydrogen-bonding interactions, the water was always oriented with its positive dipole pointed towards the flavin and acted like a positive charge in the vicinity of the flavin in terms of its stabilization effect. This means that the water’s effect on modulating the S₀-S₁ energy gap is largely electrostatic.

The effect of hydrogen bonding interactions on the second singlet excited state excitation energy, which was 3.15 eV in gas phase, was also investigated. The character of the transition was n₉₁ – π₁* in nature. The n₉₁ lone pair molecular orbital have electron density on the N₁ and N₅ atoms, therefore, hydrogen bonding nearby these two atoms is expected to stabilize the ground state more than the excited state. This is exactly what is observed experimentally. Specifically, the excitation energy of the second singlet excited state increases from 3.15 eV to 3.21 eV, 3.23 eV, 3.21 eV, and 3.19 eV for hydrogen bonding positions at N₁, N₅, C₂=O, and C₄=O, respectively. Conversely, hydrogen bonding at N₃ position, which had very little electron density in the n₉₁ frontier orbital, decreased the energy to 3.12 eV.

Similar trends were observed for the lowest π₁–π₁* and n–π₁* triplet states as observed for the lowest singlet states. These findings highlight the significance of hydrogen bonding interactions in modulating the electronic properties of molecules and provide insights into their photophysical properties.
7.5.2 The effect of geometry distortion

We examine the impact of flavin twisting (Figure 7.3(a)) and bond length alternation (BLA) (Figure 7.3(a)) on the excited states. The results are presented in Table 7.2, with the last two rows showing the effect of these changes on the energy gap of the excited states.

Regarding out-of-plane twisting, we observed that the ground state destabilizes more than both the first and second singlet excited states, which leads to a decrease in the energy gap in both cases. For instance, the energy gap decreased from 3.03 eV to 2.99 eV for the first excited state and from 3.15 eV to 3.11 eV for the second excited state. We also noticed that the lowest five triplet states show a similar decrease in excitation energy along the twisting mode compared to the ground state. These results indicate that the lowest-lying excited states in flavin are more flexible or prone to out-of-plane displacement compared to the ground state.

Furthermore, it is already known that lumiflavin's geometry relaxation in the first excited state involves a strongly coupled BLA motion that goes through the entire isoalloxazine backbone. Therefore, it is not surprising that distortion along that mode leads to a large decrease in the $S_0$-$S_1$ energy gap, from 3.03 eV to 2.53 eV. However, what is interesting is the effect of the same geometric change on other states. Specifically, we found that the energy gap also decreased for the second (dark) $n-\pi^*$ state from 3.15 eV to 2.77 eV. This is likely because the excitation involves the same $\pi^*$ LUMO state, which weakens the flavin's $\pi$-conjugated bonding when excited into. This same trend was observed for all the triplet states shown in Table 7.2, for the same reason.
Table 7.2 Vertical excitation energies of FMN in the presence of external water positions, heavy atoms, and distorted geometry.

<table>
<thead>
<tr>
<th></th>
<th>$S_1(\pi\pi^*)$</th>
<th>$S_2(n\pi^*)$</th>
<th>$T_1(\pi\pi^*)$</th>
<th>$T_2(\pi\pi^*)$</th>
<th>$T_3(n\pi^*)$</th>
<th>$T_4(n\pi^*)$</th>
<th>$T_5(n\pi^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMN</td>
<td>3.030</td>
<td>3.150</td>
<td>2.089</td>
<td>2.872</td>
<td>2.760</td>
<td>3.012</td>
<td>3.434</td>
</tr>
<tr>
<td>N1(H2O)</td>
<td>3.050</td>
<td>3.214</td>
<td>2.121</td>
<td>2.834</td>
<td>2.795</td>
<td>3.090</td>
<td>3.800</td>
</tr>
<tr>
<td>N3(H2O)</td>
<td>3.060</td>
<td>3.118</td>
<td>2.115</td>
<td>2.729</td>
<td>2.862</td>
<td>3.013</td>
<td>3.403</td>
</tr>
<tr>
<td>N5(H2O)</td>
<td>2.964</td>
<td>3.235</td>
<td>2.036</td>
<td>2.788</td>
<td>2.892</td>
<td>3.078</td>
<td>3.724</td>
</tr>
<tr>
<td>C2=O(H2O)</td>
<td>3.050</td>
<td>3.214</td>
<td>2.121</td>
<td>2.834</td>
<td>2.795</td>
<td>3.089</td>
<td>3.800</td>
</tr>
<tr>
<td>C4=O(H2O)</td>
<td>2.987</td>
<td>3.191</td>
<td>2.060</td>
<td>2.754</td>
<td>2.8624</td>
<td>3.072</td>
<td>3.394</td>
</tr>
<tr>
<td>C9(H2O)</td>
<td>3.018</td>
<td>3.177</td>
<td>2.088</td>
<td>2.830</td>
<td>2.771</td>
<td>3.042</td>
<td>3.465</td>
</tr>
<tr>
<td>C2C3(H2O)</td>
<td>3.017</td>
<td>3.175</td>
<td>2.077</td>
<td>2.847</td>
<td>2.772</td>
<td>3.039</td>
<td>3.461</td>
</tr>
<tr>
<td>C2C4(=O)(H2O)</td>
<td>3.017</td>
<td>3.175</td>
<td>2.077</td>
<td>2.847</td>
<td>2.772</td>
<td>3.039</td>
<td>3.461</td>
</tr>
<tr>
<td>FMN$_{Twisted}$</td>
<td>2.993</td>
<td>3.107</td>
<td>2.032</td>
<td>2.795</td>
<td>2.701</td>
<td>2.942</td>
<td>3.239</td>
</tr>
<tr>
<td>FMN$_{BLa}$</td>
<td>2.526</td>
<td>2.777</td>
<td>1.613</td>
<td>2.532</td>
<td>2.594</td>
<td>2.627</td>
<td>3.065</td>
</tr>
</tbody>
</table>

7.6 Conclusion

The study of how intermolecular interactions impact the photophysical properties of flavin is a crucial aspect of understanding flavoprotein photoreceptors. In this chapter, we build on previous research to explore the effect of such interactions on the low-lying $^1n\pi^*$, $^1\pi\pi^*$, $^3n\pi^*$, and $^3\pi\pi^*$ states of flavin. Our focus was on examining the influence of hydrogen bonding interactions, bond-length alternation, and out-of-plane twisting distortion on these states.

Our results showed that hydrogen bonding with the N1, N3, and C2=O atoms stabilized the ground state more than the first $^1\pi\pi^*$ and $^3\pi\pi^*$ excited states, leading to an increased energy gap. Conversely, hydrogen bonding with N5 and C4=O atoms stabilized the first excited state ($\pi_1^*$) more...
than the ground state and decreased the energy gap. The effect of hydrogen bonding with C₉, C₇C₈, and C₇C₈” atoms on the excited state was found to be similar and slightly decreased the excitation energy. When it comes to ¹nπ* and ³nπ* states, the most significant impact on the excitation energy was observed with hydrogen bonds formed with atoms where non-bonding orbitals are located.

Furthermore, we discovered that both out-of-plane and bond-length alternation distortions decreased the excitation energy of all low-lying excited states. However, bond-length alternation had a much more significant impact than out-of-plane distortions. In conclusion, our findings provide valuable insight into how intermolecular interactions impact the photophysical properties of flavin, an essential first step in the study of flavoprotein photoreceptors.
7.7 Reference


8 CONCLUSION

In this dissertation, we calculated vertical, adiabatic, and vibronic electronic excitations for the five redox/protonation states of flavin in the gas phase as well as in different polar solvents. We found that Fl’s UV/vis spectrum is not very sensitive to the solvent or protein environment. However, the second excited state (peak at ~370 nm) is relatively more sensitive than the first excited state (peak at ~448 nm). In the case of the FlH⁻, the absorption spectrum is considerably more sensitive to the electrostatic environment, such that decreasing the polarity is associated with a decrease in the intensity of the peak at ~450-500 nm and significant shifts in the relative positions/intensities of the peaks near 300-350 nm. On the other hand, Fl⁻ is far less sensitive to its electrostatic environment. Finally, the opposite trend is observed for the neutral and anionic hydroquinone. At the same time, FlH₂ is not very sensitive to the environment, FlH⁻ is significantly more sensitive, particularly in the visible range from around 350-450 nm.

We also simulated flavin’s FTIR spectrum in the presence of water molecules to investigate the effect of hydrogen bonding interaction on the frequency and intensity of flavin’s prominent peak (1450 cm⁻¹ to 1750 cm⁻¹). We found that the \( \nu_{C=\cdot} \) and \( \nu_{C=N} \) mode frequencies are not sensitive to any hydrogen-bonding interactions, but their relative intensities are affected by the dielectric environment of the solvent. On the other hand, while the \( \nu_{C_2=O} \) and \( \nu_{C_4=O} \) mode frequencies are strongly downshifted by direct (both \( \nu_{C_2=O} \) and \( \nu_{C_4=O} \)) and indirect (only \( \nu_{C_2=O} \)) hydrogen-bonding interactions, their relative intensities are less sensitive to the dielectric environment. A continuum solvation model or a cluster model with multiple water molecules is needed to reproduce the relative frequencies of the four modes in flavin.

We also have developed a new facile automated computational protocol termed electrostatic spectral tuning maps (ESTMs) to understand how a biological chromophore will
spectroscopically respond to different electrostatic environments. The map is an intuitive starting point to understand how to tune the spectroscopic properties of the chromophore by changing its environment. The map was built by moving a point charge on the chromophore’s van der Waals surface and performing single-point energy calculation in the presence of each point charge. The excitation and emission energy changes were convoluted and represented in a 3D map. Using the automated protocol, we have generated ESTMs for biological model chromophores, including the flavin cofactor in different redox states.

The first excited ESTM of flavin indicates that if we place a positive charge containing amino acid at the vicinity of N5/C4a atoms or a negative charge containing amino acid at the vicinity of the N1 atom, the flavin’s absorption spectrum will shift to the higher wavelength. Placing a positive charge amino acid (lysine) in the vicinity of the N5/C4a atom is proposed by Khrenova and coworkers. Therefore, we suggest an alternative strategy by placing a negatively charged amino acid (glutamic acid) near flavin’s N1 atom. With the help of QM/MM calculation, we predicted a single-point mutation iLOV-Q430E has a slightly red-shifted absorption and fluorescence maximum wavelength relative to iLOV. We experimentally verified our computational prediction by expressing and purifying iLOV-Q430E. We found that the Q430E mutation in iLOV results in a slight change in absorption and a 4-8 nm redshift in the fluorescence relative to iLOV.

Studying the photophysics of flavin required quantum mechanical methods that will accurately simulate ground and excited state relative energies with time efficiency. Therefore, we have performed a systematical benchmark study with different density functional and multireference methods by considering various active spaces and the number of roots in the state averaging. We found that at least [10,10] active space and three states in the state averaging are
necessary for oxidized flavin for \( \pi-\pi^* \) excitation energy calculations. We also have performed potential energy surfaces (PESs) scans using linearly interpolated geometries from \( S_0 \) minima to \( S_1 \) minima and from \( S_1 \) minima to \( S_2 \) minima and found that multi-reference perturbation methods and EOM-EE-CCSD provide more comparable curves, relative excited-state energies, and crossing geometries compared to TD-DFT methods that we tested.

Finally, we have investigated the effect of hydrogen bonding interaction and out-of-plane and bond-length alternation geometric distortions on the low-lying excited states with TD-B3LYP functional. We found that hydrogen bonding with \( N_1, N_3, \) and \( C_{2\text{-}O} \) atom/atoms stabilize the ground state more than the first excited state and increases the energy gap. In contrast, the hydrogen bonding with \( N_5 \) and \( C_{4\text{-}O} \) atom/atoms stabilizes the first excited state (\( \pi_1^* \)) more than the ground state and decreases the energy gap. The hydrogen bonding with \( C_9, C_{7\text{-}8}, \) and \( C_{7\text{-}8}'' \) atom/atoms have a similar and negligible stabilizing effect on the excited state and slightly decreases the excitation energy.

Overall, we have used reduced computational model systems to generally study the effect of intermolecular interactions between flavin and surrounding moieties (i.e., a solvent or protein) on the energetics of low-lying states and vibrational frequencies of flavin. We also report some of the developments we have contributed to the average solvent electrostatic configuration – free energy gradient (ASEC-FEG) QM/MM approach, which we used to study spectral tuning of iLOV, a flavoprotein fluorescent protein. Future directions include the application of the ASEC-FEG approach to the study of energetics and photophysics of both natural and artificial flavoprotein photoreceptors like LOV domains and their derivatives. Another future direction is the extension of this work to look not only at energetics but also at how the external environment modulates other electronic properties of flavin, like spin-orbit couplings.