Temperature Dependent Protein-Chromophore Hydrogen Bond Dynamics in the Far-Red Fluorescent Proteins by using Molecular Dynamics Simulation and Quantum Calculation

Chandra Prasad Dhakal
cdhak001@fiu.edu

Follow this and additional works at: https://digitalcommons.fiu.edu/etd

Recommended Citation

https://digitalcommons.fiu.edu/etd/4837

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.
TEMPERATURE DEPENDENT PROTEIN-CHROMOPHORE HYDROGEN BOND DYNAMICS IN THE FAR-RED FLUORESCENT PROTEINS BY USING MOLECULAR DYNAMICS SIMULATION AND QUANTUM CALCULATION

A dissertation submitted in partial fulfillment of
the requirements for the degree of
DOCTOR OF PHILOSOPHY
in
PHYSICS
by
Chandra Prasad Dhakal

2021
To: Dean Michael R. Heithaus  
College of Arts, Sciences, and Education

This dissertation, written by Chandra Prasad Dhakal, and entitled Temperature Dependent Protein-Chromophore Hydrogen Bond Dynamics in the Far-Red Fluorescent Proteins by using Molecular Dynamics Simulation and Quantum Calculation, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

__________________________________________________________________________________________
Watson J. Lees  
Prem P. Chapagain, Co-Major Professor

__________________________________________________________________________________________
Jin He  
Xuewen Wang, Co-Major Professor

Date of Defense: November 10, 2021

The dissertation of Chandra Prasad Dhakal is approved.

__________________________________________________________________________________________
Dean Michael R. Heithaus  
College of Arts, Sciences, and Education

__________________________________________________________________________________________
Andrés G. Gil  
Vice President for Research and Economic Development and Dean of the University Graduate School

Florida International University, 2021
© Copyright 2021 by Chandra Prasad Dhakal

All rights reserved.
DEDICATION

I dedicate this dissertation to my parents (Shiva P Dhakal and Late Tara Devi Dhakal), my in-laws (Ganga Prasad Kharel and Gayatri Kharel), and my beloved wife. My Ph.D. degree is their dream and would not have been possible without their sacrifice.
ACKNOWLEDGMENTS

I want to express my profound gratitude and appreciation to my Ph.D. supervisors, Professor Xuewen Wang and Professor Prem P. Chapagain. Their invaluable mentoring and support have provided me with the necessary pieces of training to excel in my research. During my graduate years, I have never felt distressed or frustrated, mainly because of their patience, encouragement, and understanding. I am very grateful to my committee members Professor Jin He and Professor Watson J. Lees, for their expert suggestions, comments, and willingness to help me always. I am thankful to the Department of Physics Faculty and Staff members. I am thankful to Barry Branch, the CASTIC team, and FIU HPC team for their computer support.

I owe more than an acknowledgment to my parents and in-laws for their continuous love and support. I always remember the support, encouragement, and care of my brother, sisters, and friends during my long education journey. My beloved wife, Srijana Kharel, has a unique place in this acknowledgment for providing unbound love, care, and understanding. She is always aware of my health during my cancer journey and took special care at the moment of Stem Cell Transplant. She does a lot for me to get my health back to normal. She is doing MBA at Atlantis University now. She provides enough time to do my research work and keeps me out of household activities. Finally, I would like to thank all of my friends (Rudra, Jeevan, Prabin, Nisha, and Lokmann) in the computational BioPhysics group for their continuous help, support, and suggestions.
ABSTRACT OF THE DISSERTATION

TEMPERATURE DEPENDENT PROTEIN-CHROMOPHORE HYDROGEN BOND DYNAMICS IN THE FAR-RED FLUORESCENT PROTEINS BY USING MOLECULAR DYNAMICS SIMULATION AND QUANTUM CALCULATION

by

Chandra Prasad Dhakal

Florida International University, 2021

Miami, Florida

Professor Xuewen Wang, Co-Major Professor

Professor Prem P. Chapagain, Co-Major Professor

Fluorescent proteins are valuable tools as biochemical markers in molecular and cell biology research for studying cellular processes. Red Fluorescent Proteins (RFPs) are highly desirable for in vivo applications in living cell imaging because they absorb and emit light in the red region of the spectrum where cellular autofluorescence. Naturally occurring fluorescent proteins with emission peaks in this region of the spectrum occur in dimeric or tetrameric forms. For their use as biochemical markers, several monomeric variants of RFP have been developed which include mCherry, dsRed, and mStrawberry. Far red-emitting FPs with large Stokes shift are especially valuables for in vivo applications due to the advantage of deep tissue penetration, longer imaging times, and low cellular autofluorescence. Examples of far-red emitting FPs include mPlum, mKate, TagRFP675, and more recently developed versions include mNeptune1, mNeptune2.5, and mCardinal2. Low-temperature experiments on mneptune1, mNeptune2.5, and mcardinal2 show a reduced Stokes shift compared to room temperature. To characterize their Stokes shift
behavior at different temperatures, I used a combination of molecular dynamics (MD) simulations and quantum mechanical calculations and investigated the dynamics of the hydrogen bonds formed due to protein-chromophore interactions at different temperatures for these FPs. Since the switching between direct and water-mediated hydrogen bonds has been shown to correlate with the Stokes shift in a related protein mPlum, I investigated the agile switching of the chromophore-PHE62 hydrogen bond between the direct and water-mediated bonding in these FPs at various temperatures. MD simulations show that while all three variants show the ability to switch the water-mediated vs. direct hydrogen bonding at 300K, mCardinal2 shows better hydrogen bond flexibility. My results provide insights into the role of thermal fluctuations on the solvation and hydrogen bonding in the chromophore environment in the FP variants. The structures of the extended chromophore obtained from MD simulations were used to calculate the excitation and emission energy/wavelength using quantum mechanical calculation. These calculations were performed separately for the direct vs. water-mediated structures extracted from the simulation trajectories. My results show that the increased flexibility of the chromophore environment at higher temperatures along with its ability to reorganize after excitation is related to the larger Stokes shift.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. MOLECULAR DYNAMICS METHOD</td>
<td>7</td>
</tr>
<tr>
<td>2.1 MOLECULAR MECHANICS: POTENTIAL ENERGY FUNCTIONS</td>
<td>8</td>
</tr>
<tr>
<td>2.1.1 BONDED INTERACTIONS AND BOND-RELATED INTERACTIONS</td>
<td>9</td>
</tr>
<tr>
<td>2.1.2 NON-BONDED INTERACTIONS</td>
<td>12</td>
</tr>
<tr>
<td>2.2 TIME INTEGRAL</td>
<td>13</td>
</tr>
<tr>
<td>2.2.1 VERLET ALGORITHM</td>
<td>13</td>
</tr>
<tr>
<td>2.2.2 LEAP-FROG ALGORITHM</td>
<td>14</td>
</tr>
<tr>
<td>2.2.3 VELOCITY VERLET ALGORITHM</td>
<td>15</td>
</tr>
<tr>
<td>2.3 MOLECULAR DYNAMICS: SYSTEM SET UP</td>
<td>15</td>
</tr>
<tr>
<td>2.4 FORCE FIELD PARAMETERS FOR CHROMOPHORE</td>
<td>19</td>
</tr>
<tr>
<td>2.5 SYSTEM AND SIMULATION</td>
<td>20</td>
</tr>
<tr>
<td>2.6 FLUORESCENT PROTEIN AND CHROMOPHORE</td>
<td>20</td>
</tr>
<tr>
<td>3. FIRST PRINCIPLE QUANTUM METHOD</td>
<td>23</td>
</tr>
<tr>
<td>3.1 MULTI-CONFIGURATION METHOD</td>
<td>24</td>
</tr>
<tr>
<td>3.11 COMPLETE ACTIVE SPACE METHOD</td>
<td>25</td>
</tr>
<tr>
<td>3.2 DENSITY FUNCTIONAL THEORY</td>
<td>29</td>
</tr>
<tr>
<td>3.2.1 HOHENBERG-KOHN THEOREM</td>
<td>31</td>
</tr>
<tr>
<td>3.2.2 KOHN-SHAM EQUATION</td>
<td>31</td>
</tr>
<tr>
<td>3.2.3 ELECTRON DENSITY</td>
<td>33</td>
</tr>
<tr>
<td>3.2.4 TOTAL ENERGY</td>
<td>34</td>
</tr>
<tr>
<td>3.2.5 EXCHANGE-CORRELATION ENERGY</td>
<td>34</td>
</tr>
<tr>
<td>3.3 TIME-DEPENDENT DENSITY FUNCTIONAL THEORY</td>
<td>36</td>
</tr>
<tr>
<td>3.3.1 RUNGE-GROSS THEOREM</td>
<td>37</td>
</tr>
<tr>
<td>3.3.2 TIME-DEPENDENT KOHN-SHAM EQUATION</td>
<td>38</td>
</tr>
<tr>
<td>3.3.3 DENSITY RESPONSE FUNCTION</td>
<td>41</td>
</tr>
<tr>
<td>3.3.4 PERTURBATION THEORY</td>
<td>43</td>
</tr>
<tr>
<td>4. MOLECULAR DYNAMICS SIMULATION RESULTS AND DISCUSSION</td>
<td>46</td>
</tr>
<tr>
<td>4.1 mNEPTUNE1</td>
<td>51</td>
</tr>
<tr>
<td>4.2 mNEPTUNE2.5</td>
<td>54</td>
</tr>
<tr>
<td>4.3 mCARDINAL2</td>
<td>57</td>
</tr>
<tr>
<td>5. QUANTUM MECHANICAL CALCULATION</td>
<td>60</td>
</tr>
<tr>
<td>5.1 DFT AND CASSCF CALCULATION ON BENZENE</td>
<td>64</td>
</tr>
<tr>
<td>5.2 TDDFT CALCULATION OF BENZENE</td>
<td>69</td>
</tr>
<tr>
<td>5.3 QUANTUM MECHANICAL CALCULATION ON CHROMOPHORE</td>
<td>70</td>
</tr>
<tr>
<td>5.3.1 HYDROGEN CAPPING AND THEIR ELECTRONIC STRUCTURE</td>
<td>70</td>
</tr>
<tr>
<td>5.3.2 RESULT ON mCARDINAL2</td>
<td>75</td>
</tr>
</tbody>
</table>
6. CONCLUSION ............................................................................................................78
REFERENCES ..............................................................................................................81
APPENDICES .............................................................................................................88
VITA ............................................................................................................................99
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Stokes shift Experimental data for Far-RFPs</td>
<td>3</td>
</tr>
<tr>
<td>Table 2</td>
<td>H-bonds to chromophore from neighboring sidechains are presented in terms of time fraction as calculated from MD simulations in mNeptune1</td>
<td>52</td>
</tr>
<tr>
<td>Table 3</td>
<td>H-bonds to chromophore from neighboring sidechains are presented in terms of time fraction as calculated from MD simulations in mNeptune2.5</td>
<td>55</td>
</tr>
<tr>
<td>Table 4</td>
<td>H-bonds to chromophore from neighboring sidechains are presented in terms of time fraction as calculated from MD simulations in mCardinal2</td>
<td>58</td>
</tr>
<tr>
<td>Table 5</td>
<td>Singlet ground ($S_0$), lowest triplet ($T_1$), and first singlet excited ($S_1$) of Benzene</td>
<td>66</td>
</tr>
<tr>
<td>Table 6</td>
<td>Singlet ground ($S_0$), lowest triplet ($T_1$), and first singlet excited ($S_1$) of Benzene with more active space</td>
<td>66</td>
</tr>
<tr>
<td>Table 7</td>
<td>The singlet ground ($S_0$), lowest triplet ($T_1$), and first singlet excited ($S_1$) of benzene</td>
<td>68</td>
</tr>
<tr>
<td>Table 8</td>
<td>Excitation and emission of benzene from singlet energy and triplet energy</td>
<td>68</td>
</tr>
<tr>
<td>Table 9</td>
<td>Excitation and emission wavelength of the chromophore</td>
<td>71</td>
</tr>
</tbody>
</table>
# Lists of Figures

<table>
<thead>
<tr>
<th>Figures</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1 Jablonski diagram illustrating the Fluorescence</td>
<td>4</td>
</tr>
<tr>
<td>Figure 2 Excitation, Emission, and Stokes shift</td>
<td>5</td>
</tr>
<tr>
<td>Figure 3 Bonded and non-bonded interactions</td>
<td>9</td>
</tr>
<tr>
<td>Figure 4 Fluorescent protein structure with chromophore inside (left figure) and chromophore with an extended chromophore (right figure)</td>
<td>20</td>
</tr>
<tr>
<td>Figure 5 Chromophore structure of mNeptune variants</td>
<td>22</td>
</tr>
<tr>
<td>Figure 6 Chromophore maturation process (adapted from reference 45). [O] and [H] represent the oxidation and reduction respectively</td>
<td>23</td>
</tr>
<tr>
<td>Figure 7 Potential and its corresponding wave function of all-electron (blue) compared that of pseudo electron (red)</td>
<td>33</td>
</tr>
<tr>
<td>Figure 8 Red Fluorescent Protein</td>
<td>48</td>
</tr>
<tr>
<td>Figure 9 Chromophore-hydrogen bond with nearby amino acid residues</td>
<td>50</td>
</tr>
<tr>
<td>Figure 10 Direct hydrogen bond and water-mediated hydrogen bond between extended chromophore (62) of mNeptune1 and nearby amino acid residues</td>
<td>51</td>
</tr>
<tr>
<td>Figure 11 Distance time trajectories between the extended chromophore and nearby amino-acid residues of mNeptune1</td>
<td>53</td>
</tr>
<tr>
<td>Figure 12 Direct hydrogen bond and water-mediated hydrogen bond between extended chromophore of mNeptune2.5 and nearby amino-acid residues</td>
<td>54</td>
</tr>
<tr>
<td>Figure 13 Distance time trajectories between the extended chromophore and nearby residues of mNeptune2.5</td>
<td>56</td>
</tr>
<tr>
<td>Figure 14 Direct hydrogen bond and water-mediated hydrogen bond between extended chromophore of mCardinal2 and nearby amino-acid residues</td>
<td>57</td>
</tr>
<tr>
<td>Figure 15 Distance time trajectories between the extended chromophore and nearby residues of mCardinal2</td>
<td>60</td>
</tr>
<tr>
<td>Figure 16 Benzene hexagonal ring</td>
<td>65</td>
</tr>
</tbody>
</table>
Figure 17 Chromophore structures based on the different terminus .................................................72

Figure 18 Histogram Plot of Excited-State of Chromophore (a) Unoptimized Hydrogen Cap (b) Optimized Hydrogen Cap ........................................................................................................74

Figure 19 Double Histogram Plot of Direct Hydrogen Bond and Water-mediated Hydrogen Bond at temperatures (a) 300K and (b) 240K ..........................................................77
### ABBREVIATIONS AND ACRONYMS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPs</td>
<td>Fluorescent Proteins</td>
</tr>
<tr>
<td>RFPs</td>
<td>Red Fluorescent Proteins</td>
</tr>
<tr>
<td>YFP</td>
<td>Yellow Fluorescent Protein</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescent Protein</td>
</tr>
<tr>
<td>NIR</td>
<td>Near Infra-Red</td>
</tr>
<tr>
<td>CHARMM</td>
<td>Chemistry at Harvard Molecular Mechanics</td>
</tr>
<tr>
<td>LJ</td>
<td>Lennard-Jones</td>
</tr>
<tr>
<td>VDW</td>
<td>Van Der Waals</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>MM</td>
<td>Molecular Mechanics</td>
</tr>
<tr>
<td>QM</td>
<td>Quantum Mechanics</td>
</tr>
<tr>
<td>PME</td>
<td>Particle Mesh Ewald</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>PBC</td>
<td>Periodic Boundary Condition</td>
</tr>
<tr>
<td>NAMD</td>
<td>Nanoscale Molecular Dynamics</td>
</tr>
<tr>
<td>VMD</td>
<td>Visual Molecular Dynamics</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>DFT</td>
<td>Density Functional Theory</td>
</tr>
<tr>
<td>CI</td>
<td>Configuration Interaction</td>
</tr>
<tr>
<td>MCSCF</td>
<td>Multi-Configuration Self-Consistent Field</td>
</tr>
<tr>
<td>CASSCF</td>
<td>Complete Active Space Self-Consistent Field</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>B3LYP</td>
<td>Becke 3-Parameter Lee-Yang-Parr</td>
</tr>
<tr>
<td>LDA</td>
<td>Local Density Approximation</td>
</tr>
<tr>
<td>GGA</td>
<td>Generalized Gradient Approximation</td>
</tr>
<tr>
<td>GTOs</td>
<td>Gaussian Type of Orbitals</td>
</tr>
<tr>
<td>HPC</td>
<td>High-Performance Computational</td>
</tr>
<tr>
<td>HF</td>
<td>Hartree-Fock</td>
</tr>
<tr>
<td>HK</td>
<td>Hohenberg Kohn</td>
</tr>
<tr>
<td>KS</td>
<td>Kohn Sham</td>
</tr>
<tr>
<td>TDDFT</td>
<td>Time-Dependent Density Functional Theory</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

Fluorescent proteins (FPs) are a valuable tool in molecular and cell biology research. They have revolutionized the field of molecular biology as they allow tracking of cellular and physiological processes in living cells by visualizing protein localization and interactions [1]. Green fluorescent proteins (GFPs) extracted from jellyfish Aequorea victoria was discovered in 1960 [2] and at the same time, researchers could exhibit the GFP gene in other cells [3, 4]. It has found many applications in protein biology such as protein tags, reporters of gene expression, and cell lineage tracers. Researchers studied carefully and found a new improved monomeric variant having a wide spectral range of visible light [5]. Different mutations were done on the chromophore of wild-type GFP such as S65T, Y66H, and Y66W to get Enhanced GFP, Blue FP, and a Cyan FP respectively [6, 7]. Similarly, a mutation in barrel T203Y yielded Yellow FP [8]. But it was difficult to achieve the Red Fluorescent Proteins (RFPs) from the mutation of wild-type GFP. For live-cell imaging of mammalian cells, RFPs that excite and emit at a wavelength in the red region of the spectrum are preferred over GFPs. The naturally occurring RFPs such as DsRed are in dimeric or tetrameric forms [9, 10], which tend to oligomerize [11, 12] and are not suitable for fusion tagging [13]. My present work focuses on monomeric variants RFPs. The monomeric variants of RFPs such as mCherry avoid the issue of oligomerization and are used for in vivo applications in living cell imaging [14]. To overcome the issues of cellular autofluorescence and deeper tissue imaging, next-generation fluorescent proteins with excitation and emission of light in the far-red region of the spectrum (emission beyond 650 nm), and possessing enhanced fluorescence quantum yield are highly sought after [15-17]. Compared to shorter wavelength FPs, far-red FPs have lower autofluorescence, lower
scattering, and deeper tissue penetration. These include the FP variants mPlum, tagRFP, and mKate. More recent variants of such RFPs are mNeptune, mNeptune1, mNeptune2, mNeptune2.5, mCardinal, mCardinal1, and mCardinal2.

Molecular Dynamics (MD) simulations of Red Fluorescent Protein (RFP) reveal innovative ideas into the dynamics of the interaction between the E16 and I65 sidechains in mPlum and a possible explanation for its large Stokes shift relative to RFPs. The simulations represent interconversion between direct and water-mediated hydrogen bonding sub-states between the 16 and 65 side chains [18]. There is a direct correlation between the Stokes shift and hydrogen bond interconversion. The RFPs with the greatest propensity to interconvert between hydrogen bond states have the largest Stokes shift. This dynamic model of the chromophore mainly assigned its red-shifted emission to direct stabilization of the excited state through a hydrogen bond between E16 and the N-acylimine oxygen of I65. The hypothesis of such dynamics facilitates the emission wavelength within Near Infra-Red (NIR) region through a combination of excited state solvation and regulation of chromophore flexibility [19]. This model provides proper guidance for my computational research related to the temperature dependence of protein chromophore interactions of far RFPs in mNeptune variants mneptune1, mNeptune2.5, and mCardinal2. I have also done a different calculation by using Quantum Mechanics (QM) calculation. In this process, I have taken direct hydrogen bond and water-mediated hydrogen bond from MD simulation under multiple time frames at the temperature of 300K and performed excited-state calculation using Time-Dependent Density Functional Theory (TDDFT).
There are several experimental research done in Fluorescent Proteins (FPs) particularly on another mNeptune variants as shown in Table 1. These are different variants than that I am showing in my dissertation work. Experiment on mNeptune681 and mNeptune684 shows a more red shift than mNeptune because of the extensive hydrogen bond network around the chromophore as a result of barrel mutation [20]. Their emission wavelength lies within the NIR range. These proteins are allowed to use in medical research, especially for cancer diagnosis and their treatment. The mammalian tissues became transparent to NIR and can easily penetrate inside the body tissues to scan the cancer tumors. Also, these experiment indicates enough space for improving the FPs with desired properties.

<table>
<thead>
<tr>
<th>Far-RFP</th>
<th>$E_{ex}$ (nm)</th>
<th>$E_{em}$ (nm)</th>
<th>Stokes shift (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mPlum</td>
<td>590</td>
<td>649</td>
<td>59</td>
</tr>
<tr>
<td>mNeptune</td>
<td>600</td>
<td>650</td>
<td>50</td>
</tr>
<tr>
<td>mNeptune681</td>
<td>605</td>
<td>685</td>
<td>80</td>
</tr>
<tr>
<td>mNeptune684</td>
<td>604</td>
<td>684</td>
<td>80</td>
</tr>
</tbody>
</table>

When a molecule absorbs radiation, it changes into an excited state. The excited molecule returns to the ground state with the release of energy in the form of radiation of a particular wavelength. This process of emission of radiation is called Fluorescence. In this phenomenon, the energy of emitted radiation is smaller than the excitation energy [21]. In other words, the wavelength of emitted radiation is longer than the wavelength of excitation during the fluorescence process. I am going to describe the fluorescence phenomenon in Fluorescent Protein (FPs), particularly due to the chromophore. This
process is elucidated by a well-known three-stage electronic-state diagram, which is known as the Jablonski diagram as shown in Figure 1.

![Jablonski diagram](image)

*Figure 1* Jablonski diagram illustrating the Fluorescence

A photon of energy $h\nu_{EX}$ is absorbed by a molecule at singlet ground state, $S_0$ and get electronically excited state $S'_1$. The existence of an excited state occurs for a small finite time of range 1ns—10ns where a molecule is allowed to a profusion of possible interactions with its both electronic and molecular environment along with its conformational changes. This mechanism is carried out by partial dissipation of $S'_1$ the energy that helps to obtain a relaxed singlet excited $S_1$ from which the emission occurs.

A photon of energy $h\nu_{EM}$ is emitted, when $S_1$ undergoes the initial ground state $S_0$. The energy of an emitted photon is lower than an excited photon because of the energy dissipation of vibrational mode in the excited state lifetime. Therefore, the wavelength of emitted radiation is longer than the wavelength of excitation. The difference of wavelength between these two radiations is called the Stokes shift.
If $\lambda_{ex}$ and $\lambda_{em}$ are the wavelength of excitation and wavelength of emission, then Stokes shift is given by

$$\Delta \lambda = \lambda_{em} - \lambda_{ex}$$

We can write a Stokes shift formula is the difference between the peak value of emitted radiation and the corresponding value of excitation wavelength as mentioned in Figure 2.

$$\Delta \lambda = \lambda_{em}^{peak} - \lambda_{ex}^{peak}$$

The chromophore of FP lies within the 11-stranded beta-barrel and it is autocatalytically formed by fusing three amino acids in the presence of oxygen. The DsRed-like chromophore absorbs at 560–580 nm and emits at 570–610 nm [22], with a Stokes shift of 20 nm. Excitation wavelength in fluorescent proteins is determined by the covalent structure of the chromophore and influenced by hydrogen bonding and electronic interactions of the chromophore with nearby amino acid residues [23]. The chromophore excitation causes the electron density to shift from the phenoxy ring towards the acylimine.

Figure 2 Excitation, Emission, and Stokes shift
region across the π-conjugated system [24, 25]. The changes in the chromophore environment can perturb the emission spectra due to the change in the excited state dynamics [26-30]. The molecular engineering of the most recent RFPs with large Stokes shifted emissions could be a vital topic. Chromophore-protein sidechain hydrogen bonding interactions are observed in the crystal structure of some far-red fluorescent proteins. There are also internal water molecules trapped within the chromophore environment. Hydrogen-bonding to the extended chromophore (N-acylimine carbonyl of the chromophore) has been observed in most red-emitting RFPs [31, 32]. This bonding strongly correlates with the large shift observed in far RFPs, but the detailed molecular mechanism is still unclear.

Low-temperature experiments on mNeptune1, mNeptune2.5, and mcardinal2 show a reduced stokes shift compared to room temperature. This suggests that the increased flexibility of the chromophore environment at higher temperatures along with its ability to reorganize after excitation is expounded to the larger Stokes shift. I used molecular dynamics (MD) simulations to investigate the dynamics of the hydrogen bonds formed in the chromophore regions of the far-red fluorescent proteins mNeptune1, mneptune2.5, and mcardinal2. I explored the protein-chromophore hydrogen bond pattern at various temperatures and correlate the hydrogen bond dynamics to the experimentally observed Stokes shifts.

The residue next to the chromophore (e.g. PHE62 in mNeptune and its variants), called the extended chromophore, plays an especially important role in determining the Stokes shift. I used the MD simulations method to investigate the different types of hydrogen bonding (direct hydrogen bond and water-mediated hydrogen bond) between the chromophore and nearby amino acid residues and chromophore flexibility at different
temperatures. Specifically, the hydrogen bond between the extended chromophore (residue 62) and other amino acid residues or an internal water molecule dictates the nature of Stokes shift. Previous research has shown that the ability of protein-chromophore conformations to interconvert between two hydrogen bond states (direct and water-mediated hydrogen bond) allows the FP to possess an extended Stokes shift.

In this research, I investigate how the hydrogen-bonding pattern in the acylimine region of the extended chromophore is correlated with the observed Stokes shifts in three far-red RFPs mNeptune1, mNeptune2.5, and mcardinal2. This result was verified by performing electronic structure calculations. In this process, I calculated the electronic structure of the isolated chromophore, the electronic structure of both direct hydrogen bond and water-mediated hydrogen bond of mCardinal2 in both the ground and excited states by using DFT and time-dependent DFT. From this calculation, excitation and emission energy and wavelengths of multiple structures at different frames were calculated. Finally, numerically data of Stokes shift of different conformations were obtained. The results provide a basis for the interpretation of the experimental observations, illuminate the character of the extended Stokes shift observed in far RFP and reveal important information in engineering next generation far-red FPs.

2. MOLECULAR DYNAMICS METHOD

The Scientific study of the relationship between molecular structure, dynamics, and function in a biological system is greatly assisted by MD simulations [33]. More detailed and accurate characteristics of FPs are given by Quantum Mechanical (QM) calculation [34] in comparison with classical MD. FPs have a large number of atoms and QM calculations are expensive (requiring longer time) in terms of computational cost. I
carried out the structure consisting of a few atoms (chromophore atoms and nearby amino-acid residue that took part in bonding) only. So, MD simulation is feasible and is broadly used in the study of macromolecular properties. The MD simulation can find out important dynamics of a macromolecule at the molecular level but is not as detailed as the QM-like electronic level.

2.1 MOLECULAR MECHANICS: POTENTIAL ENERGY FUNCTIONS

The atoms possess in the simulation experience exact potential energy due to the force field used. The force field and the potential taken are comparable. The validity of the force fields taken compared to the exact potentials is critically important in determining the reliability of the sample [35] in replicating the actual behavior of a system. These are crucial in MD calculation to investigate the accuracy of protein conformation [36]. CHARMM is an important simulation package [37] and it has a potential energy function as given in Equation 2.1. Bonded interactions between atoms (Figure 3) are modeled as harmonic oscillator potential functions except for the dihedral energy term. Non-bond interactions are also included.

\[
U(R) = \sum_{\text{bonds}} K_b (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{UB}} K_{UB} (S - S_0)^2 + \\
\sum_{\text{dihedrals}} K_{\phi} (1 + \cos (n \phi - \delta)) + \sum_{\text{impropers}} K_{\omega} (\omega - \omega_0)^2 + \\
\sum_{\text{non-bonded}} \left\{ \epsilon \left( \frac{\sigma}{r_{ij}} \right)^{12} - 2 \left( \frac{\sigma}{r_{ij}} \right)^6 + \frac{q_i q_j}{4 \pi \varepsilon_0 \varepsilon_{r_{ij}}} \right\} + \sum_{\text{Residues}} UCMA \left( \varphi, \psi \right) \quad (2.1)
\]

Where \(K_b\), \(K_\theta\), \(K_{UB}\), \(K_{\phi}\), and \(K_{\omega}\) are the force constants for bond stretching, bond angle changes, bond-angle bending (non-bonded Urey-Bradely energy), dihedral angle rotation, and improper dihedral angle changes respectively. Similarly, \(r_0\), \(\theta_0\), \(S_0\), and \(S_0\) are corresponding equilibrium values. Equation 2.1 can be divided into bonded energy terms
and non-bonded energy terms. The first four terms and the last term in Equation 2.1 are bonded energy terms, whereas the fifth term with brackets contains two non-bonded terms representing van der Waals and electric-charge interactions. The derivative of the potential energy with respect to spatial coordinates provides forces that act on the atoms. All of these terms are explained in more detail below.

![Figure 3 Bonded and non-bonded interactions](image)

**2.1.1 BONDED INTERACTIONS AND BOND-RELATED INTERACTIONS**

**Bond Stretching Energy**

Figure 3 shows a system having four atoms $i$, $j$, $k$, and $l$. Bond stretching from their equilibrium lengths can occur between pairs of atoms $i$-$j$, $j$-$k$, and $k$-$l$. The bond stretching energy between two covalently is determined by the first term in Equation 2.1, i.e. $\sum_{\text{bonds}} K_b (r - r_0)^2$. The force constant $K_b$ provides the strength of the bond. The $r_0$ is the equilibrium bond length. Parameters $r_0$ and $K_b$ are fixed to the recognition of atoms in pairs.
Angle Bending

A bond angle is an angle between pair of atoms $i$-$j$ and $j$-$k$ as shown in Figure 4. The angle bending energy is the energy required to bend a bond from its equilibrium angle, $\theta_0$. The second term $\sum_{\text{angles}} K_{\theta} (\theta - \theta_0)^2$ in Equation 2.1 is the energy bending with angular force constant $K_{\theta}$. The force constant $K_{\theta}$ is responsible to produce the energy in-plane bending. The energy termed mentioned above is in the harmonic form given by its potential

$$U(\theta) = \frac{k}{2} (\theta - \theta_0)^2$$

Where $k$ is the force constant. It provides the lowest energy to cause the deviation from equilibrium geometry.

Urey-Bradley Energy

The Urey-Bradley energy term, $\sum_{UB} K_{UB} (S - S_0)^2$ in Equation (2.1) is responsible for the development of the CHARMM force field. This energy helps to set up calculated data in close agreement with the experimental vibrational frequencies. This energy term is also called harmonic potential with equilibrium distance $S_0$ and force constant $K_{UB}$. Here $S_0$ is the distance between atoms $i$ and $k$ in Figure 1. The existence of this energy term is possible only when atoms $i, j, k$ are identified, otherwise, $K_{UB} = 0$ and the term vanishes.

Torsional Energy

The rotation of atoms within the molecules changes the energy of the system and its conformations. This rotation develops a torsion around the four atoms $i, j, k$, and $l$. The torsional energy is represented by the sum of cosines functions and can be written as
\[ \sum_{\text{dihedrals}} K_\varphi (1 + \cos(n\varphi - \delta)) \] in Equation (2.1) as the fourth energy term with multiplicities \( n = 1, 2, 3 \ldots \), amplitude \( K_\varphi \), phase \( \delta \) and \( \varphi \) is the torsion angle. This energy term arises due to the rotation of covalent bonds around four atoms \( i, j, k, \) and \( l \). Amplitude \( K_\varphi \) is the constant which gives the strength of the potential and the barrier height resisting full rotation, angular phase factor \( \delta \) reflects the most favorable torsional angle, and the multiplicity \( n \) represents the number of potential energy barriers experienced in one full rotation of the angle \( \varphi \). The torsion angle is set up between plane \( i-j-k \) and \( j-k-l \). This term is periodic.

**Improper Energy**

The improper energy term \( \sum_{\text{impropers}} K_\omega (\omega - \omega_0)^2 \) is used to explain the planarity of confirmations when needed. This energy is out of plane energy function to get the proper geometry. The improper angle \( \omega \) is identical to the torsional angle \( \varphi \) for the equivalent geometry composed of atoms \( i, j, k, \) and \( l \). But the form of potential is different with force constant \( K_\omega \). The explanation of improper angle is similar to the angle of torsion, but atoms are not bonded in a sequence. The terms used in the improper energy represent the tendency of the system to remain planar.

**CMAP Correction**

The last term in Equation (2.1) is \( \sum_{\text{Residues}} UCMAP(\varphi, \psi) \) a numerical value of energy added to the CHARMM force field to improve the ability by correcting the backbone dihedral angles \( \varphi \) and \( \psi \) to experimentally observe secondary structure. CHARMM36 is the new version of CHARMM. It contains both phi-psi backbone dihedral CMAP correction term and side-chain dihedral angles optimization. It equilibrates the
system particles between α-helices and β-region. CMAP correction term was not included in CHARMM22, but the α-helices bias of its force fields [38] are corrected as a result of optimization.

2.1.2 NON-BONDED INTERACTIONS

In MD simulation, the trajectory of particles contains both bonded and non-bonded interactions in their force field. The term in Equation (2.1)

\[ \Sigma_{\text{non-bonded}} = \epsilon \left( \frac{\sigma}{r_{ij}} \right)^{12} - 2 \left( \frac{\sigma}{r_{ij}} \right)^6 \right) + \frac{q_i q_j}{4 \pi \varepsilon_0 \varepsilon r_{ij}} \]

represents non-bonded interactions where \( \sigma \) is collision diameter between two atoms is, \( \epsilon \) is the potential well depth and \( r_{ij} \) is the intra-atomic distance. This interaction contains both Van Der Waals (VDW) term and electrostatic term. The electrostatic interaction with total charge \( q_i \) produces due to unequal distribution of charges in the system of particles, which is given by coulomb potential. The VDW interaction consists of all types of interactions between systems of particles except the electrostatic term. It means, all atoms pair set up non-bonded energy with potential function via VDW interaction. This interaction is represented by a Lennard-Jones (LJ) potential [39]. The VDW term contains both repulsive \( \left( \frac{1}{r_{ij}^{12}} \right) \) and attractive \( \left( \frac{1}{r_{ij}^6} \right) \) term. The repulsive term can be explained by Pauli Exclusion Principle and induced dipole moment is responsible for attractive interactions. MD simulation is done with the help of the latest version of CHARMM, the CHARMM36 force field that accounts for hydrogen bond calculation from electrostatic and VDW interaction, without any explicit term in Equation (2.1) for it. MD simulation for non-bonded interaction is more expensive in terms of computational cost (time). Cut off distance should be set up higher value \( r_{ij} \) to limit the number of non-bonded interactions and interaction greater than cut-off distance should be
introduced as zero without any calculation. This process lowers the computational time during MD simulation.

2.2 TIME INTEGRAL

MD simulation contains the system of particles that uses Newton’s laws of motion to predict the position of a particle as a function of time. From Newton’s law, \( \vec{F} = m\vec{a} \), which is used to determine the acceleration of each particle in the protein system at any time. Here, \( \vec{F} \) is the force acting on the particle at time \( t \) and \( m \) is the mass of the particle. Once calculate acceleration, their momenta are also determined. As the process continues, these values are useful to calculate the new position of a particle after time step \( \Delta t \). Some of the algorithms used in MD simulations are explained below.

2.2.1 VERLET ALGORITHM

Verlet Algorithm is used to calculate the trajectories of the system of particles in MD simulation by integrating Newton’s Equation of motion. Suppose the position of the particles is the function of time and the small-time step during simulation is \( \Delta t \). Then, the Taylor series expansion of position is expressed as:

\[
\begin{align*}
 r(t + \Delta t) &= r(t) + \frac{dr}{dt} \Delta t + \frac{1}{2} \frac{d^2r}{dt^2} (\Delta t)^2 + \cdots \quad (2.2) \\

 r(t - \Delta t) &= r(t) - \frac{dr}{dt} \Delta t + \frac{1}{2} \frac{d^2r}{dt^2} (\Delta t)^2 + \cdots \quad (2.3)
\end{align*}
\]

By adding Equations (2.2) and (2.3), we get

\[
r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t)(\Delta t)^2 \quad (2.4)
\]

The new position of the particle at time \( t + \Delta t \) is given by Equation (2.4), which contains second order of time \((\Delta t)^2\) only and depends explicitly on velocity. Therefore,
the final position of the particle is obtained from its position and acceleration at current time \( t \) and previous time \( t - \Delta t \). This is the advantage of the given algorithm to calculate the correct position of the particle in a trajectory in absence of the first-order term in \( \Delta t \).

### 2.2.2 LEAP-FROG ALGORITHM

In MD simulation, this algorithm is obtained by integrating second-order differential Equation \( \frac{d^2r}{dt^2} = a(t) \) to get the new position of particles during simulation.

The position of the system of particles at time \( t + \Delta t \) is calculated from Equation (2.2) as

\[
r(t + \Delta t) = r(t) + \frac{dr}{dt} \Delta t + \frac{1}{2} \frac{d^2r}{dt^2}(\Delta t)^2 = r(t) + \frac{dr}{dt} \left( t + \frac{\Delta t}{2} \right) \Delta t
\]

\[
= r(t) + v(t + \frac{\Delta t}{2}) \Delta t \tag{2.5}
\]

Where \( v \left( t + \frac{\Delta t}{2} \right) = \frac{dr}{dt} \left( t + \frac{\Delta t}{2} \right) \)

On expanding the velocity term in time step \( \Delta t \) is given by

\[
v \left( t + \frac{\Delta t}{2} \right) = v(t) + \frac{dv(t)}{dt} \frac{\Delta t}{2} + \frac{1}{2} \frac{d^2v(t)}{dt^2} \left( \frac{\Delta t}{2} \right)^2 + \cdots \tag{2.6}
\]

\[
v \left( t - \frac{\Delta t}{2} \right) = v(t) - \frac{dv(t)}{dt} \frac{\Delta t}{2} + \frac{1}{2} \frac{d^2v(t)}{dt^2} \left( \frac{\Delta t}{2} \right)^2 + \cdots \tag{2.7}
\]

By subtraction Equations (2.6) and (2.7), we get

\[
v \left( t + \frac{\Delta t}{2} \right) = v \left( t - \frac{\Delta t}{2} \right) + a(t)\Delta t \tag{2.8}
\]

Here, \( a(t) = \frac{dv(t)}{dt} \)

From Equations (2.5) and (2.8), the position of particles at time \( t + \Delta t \) can be calculated. The velocities of the system of particles at time \( t \) can be approximated by

\[
v(t) = \frac{1}{2} \left[ v \left( t + \frac{\Delta t}{2} \right) + v \left( t - \frac{\Delta t}{2} \right) \right] \tag{2.9}
\]
Therefore, this algorithm is useful to calculate the accurate position of particles at a time, \( t + \Delta t \).

### 2.2.3 VELOCITY VERLET ALGORITHM

This algorithm is used in MD simulation to update the position and velocity of particles at time \( t + \Delta t \) and position of particles during trajectories can be found as

\[
\begin{align*}
    r(t + \Delta t) &= r(t) + v(t)\Delta t + \frac{1}{2} a(t)(\Delta t)^2 \\
    v(t + \Delta t) &= v(t) + \frac{1}{2} [a(t) + a(t + \Delta t)]\Delta t
\end{align*}
\]

This algorithm is considered to be more accurate to calculate positions and velocities of particles by taking average acceleration at time \( t \) and \( t + \Delta t \).

### 2.3 MOLECULAR DYNAMICS: SYSTEM SET UP

The FPs system contains a large number of atoms and molecules. During the simulations, potential energy function, integration algorithms, and Newton’s equation of motion are crucial to developing the trajectory of the systems. The computational system is summarized by taking the initial time \( t=0 \) of the system of atoms from x-ray crystallographic structure, their corresponding positions \( x, y, z \), and velocities are defined as soon as user-defined a small time step. This section is used to explain the summary of MD simulation. All of these procedures are described based on the CHARMM simulation protocol.

Biomolecules are an essential component in biophysics to study MD simulation. I need an initial structure for simulation that is available online in Protein Data Bank (PDB). The structure obtained from PDB is found in a format known as a PDB file. The geometry found in PDB is x-ray crystallography structure or NMR structure. The biomolecule such
as a protein of interest in the form of a PDB file is downloaded from PDB and it should be studied carefully. Sometimes, the residues and atoms may be missing in the PDB file. The CHARMM-GUI is crucial to fill up the necessary missing residues and atoms in the file using the PDB input generator method from the drop-down menu in the tool (CHARMM-GUI) [40]. Also, the BUILD command in the tool adds the necessary hydrogen atoms required to the protein geometry thus getting the complete structure of the desired protein. Finally, the notations of the name of a few residues or atoms may need to change in the PDB file. It makes the simulation and calculation of protein compatible with MD tools.

The files should be prepared for the production run. There are several steps to make it suitable for calculation. The first step of the MD simulation is solvation. The water molecules are important in the solvation process to set up the initial step of the simulation. The proteins and nucleic acids act as the real biomolecules, while they are made to surround by water molecules in the form of a water solvation box. The size of the water solvation box should be huge enough to cover up all protein surfaces during simulation time. The approximate size of the water solvation box may require twice the size of the initial protein structure in each direction (x, y, z) that can increase the computational costs. Another alternative method to make a water solvation box is using Periodic Boundary Condition (PBC) where the size of the box is a little bigger than the protein structure. During simulation, residues reach the edge of the solvation box and experience the solvation effect of water molecules lying at the opposite faces of the box. So, there are greater chances of physical interactions of residues lying between two opposite faces. To prevent these interactions, the size of the solvation water box should be corrected which is equal to the length of the longest axis of protein plus double the cut-off distance for non-bonded
interactions. This size of the water solvation box enables the protection of the unwanted interactions among residues during MD simulation.

The next step of the simulation is the neutralization of the protein system. The amino acids lying in the protein system have either positive or negative charges. It means overall protein may have a total positive or negative charge. These charges enable the long-range electrostatic interaction in presence of PBC that is measured by the Particle Mesh Ewald (PME) method [41]. The counter ions are added to the protein to make the overall system-neutral that allowing using the PME method.

After neutralization of the protein system, MD simulation should be done, but the initial positions of atoms may not be real. So there are greater chances of stopping the simulation after a few time steps. The positions of atoms in the protein system are conjectured as a result of x-ray crystallography, CHARMM-GUI, and solvation. If the initial approximation of the position of atoms is spaced by a fraction of an angstrom, the potential energy produced in some locations is responsible to create large forces and acceleration without any physical interactions. This measurement makes an MD simulation unstable such as breaking bonds due to higher velocities and greater changes in positions of atoms. The MD simulation has set up the tests to evaluate whether large energy and energy gradient are key factors to crash MD simulation or not. In order to correct this problem, the MD simulation tools have a technique to develop minimum molecular energy of systems by enabling small separation relative to their initial position. This initial minimum energy geometry may be a local minimum, but not the global minimum energy structure. This initial energy minimized configuration is a required geometry to perform MD simulation. The energy minimization steps are used to run the full MD simulations.
because of their similarity in the procedure. The only difference in energy minimization is that the MD simulation inhibits large energies and forces to maintain the smaller possible spacing of atoms. Also, planned MD simulation follows the same PME basis and potential energy cut-off parameters used at the time of energy minimization. Once the simulation process achieved the minimum change in energy (less than 0.001 kcal/mol), as a result of jumping from one time step to the next time step, then the system is said to converge to a local energy minimum. The number of time steps to converge the configuration is not finite and minimization steps are terminated once achieve the minimum energy change.

After getting minimum energy of protein with solvated water box, the configuration is ready to subject for equilibration which is the previous step before pursuing actual MD simulation. The velocity of each atom in configurations is provided by Maxwell-Boltzmann distribution at the temperature of 300K. This step requires the large velocity correction of some atoms that may have a higher probability of getting unphysical motion. The interaction from such motion impractically changes the protein configurations. Therefore, atoms in the system are assigned smaller velocity at the beginning of simulation at low temperature (<50K) and subjected to run some smaller steps (such as 500) to interact with atoms gently to achieve new equilibrium positions. The system is built up in such a way that every increase in velocity of the system of atoms correspondingly increases a small amount of temperature (5K, 10K, or 15K) and the system is allowed to relax again. This process is repeated until the simulation temperature of interest is obtained, which develops the small incremental heating on the system. This process is helpful to prevent the configurations from structural instability and distortion due to unexpected enormous changes in velocity at high temperatures. The equilibrations of the system are achieved till
it gets the final temperature and system properties (pressure, temperature, and energy) become stable with time.

The last step of MD simulation is the production run. The dynamics of amino acid residues in a protein are obtained as soon as the actual run “production run” is finished. This step is expected to be performed for a certain length of time for a protein to show some useful changes in configuration. The protein system contains a large number of atoms in different positions \((x, y, z)\) and velocities at each time step. The time evolution of this system of configuration is called trajectory which can be used to analyze properties of amino acid residues within the system. Some properties in this simulation are difficult to perform because of the lack of computer power and system. Another restriction in this simulation is the higher computational cost that may require a couple of weeks, even a month.

2.4 FORCE FIELD PARAMETERS FOR CHROMOPHORE

The potential energy of the protein system is given by equation (2.1) that has different physical parameters for individual pairs of the system of atoms. The values of parameters like force constants, equilibrium bond angles, and their distances are mentioned for all pairs of atoms in any protein in the MD package. Sometimes bonding geometry of atoms of proteins may not be similar, at that time numerical values of the parameter for an individual protein (few related proteins) are required, which are distinct to notable configuration. The chromophore is the main responsible part of FP to show stokes shift. So, the numerical values of the parameter of the FP chromophore should be needed and obtained from the CHARMM force field parameterization technique [42]. Other remaining parameters were adopted from the CHARMM36 force field.
2.5 SYSTEM AND SIMULATION

Initial protein structure (mNeptune, PDB code:3IP2) was taken from the protein data bank. Missing residues were added by the CHARMM-GUI PDB input generator tool. Some Far RFP proteins do not have their PDB file in the protein data bank. Necessary mutations of mNeptune were done to get variants mNeptune1, mNeptune2.5, and mCardinal2 with the help of CHARMM-GUI. MD simulations were carried out by using CHARMM and NAMD packages where an all-atom CHARMM36 force field was adopted. All three far RFP were solvated with water molecules (TIP3P) in a water box of cut-off size 10 Å in each simulation.

2.6 FLUORESCENT PROTEIN AND CHROMOPHORE

The GFP was discovered in the early 1960s, as a byproduct of photo-protein, Aequorin and it is a protein extracted from the jellyfish named Aequorea Victoria. The famous researcher Osamu Shimomura is the first scientist to study the GFP and reveal the possibility of getting the fluorescent mechanism. GFP crystal structure was shown in
Protein Data Bank (PDB) in 1966 [43]. I studied the monomeric variant of RFP mNeptune that are mNeptune1, mNeptune2.5, and mCardinal2. Each variant is composed of 11 beta-strands that are arranged in the cylindrical shape of length 40 Angstrom and diameter 30 Angstrom as shown in Figure 4. Loops and short Alpha helices form caps on both ends of beta can shape that can prevent the entry of unwanted matter inside the barrel. The helices run down the geometric center of the β-barrel and can hold a chemically modified tripeptide called the chromophore. The letter B indicated by the arrow sign on the right side of Figure 4 are points connected to the barrel with the chromophore. Each variant is composed of 220 to 240 amino acids and some water molecules may contain close to the chromophore. The chromophore is a part of FPs responsible for the emission of light and the energy difference between two molecular orbitals (occupied and unoccupied) lie within the visible spectrum.
Figure 5 Chromophore structure of mNeptune variants

The FPs chromophore is auto-catalytically formed by the chemical modification of three amino acid residues. In Far RFPs, the amino-acid residues MET- TYR-GLY (positions 63-65) are the tripeptide building block of the chromophore. The chromophore formation of RFPs is a complex process that takes place by chemical mechanisms, which are cyclization, dehydration, and oxidation. At the beginning of the mechanism, nucleophilic reactions occur on carboxyl at position 65. This reaction takes place with the action of amide nitrogen of glycine atom of residue 67, which sets up the dehydration of the system. This process enables the formation of an imidazoline-5-one heterocyclic ring. The imidazoline-5-one heterocyclic ring and tyrosine phenol ring are conjugated as a result of the oxidation of the tyrosine $C_\alpha-C_\beta$ bond. In the case of the RFP chromophore, there is
the next oxidation step with C\textsubscript{α} and amide nitrogen of residue 63. This process also increases the extended \(\pi\)-bonding electron system to attach the carboxyl group of residue 62 [44]. The schematic representation of the chromophore formation is mentioned below with clear steps in Figure 6 [45].

![Figure 6 Chromophore maturation process (adapted from reference 45). [O] and [H] represent the oxidation and reduction respectively.](image)

**3. FIRST PRINCIPLE QUANTUM METHOD**

The classical MD simulation can explain the results in terms of molecular level such as hydrogen bonding, residue interaction, chromophore flexibility, but is unable to explain the results from electronic aspects. Therefore, it was sought to search for a suitable electronic structure method and found a QM as the best fit in multiple scientific research. QM method is a type of computation that provides information about the quantum phenomena such as excited-state energy. One can obtain the electronic level of information from the QM technique.
3.1 MULTI-CONFIGURATION METHOD

The Born-Oppenheimer approximation is defined based on the nuclei-electrons interactions because the nuclei are considerably more massive than electrons, the response of electrons to nuclear motion is instantaneous. It is reasonable to consider the position of the nucleus fixed when calculating the electronic wave functions. As a result, the wave function of the molecule can be separated into nuclear and electrons \[ \Psi_{total} = \Psi_{\text{electronic}} \times \Psi_{\text{nuclear}} \] (3.1)

Also, the nuclear degree of freedom in Hamiltonian are eliminated as

\[ \hat{H} = T_e + V_{ne=ne} + V_{ee=ee} \] (3.2)

Here, \( V \) and \( T \) denote the potential energy and kinetic energy, respectively. Moreover, \( e \) and \( n \) represent electron and nuclear interaction, correspondingly.

From quantum mechanics, the information about a given system is explained in terms of the wave function, \( \Psi \). In condensed matter physics the systems of interest are usually composed of electrons and nuclei. For more than one electron i.e. many-body problems, Schrödinger’s equation becomes:

\[
\left[ \sum_i^N \left( -\frac{\hbar^2 \nabla_i^2}{2m} + V(r_i) \right) + \sum_{i<j} U(r_i, r_j) \right] \psi(r_1, r_2, \ldots, r_N) = E \psi(r_1, r_2, \ldots, r_N) \] (3.3)

Here, \( N \) is the number of electrons and \( U(r_i, r_j) \) is electron-electron interaction. Several powerful methods for solving Schrödinger’s equation for many-body problems have been developed. Due to the Pauli principle, the wave function needs to be antisymmetric under the exchange of any two electrons.

\[
\Psi(\vec{r}_1, \ldots, \vec{r}_n, \ldots, \vec{r}_m, \ldots, \vec{r}_N) = -\Psi(\vec{r}_1, \ldots, \vec{r}_N, \ldots, \vec{r}_m, \ldots, \vec{r}_N) \] (3.4)
This is the problem we need to solve. Clearly, as a $3 \times N_e$ dimensional partial differential equation, this is mathematically a very hard problem. If, however, we were able to solve it, we could then use the wave function to calculate any physical properties $O$ of the system as the expectation value of the corresponding Hermitian operator $\hat{O}$

$$O = \langle \Psi | \hat{O} | \Psi \rangle = \int d^3r_1 \ldots \int d^3r_{N_e} \; \Psi^*(\vec{r}_1, \ldots, \vec{r}_{N_e}) \; \hat{O} \; \Psi(\vec{r}_1, \ldots, \vec{r}_{N_e})$$  \hspace{1cm} (3.5)

This is where density functional theory comes in and provides a shortcut to the solution of our problem: It allows us to calculate expectation values without knowledge of the many-body wave function. It is worth mentioning that The Born- Oppenheimer approximation alone is not enough to solve Schrödinger’s equation, and further explanation and simplification are sought. Slater determinant describes the multi-electronic wave function in quantum mechanics. It arises from the assumption of collection of electrons having wave function known as spin-orbital. It contains a pair of electrons with the same spin-orbital that corresponds to a wave function and zero everywhere. The Hartree-Fock (HF) method approximates the N-body wave function using a single Slater determinant. The $E_x^{HF}$ is Hartree-Fock energy that can be written as,

$$E_x^{HF} = -\frac{1}{2} \sum_{i,j} \int \int \Phi_i^*(r_1)\Phi_j^*(r_1) \frac{1}{r_{12}}\Phi_i(r_2)\Phi_j(r_2) dr_1 dr_2$$ \hspace{1cm} (3.6)

### 3.11 COMPLETE ACTIVE SPACE METHOD

There are different methods identified in solid-state physics to find the approximation for the exact solution. The complete Active Space Self-Consistent Field (CASSCF) method is used to calculate the electronic structure of systems having different configurations. The requirements of the CASSCF method are due to the search of accuracy in our calculation, but this method has to spend a higher computational cost implementing
in Gaussian 09 i.e full configuration interaction (CI). The full CI method expands with the systems with multiconfigurational features to build up the wave function. One such method is CASSCF which is a powerful method for studying the excited states of molecules [47].

In the CASSCF method, the wave function is the linear combination of multiple determinants as,

\[ \Psi = C_0 \Psi_0 + C_1 \Psi_1 + C_2 \Psi_2 + \cdots \]  \hspace{1cm} (3.7)

\[ \Psi_{MCSCF} = \sum K C_K \Psi_K \]  \hspace{1cm} (3.8)

Where coefficient \( C_K \) is the weight of configuration during expansion and normalization.

From equation (3.8), we can say that the wave function has only one configuration as in Hartree-Fock (HF). This method provides the highly descriptive nature of the electronic system. The explanation of one electron and N-electron may create a problem in the study of their physical studies, where the CASSCF method is a powerful process to solve such kinds of issues. The minimum energy at ground state

\[ E_{MCSCF} = \min \frac{\langle \Psi_{MCSCF} | H | \Psi_{MCSCF} \rangle}{\langle \Psi_{MCSCF} | \Psi_{MCSCF} \rangle} \]  \hspace{1cm} (3.9)

The choice of active space in the study of the CASSCF method is an important tool that determines the popularity of wave function.

The first part of CASSCF calculation is the structure of our interest and then the selection of orbitals following the geometry. The correctness of this method depends upon the precise choice of active space. The natural orbital analysis is the best way to pick out the active space and those orbitals are introduced by diagonalization of density matrix [48]. The advantage of using natural orbital is the rapid convergence of CI expansion. Also, natural orbitals form fewer configurations to obtain accurate calculations.
I know that the HF method cannot describe the systems with multi-configuration attributed (having excited state) due to the lack of electron-correlation effect in the wave function. To get the excited state feature, multiple numbers of configurations are combined to correct the wave function and can recover the electronic correlation energy. Mathematically it can be expressed as the linear combination of excited determinants,

$$\Psi_{CI} = C_{HF}\Psi_{HF} + \sum_{a} \sum_{r} C_{a}^{r}\Psi_{a}^{r} + \sum_{b} \sum_{s} \sum_{r} \sum_{s} C_{ab}^{rs}\Psi_{ab}^{rs} + \cdots = \sum_{i} C_{i}\Psi_{i} \quad (3.10)$$

Where a, b are occupied orbitals in HF wave function, r and s are virtual molecular orbitals in HF wave function. The additional term that appears in equation (3.10) is developed due to the excitation of electrons. The electron excitation from the occupied orbital is denoted by subscripts and the virtual orbitals are indicated by superscripts. Therefore, configurational expansion coefficients are variationally optimized as a result of building a single CI method. The ground state configurations are accomplished and energy minimization of CI,

$$E_{CI} = \min_{C} \frac{\langle \Psi_{CI} | H | \Psi_{CI} \rangle}{\langle \Psi_{CI} | \Psi_{CI} \rangle} \quad (3.11)$$

And equivalent eigenvalue equation,

$$H\Psi_{CI} = E_{CI}\Psi_{CI} \quad (3.12)$$

Here, H is the Hamiltonian matrix with elements.

$$H_{KL} = \langle \Psi_{K} | H | \Psi_{L} \rangle = \langle K | H | L \rangle \quad (3.13)$$

K and L are used to expand the arbitrary configurations. The solution of equation (3.13) is comparable to the diagonalization of the Hamiltonian matrix H that allows the calculation of CI coefficients associated with energy. The matrix elements of equation (3.13) can be calculated using Slater determinants as a result of Condon-Slater rules [49].
The active space selection in the CASSCF method is represented as \((N, M)\) with \(N\) number active electrons and \(M\) represents the number of active orbitals. The number of structural configurations is included in the system and that is the function of active space in the CASSCF method. Consider the number of electrons with \(\alpha\)-spin and \(\beta\)-spin as \(N_\alpha\) and \(N_\beta\) with the total number of electrons \(N = N_\alpha + N_\beta\), then the slatter determinant can be written as,

\[
N_{SD} = \binom{M}{N_\alpha} \binom{M}{N_\beta}
\]

(3.14)

The quantity in parentheses are binomial coefficients and obtained as

\[
\binom{M}{N} = \frac{M!}{N!(M-N)!}
\]

(3.15)

The number of configurations in singlet basis and triplet basis is given by,

\[
N_{Singlet} = \binom{M}{N/2} \left( \binom{M}{N/2} + 1 \right) \quad \text{(3.16)}
\]

\[
N_{Triplet} = \binom{M}{N/2} \left( \binom{M}{N/2} - 1 \right) \quad \text{(3.17)}
\]

In the CASSCF method, it is said to be a full CI method, if all orbitals of configuration are chosen as active space and all electrons are taken as active electrons. This method is used to solve the time-independent nonrelativistic Schrödinger equation to get an exact solution with an infinite basis set [50]. So, the set of all possible configurations is satisfied without requiring orbital optimization. But the simplest system may have a large number of configurations and full CI is not able to support such unmanageable configuration which is considered as requiring higher-level theory. The CASSCF is supportive of full CI with some restrictions with varying optimized active orbitals. Also, the performance of CASSCF in Gaussian is poor with higher computational costs, while
using the CI approach. As mentioned above, full CI is possible by taking all orbitals and
electrons of a configuration. This type of calculation is difficult to achieve. To make
CASSCF calculation compatible, the only subset of the configuration is included in the
wave function instead of all configurations. But this method still requires higher
computational cost and is difficult to converge, while including the requiring number of
electrons and orbitals inactive space. Therefore, I am seeking a method that can do the
calculations taking a large number of atoms without mentioning their active space. I found
a DFT method to calculate the ground state energy of the many-electron system.

3.2 DENSITY FUNCTIONAL THEORY

Density-functional theory (DFT) is a computational quantum method used in
physics, chemistry, and materials science to investigate the electronic structure of many-
body systems, in particular atoms, molecules, and condensed phases. Ground state DFT is
a much more mature theory that has enjoyed tremendous success due to exact conditions
obtained from the uniform electron gas. Using this theory, the properties of a many-electron
system can be determined by using functional, i.e. functions of another function. In the
case of DFT, these are functions of the spatially dependent electron density. DFT is among
the most popular and versatile methods available in condensed-matter physics,
computational physics, and computational chemistry.

The DFT method was first introduced in two influential papers in the ’60s [51, 52]. Walter Kohn was awarded the Nobel Prize in Chemistry in 1998 for his development
of the DFT. It has been very popular for calculations in solid-state physics since the 1970s.
However, DFT was not considered accurate enough for calculations in quantum chemistry
until the 1990s, when the approximations used in the theory were greatly refined to better
model the exchange and correlation interactions. Computational costs are relatively low when compared to traditional methods, such as exchange only Hartree–Fock theory and its descendants that include electron correlation. Since, DFT has become an important tool for methods of nuclear spectroscopy such as Mössbauer spectroscopy or perturbed angular correlation, to understand the reason for specific electric field gradients in crystals. Despite recent improvements, there are still difficulties in using density functional theory to properly describe: intermolecular interactions (of critical importance to understanding chemical reactions), especially van der Waals forces (dispersion); charge transfer excitations; transition states, global potential energy surfaces, dopant interactions, and some strongly correlated systems; and in calculations of the bandgap and ferromagnetism in semiconductors. The incomplete treatment of dispersion can adversely affect the accuracy of DFT in the treatment of systems that are dominated by dispersion (e.g. interacting noble gas atoms) or where dispersion competes significantly with other effects (e.g. in biomolecules). The development of new DFT methods designed to overcome this problem, by alterations to the functional or by the inclusion of additive terms is a current research topic.

However, it seems impossible to apply them to large and complex systems. The wave function for a single electron moving in potential energy \( V(r) \) is calculated from non-relativistic Schrödinger's equation:

\[
\hat{H}\psi(r) = \left[-\frac{\hbar^2\nabla^2}{2m} + V(r)\right]\psi(r) = E\psi(r)
\] (3.18)

Here, \( \hat{H} \) is the Hamiltonian operator, \( r \) is the position vector, \( \hbar \) is the reduced Planck constant, \( m \) is particle mass, \( \nabla^2 \) is the Laplacian, and \( E \) is the energy of the state. DFT provides
a way of reducing this problem to three spatial dependencies by presenting precise simplifications and approximations which will be introduced in the later sections.

### 3.2.1 HOHENBERG-KOHN THEOREM

There is two Hohenberg and Kohn theorem in 1964 in the existence of Density Functional Theory (DFT). The external potential is the unique function of the electron density $\rho(r)$ is the statement of the first Hohenberg-Kohn theorem. So, the ground state wave function $\Psi$ and total energy $E[\langle \psi \rangle]$ of a system can be determined by ground-state electron density as

$$E[\langle \psi \rangle] = E[\rho(r)] \quad (3.19)$$

The second Hohenberg-Kohn theorem indicates that the ground-state electron density is the exact density that minimizes the total energy. In another way, if Schrödinger’s equation is solved for the ground-state electron density, the corresponding wave function can be determined In this way, the observable wave function is interpreted by the observable electron density, which is the number of electrons at a specific location is. As a result, the system can be fully described. Although the Hohenberg-Kohn theorems are immensely powerful, the ground state density can’t be computed using them.

### 3.2.2 KOHN-SHAM EQUATION

Kohn-Sham theorem proposed the introduction of an auxiliary system of non-interacting particles to compute the electron density. The Schrodinger equation of a fictitious system of particles in the form of equation (3.2) is given by

$$\left( -\frac{1}{2m} \hbar^2 \nabla^2 + V_{\text{eff}}(r) \right) \psi_\alpha(r) = \varepsilon_\alpha \psi_\alpha(r) \quad (3.20)$$
Here, $\varepsilon_\alpha$ is the orbital energy of the corresponding Kohn-Sham orbital, $\Psi_\alpha$. In addition, $V_{\text{eff}}(r)$ is the Kohn-Sham potential in which the non-interacting particles move and can be written as

$$V_{\text{eff}}(r) = V_H(r) + V_{xc}(r) + V_{\text{ext}}(r) \quad (3.21)$$

The first two terms $V_H(r)$ and $V_{xc}(r)$ are related to the interactions of the electrons (interaction with other electrons). The $V_H(r)$ is the Hartree potential as a result of the mean-field electrostatic interaction, and $V_{xc}(r)$ represents the exchange-correlation potential caused by the quantum mechanical nature of the electrons. The Hartree potential is also termed as the classical electrostatic potential that can be calculated from the electron density as,

$$\nabla^2 V_H(r) = -4\pi \rho(r) \quad (3.22)$$

Also, the exchange-correlation potential is given by,

$$V_{xc}(r) = -\frac{\delta E_{xc}(\rho)}{\delta \rho(r)} \quad (3.23)$$

Here, $E_{xc}(\rho)$ represents the exchange-correlation energy, which will be introduced in detail later. The third term, $V_{\text{ext}}(r)$ is associated with any other electrostatic interactions in the system. This term can be caused by the ion potentials (described by norm-conserving pseudopotential) and electrostatic calculations by replacing the effects of core interactions with the external electrostatic field. Pseudopotential is an attempt to simplify the electrons of an atom with an effective potential. This approximation was first introduced by Hans Hellmann in 1934 [53]. As can be seen in Figure 7 (113) the wave function oscillates rapidly in the core electrons region because of the strong ionic potential. By replacing the real potential (red solid line) with a weaker pseudo-
potential (blue dashed line), a pseudo wave function with a smooth curve is constructed as shown in Figure 7 [54].

Figure 7 Potential and its corresponding wave function of all-electron (blue) compared that of pseudo electron (red)

Here, \( r_c \) is the radius at which all-electron and pseudo-potential values match

As a result, the Kohn-Sham theorem considers a fictitious system of non-interacting particles within an effective potential to study a real system in which particles interact within an external potential [55].

3.2.3 ELECTRON DENSITY

Consider a periodic system consisting of \( N \)-particles. Then the electron density of particles on the occupied eigenstates can be written as,

\[
\rho(r) = \sum_{\alpha=1}^{N} |\Psi_{\alpha}(r)|^2 \, f \left( \frac{\epsilon_{\alpha} - \epsilon_f}{kT} \right)
\]  

(3.24)

Consider the Fermi function \( f(x) = \frac{1}{1 + e^{x}} \), \( \epsilon_f \) is the Fermi energy, \( T \) is electron temperature, and \( k \) is the Boltzmann constant. For our easiness, the electron density can be written in terms of density matrix \( (D_{ij}) \),
\[ \rho (r) = \sum_{\alpha} D_{ij} \phi_i(r) \phi_j(r) \]  \quad (3.25)

In terms of basis set expansion coefficient, \( D_{ij} \) is defined as

\[ D_{ij} = \sum_{\alpha \beta} c_{i\alpha}^* c_{\beta j} f \left( \frac{E_{\alpha} - E_{\beta}}{kT} \right) \]  \quad (3.26)

Moreover, the electron difference density can be calculated by comparing the electron density with a superposition of atom-based densities as,

\[ \Delta \rho (r) = \rho (r) - \sum_{\mu} \rho_{\text{atom}}^\mu (r - r_{\mu}) \]  \quad (3.27)

### 3.2.4 TOTAL ENERGY

The Kohn-Sham potential and electron density are self-consistently calculated up until the electron density which minimizes the total energy is found. The total energy of a system as a function of electron density is expressed as follows:

\[ E(\rho) = T(\rho) + E_H(\rho) + E_{\text{ext}}(\rho) + E_{xc}(\rho) \]  \quad (3.28)

Here, \( T(\rho) \) is the kinetic energy of non-interacting particles and given by,

\[ T_s(\rho) = \sum_{\alpha=1}^N \int dr \psi_{\alpha}^*(r) \left( -\frac{\hbar^2}{2m} \right) \psi_{\alpha}(r) f \left( \frac{E_{\alpha} - E_{\beta}}{kT} \right) \]  \quad (3.29)

Also, the Hartree energy \( E_H(\rho) \) is given by

\[ E_H(\rho) = \frac{e^2}{2} \int dr \int dr' \frac{\rho(r)\rho(r')}{|r-r'|} \]  \quad (3.30)

The energy-related to interactions with the pseudopotential ions and other electrostatic external potential is given by,

\[ E_{\text{ext}}(\rho) = \int dr V_{\text{ext}}(r) \rho (r) \]  \quad (3.31)

### 3.2.5 EXCHANGE-CORRELATION ENERGY

The quantum mechanical effect of the other electrons in the system is included in the exchange-correlation energy term. There are various approximations for the exchange-correlation energy. The choice of the appropriate function greatly depends
on the system at hand. There have been attempts at designing exchange-correlation functional that exclude the so-called self-interaction in multi-electron systems [56, 57]. These methods have so far not become standard, and most approximate exchange-correlation functional in use today suffer from a spurious self-interaction. The oldest and simplest approximation for the exchange-correlation functional is known as the local density approximation (LDA). The exchange-correlational energy depends only on the electronic density at a point in space [58]. The LDA for the exchange-correlation energy is given by

$$E_{xc}^{LDA}[\rho] = \int \rho(r) \epsilon_{xc}(\rho(r)) dr \quad (3.32)$$

Where $\epsilon_{xc}(\rho)$ is the per-particle exchange-correlation energy density of a homogeneous electron gas of density $\rho$, which can be calculated using quantum Monte Carlo methods [59, 60]. While LDA usually gives reasonable results, better approximations for the exchange-correlation functional can be constructed if one not only takes the value of the density into account but also its gradient: The generalized gradient approximation (GGA) [60] is a functional in which both the density and its gradient at each point are considered. With this additional information, GGA is usually more accurate than LDA. The GGA for the exchange-correlation energy is given by

$$E_{xc}^{GGA}[\rho] = \int \rho(r) \epsilon_{xc}(\rho(r), \nabla \rho) dr \quad (3.33)$$

Here, $\epsilon_{xc}(\rho(r), \nabla \rho)$ is the exchange-correlation energy as a function of density and its gradient. These functionals are known as generalized gradient approximation (GGA). More sophisticated functional can be constructed if one, in addition to its gradient, also includes the Laplacian of the density (or of the kinetic energy density), which results in the so-called meta-GGA functional. A review of the classes of exchange-correlation functional and the
methods used to construct them can be found in reference [61]. A special class of functional is the so-called hybrid functional, where the Hartree-Fock exchange is partially used as the exchange portion of the exchange-correlation functional. The correlation from other sources such as ab initio or empirical B3LYP (Becke, 3-parameter, Lee-Yang-Parr) is one of the most frequently used functionals of this type [62, 63]. The exchange-correlation energy of B3LYP functional is expressed as:

\[
E_{xc}^{B3LYP} = E_x^{LDA} + a_0(E_x^{HF} - E_x^{LDA}) + a_x(E_x^{GGA} - E_x^{LDA}) + E_c^{LDA} + a_c(E_c^{GGA} - E_c^{LDA})
\]  

(3.34)

Here, \(a_0, a_x, a_c\) are 0.20, 0.72, and 0.81 respectively. \(E_x^{LDA}\) and \(E_x^{GGA}\) are the exchange energies of LDA-VWN and GGA-Beke 88 [58, 64]. Correspondingly, \(E_c^{LDA}\) and \(E_c^{GGA}\) represent the correlation energy of LDA-VWN and GGA-Beke 88 functionals, respectively. Therefore, The DFT method is enough to calculate the ground-state properties of the system but is unable to show the excited state properties of the same system. We sought another method to explain the higher state than the ground state and the TD-DFT method was found to be suitable to explain the excited states of the many-body system [65, 66].

3.3 TIME-DEPENDENT DENSITY FUNCTIONAL THEORY

Time-dependent density-functional theory (TDDFT) is an extension of the static ground state DFT [67]. It is the basic tool to solve the many-electron problem in quantum mechanics. It is widely used in solid-state physics to get the accuracy between computational efficiency and practical balance [68-70]. The many-electron problem in the TDDFT method covers the different phenomena including optical response properties of atoms, molecules, solids, strong-field ionization physics, and quantum optimal control.
theory. It is difficult to control the computation in the above-mentioned applications to get the full solution. Therefore it should be imposed to approximate the efficient method to resolve this problem. The popularity of the TDDFT method in electronic structure calculation is due to the precise output and is comparable to the experimental value. The theory used in the TDDFT method is real theory and linked to helpful mathematical theorem [71]. Therefore, the study of the simple model problem in TDDFT is used to develop the approximate functional from the known solution, which is supposed to be the key factor in total for the approximation [72-75]. Here, below is the summary of the foundation of TDDFT and its connection with the excited state properties and many-body perturbation theories.

3.3.1 RUNGE-GROSS THEOREM

The foundation of TDDFT is Runge-Gross (RG) theorem. There are more similarities between time-dependent potential and its density. A wave function is equivalent to electron density. So, the many-body wave function is determined by the density of electrons. Time-dependent Schrödinger equation for system is given by

\[
i \frac{\partial}{\partial t} \psi(t) = \hat{H}(t)\psi(t) = \left[ -\frac{\hbar^2}{2m} \nabla^2 + V(r) \right] \psi(t) = E \psi(t)\] (3.35)

At initial condition \( t = 0 \), many-body time-dependent Schrödinger equation can be written as

\[
\psi(t_0) = \psi\rangle \] (3.36)

Which is the starting point of the Schroedinger equation. The RG theorem explains that the electron density from equation (3.36) uniquely determines the external potential.

We are doing the electronic structure calculation of molecules in biophysics. The
Hamiltonian $\hat{H}(t)$ is given by

$$\hat{H} = \hat{T} + \hat{W} + \hat{V}_e(t) \tag{3.37}$$

Where $\hat{T}$ is the kinetic energy operator, $\hat{W}$ the electron-electron interaction, and $\hat{V}_e(t)$ the external potential of interacting electrons and nuclei.

$$=-\frac{1}{2} \sum_{n=1}^{N_e} \nabla_n^2 + \frac{1}{2} \sum_{n \neq m} \frac{1}{|\mathbf{r}_n - \mathbf{r}_m|} + \sum_{n=1}^{N_e} v_{ext}(\mathbf{r}_n, t) \tag{3.38}$$

Here, the external potential $v_{ext}(\mathbf{r}_n, t)$ is measured by taking the sum of the Coulomb potential of nuclei and external electric field.

### 3.3.2 TIME-DEPENDENT KOHN-SHAM EQUATION

The electric field is originate as a result of the perturbation of incoming waves and assuming fixed nuclei. The external potential becomes

$$v_{ext}(\mathbf{r}_n, t) = -\sum_{A=1}^{N_{atoms}} \frac{Z_A}{|\mathbf{r}_A - \mathbf{r}_n|} + f(t) \sin(\omega t) \mathbf{E} \cdot \mathbf{r} \tag{3.39}$$

Where $\mathbf{E}$ is the strength of the electric field in infinite direction and $f(t)$ balances the intensity of the field at time $t$. The magnetic effect in an electromagnetic field is small enough to neglect in comparison to an electric field. Then, the external potential can be written as

$$v_{ext}(\mathbf{r}_n, t) = v_{ext}(\mathbf{r}_n) + \delta v_{ext}(\mathbf{r}_n, t) \tag{3.40}$$

Where $\delta v_{ext}(\mathbf{r}_n, t)$ sufficiently small potential to compare with time-dependent perturbation theory.

Hohenberg-Kohn theorem has been made enormously applicable to the time-dependent problem by RG. Initial state $\psi(t_0)$ is used to set up the one to one correspondence between time-dependent electron density $\rho(\mathbf{r}, t)$ and time-dependent external potential $v_{ext}(\mathbf{r}, t)$ of a system by RG theorem. Also, the Hohenberg-Kohn
Theorem is enough to find the initial state of the system by using density only, if the external potential is time-dependent in beginning. Therefore, the dependence of the initial state is not a real problem as it appears. This covers many experimental procedures where initial state dependence has to be neglected. So Runge-Gross theorem can be written as,

\[ \rho(r, t) \leftrightarrow v_{\text{ext}}(r, t) \leftrightarrow \psi(t) \quad \text{for} \quad v_{\text{ext}} \frac{\partial v_{\text{ext}}(r, t)}{\partial t} \bigg|_{t \leq t_0} = 0 \quad (3.41) \]

We know, functional is the function of a function. It is called electron density \( \rho \) which is the function of space and time. If a state is the functional of electron density \( \rho(r, t) \), the expectation value of any observable

\[ O[\rho](t) = \langle \psi[\rho](t) | \hat{F} | \psi[\rho](t) \rangle \quad (3.42) \]

In TDDFT, there is no principle used to calculate energy like in DFT such as the second Hohenberg-Kohn theorem. Therefore action principle is used to derive the stationary condition as,

\[ A[\rho] = \int_{t_0}^{t} dt' \left( \psi[\rho](t') \left| \frac{\partial}{\partial t'} + \hat{H}(t') \right| \psi[\rho](t') \right) \quad (3.43) \]

At stationary condition

\[ \frac{\delta A[\rho]}{\delta \rho(r, t)} = 0 \quad (3.44) \]

The equation (1.9) can be written as,

\[ A[\rho] = \int_{t_0}^{t} dt' \left( \psi[\rho](t') \left| i \frac{\partial}{\partial t'} + \hat{T} + \hat{W} \right| \psi[\rho](t') \right) + \int_{t_0}^{t} dt' \int d^3r \, \rho(\vec{r}, t') v_{\text{ext}}(\vec{r}, t') \quad (3.45) \]

The kinetic energy term for non-interacting system and Hartree term for time-dependent densities can be obtained with the introduction of universal action functional \( B[\rho] \) and given by,
\[ B[\rho] = \int_{t_0}^{t} dt' \left( \psi[\rho](t') \left| i \frac{\partial}{\partial t'} + \hat{T} + \hat{W} \right| \psi[\rho](t') \right) \]  

(3.46)

With the inclusion of exchange-correlation action functional \( A_{xc}[\rho] \), equation (3.46) becomes,

\[ B[\rho] = -\frac{1}{2} \sum_{i=1}^{N_e} \int d t' \int d^3 r \phi_i^*(\vec{r}, t') \nabla^2 \phi_i(\vec{r}, t') \]

\[ + \frac{1}{2} \int_{t_0}^{t} dt' \int d^3 r \int d^3 r' \frac{\rho(\vec{r}, t) \rho(\vec{r}', t')}{|\vec{r} - \vec{r}'|} + A_{xc}[\rho] \]  

(3.47)

Arguments from equation (3.47) are similar to the ground state. We can interpret the motion of particles the time-dependent effective potential as,

\[ v_{eff}[\rho](\vec{r}, t) = v_{ext}(\vec{r}, t) + \int d^3 r' \frac{\rho(\vec{r}, t)}{|\vec{r} - \vec{r}'|} + \frac{\delta A_{xc}[\rho]}{\delta \rho(\vec{r}, t)} \]

\[ = v_{ext}(\vec{r}, t) + v_H[\rho](\vec{r}, t) + v_{xc}[\rho](\vec{r}, t) \]  

(3.48)

Equation (3.48) generates the density \( \rho(\vec{r}, t) \) same as that of interacting particles. The time-dependent Kohn-Sham equation can be written as,

\[ i \frac{\partial \phi_i(\vec{r}, t)}{\partial t} = \left[ -\frac{1}{2} \nabla^2 + v_{eff}[\rho](\vec{r}, t) \right] \phi_i(\vec{r}, t) \]  

(3.49)

Hartree potential \( v_H[\rho](\vec{r}, t) \) in equation (3.48) depends on density at time \( t \) and is independent of the history of electron density. But exchange-correlation functional \( A_{xc}[\rho] \) entirely depends upon the history of electron density. For slowly varying densities, exchange-correlation action functional \( A_{xc}[\rho] \) can be written as the exchange-correlation functional \( E_{xc} [\rho] \).

\[ \lim_{\delta \rho \to 0} \frac{\partial}{\partial \tau} A_{xc} [\rho] = \int_{t_0}^{t} dt' E_{xc} [\rho(t')] \]  

(3.50)
It is assumed that densities are slowly varying to be a precise approximation which is the adiabatic approximation. This approximation works as an exact in the limit of adiabatically changing densities. The $\rho(\vec{r}, t)$ approximated at time $t'$ as $\rho_{t'}$. The functional derivatives of $A_{xc}[\rho]$ in equation (4.48) as a result of adiabatic approximation,

$$v_{xc}[\rho](\vec{r}, t) = \left. \frac{\delta A_{xc}[\rho]}{\delta \rho(\vec{r}, t)} \right|_{\rho_{t}} \approx \left. \frac{\delta E_{xc}[\rho_{t}]}{\delta \rho(\vec{r})} \right|_{\rho_{t}} = v_{xc}[\rho_{t}](\vec{r})$$ (3.51)

Equation (4.51) generates the effective potential $v_{eff}[\rho](\vec{r}, t)$ with the help of density at time, but without the history of electron density. The adiabatic approximation establishes a smaller error than other approximations in actual practice \[76\]. If we compare with local density approximation for exchange-correlation, it becomes exact for spatially uniform density, but not like real systems like solids.

### 3.3.3 DENSITY RESPONSE FUNCTION

Consider external potential $v_{ext}$ that can be expressed as a time-independent potential and a time-dependent perturbation.

$$v_{ext}(\vec{r}, t) = v_{ext}^0(\vec{r}) + \delta v_{ext}(\vec{r}, t)$$ (3.52)

Similarly, the time-dependent electron density is written as the sum of electronic density without perturbation and the same density due to perturbation.

$$\rho(\vec{r}, t) = \rho^0(\vec{r}) + \delta \rho(\vec{r}, t)$$ (3.53)

Consider the response function $\chi(\vec{r}, \vec{r}', t - t')$ with a small change in external potential $\delta v_{ext}(\vec{r}', t')$ so that this potential changes the electron density by a small amount $\delta \rho(\vec{r}, t)$.

$$\delta \rho(\vec{r}, t) = \int dt' \int d^3r' \chi(\vec{r}, \vec{r}', t - t') \delta v_{ext}(\vec{r}', t')$$ (3.54)

$$\chi(\vec{r}, \vec{r}', t - t') = \frac{\delta \rho(\vec{r}, t)}{\delta v_{ext}(\vec{r}', t')}$$ (3.55)
Because of the assumption of uniform physics laws, response function depends on the time difference, $t - t'$ and it is independent of separate variations in, $t$ and $t'$. Let us construct the density $\rho$, and density change $\delta \rho$ are equal in real and Kohn-Sham system, where the response function $\chi$ of the real system is unknown that is unfortunate. So, density change of non-interacting electrons is given by Kohn-Sham as,

$$
\delta \rho(\vec{r}, t) = \int dt' \int d^3r' \chi_{KS}(\vec{r}, \vec{r}', t - t') \delta v_{eff}(\vec{r}', t') \tag{3.56}
$$

$$
\chi_{KS}(\vec{r}, \vec{r}', t - t') = \frac{\delta \rho(\vec{r}, t)}{\delta v_{eff}(\vec{r}, t)} \tag{3.57}
$$

By substituting individual terms of Kohn-Sham effective potential, we get

$$
\delta \rho(\vec{r}, t) = \int dt' \int d^3r' \chi_{KS}(\vec{r}, \vec{r}', t - t')[\delta v_{ext}(\vec{r}', t') + \delta v_{H}(\vec{r}', t') + \delta v_{xc}(\vec{r}', t')] \tag{3.58}
$$

Now, the chain rule for the functional derivative is used to evaluate the changes in the Hartree potential and exchange-correlation potential using density change $\delta \rho$.

$$
\delta v_{H}(\vec{r}, t) = \int d^3r' \frac{\delta v_{H}(\vec{r}, t)}{\delta \rho(\vec{r}, t)} \delta \rho(\vec{r}', t) = \int d^3r' \frac{1}{|\vec{r} - \vec{r}'|} \delta \rho(\vec{r}, t) \tag{3.59}
$$

$$
\delta v_{xc}(\vec{r}, t) = \int d^3r' \int dt' \frac{\delta v_{xc}(\rho)(\vec{r}, t)}{\delta \rho(\vec{r}, t')} \delta \rho(\vec{r}', t') \approx \int d^3r' \frac{\delta^2 v_{xc}(\rho^0)}{\delta \rho(\vec{r}) \delta \rho(\vec{r'})} \delta \rho(\vec{r}', t) \tag{3.60}
$$

$$
= \int d^3r' f_{xc}(\vec{r}, \vec{r}') \delta \rho(\vec{r}', t) \tag{3.61}
$$

Here, $\rho^0$ is the unperturbed density that replaces the first-order density $\rho^t$. $f_{xc}(\vec{r}, \vec{r}')$ is the time-dependent exchange-correlation kernel.

$$
\delta \rho(\vec{r}, t) = \int dt' \int d^3r' \chi_{KS}(\vec{r}, \vec{r}', t - t') \left[ \delta v_{ext}(\vec{r}', t') + \right.
$$

$$
\int d^3r'' \left( \frac{1}{|\vec{r} - \vec{r}'|} + f_{xc}(\vec{r}, \vec{r}'') \right) \delta \rho(\vec{r}'', t') \right] \tag{3.62}
$$

Using convolution theorem

$$
h(t) = \int dt' g(t - t')f(t') \Leftrightarrow h(\omega) = g(\omega)f(\omega) \tag{3.63}
$$
Therefore, density response in the frequency domain

\[
\delta \rho(\vec{r}, \omega) = \int d^3 r' \chi_{KS}(\vec{r}, \vec{r}', \omega) \left[ \delta v_{ext} \left( \vec{r}', \omega \right) + \int d^3 r'' \left( \frac{1}{|\vec{r}' - \vec{r}'|} + f_{xc} (\vec{r}', \vec{r}'') \right) \delta \rho(\vec{r}'', \omega) \right]
\]  

(3.64)

#### 3.3.4 PERTURBATION THEORY

The perturbation in the Hamiltonian of the real system is calculated with the help of the first-order time-dependent perturbation theory.

\[
\delta \hat{H}(t) = e^{\eta t} \sum_{n=1}^{N_e} \delta v_{ext} (\vec{r}_n, t)
\]  

(3.65)

Equation (4.65) shows the adiabatic approximation with \(0 < \eta \leq 1\) and gently changing from unperturbed Hamiltonian at \(t = -\infty\). From the definition of the density operator

\[
\hat{\rho}(\vec{r}) = \sum_{i=1}^{N_e} \delta (\vec{r}_i - \vec{r})
\]

and Fourier transform \((f(w) = \int dt \ e^{i\omega t} f(t))\) and

\[
\frac{1}{2\pi} \int dw \ e^{i\omega t} f(w)\),
\]

\[
\delta \hat{H}(t) = \int d^3 r \int \frac{dw}{2\pi} e^{-i\omega t} \delta v_{ext}(\vec{r}, \omega) \hat{\rho}(\vec{r})
\]  

(3.66)

Assuming \(\tilde{\omega} = \omega + i\eta\)

The total Hamiltonian due to perturbation

\[
\hat{H}(t) = \hat{H}^0 + \delta \hat{H}(t)
\]  

(3.67)

The time-dependent Schroedinger equation is given by

\[
i \frac{\partial \psi(t)}{\partial t} = \left( \hat{H}^0 + \delta \hat{H}(t) \right) \psi(t)
\]  

(3.68)

The solution of equation (3.68) in terms of time-dependent perturbation is given by

\[
\psi(t) = \sum_i a_i(t) e^{-iE_i t} \Psi_i^0
\]

(3.69)

Where \(\hat{H}^0 \Psi_i^0 = E_i \Psi_i^0\). The expansion coefficient \(a_i(t)\) can be written as
\[ a(t) = a(-\infty) - i \sum_{\gamma} \int_{-\infty}^{t} dt' \langle \Psi_{\gamma}^0 | \delta \hat{H}(t') | \Psi_{\gamma}^0 \rangle a_{\gamma}(t') e^{-i(E_{\gamma} - E) t'} \]  

(3.70)

The expression for the coefficient \( a(t) \) is looking complex for its use. If we further explore with the help of first-order approximation, we get

\[ a(t) \approx a(-\infty) - i \sum_{\gamma} \int_{-\infty}^{t} dt' \langle \Psi_{\gamma}^0 | \delta \hat{H}(t') | \Psi_{\gamma}^0 \rangle \delta_{\gamma} e^{-i(E_{\gamma} - E) t'} \]  

(3.71)

A system with unperturbed Hamiltonian is in the ground state \( \psi_{0}^0 \) at \( t = -\infty \). Then coefficient \( a_{\gamma}(-\infty) = \delta_{\gamma 0} \). Then

\[ a(t) = \delta_{\gamma 0} - i \sum_{\gamma} \int_{-\infty}^{t} dt' \langle \Psi_{\gamma}^0 | \delta \hat{H}(t') | \Psi_{\gamma}^0 \rangle \delta_{\gamma 0} e^{-i(E_{\gamma} - E) t'} \]  

(3.72)

\[ = \delta_{\gamma 0} - i \int_{-\infty}^{t} dt' \langle \Psi_{\gamma}^0 | \delta \hat{H}(t') | \psi_{0}^0 \rangle e^{i \Delta t \tau} \]  

(3.73)

Where \( \Delta t = E_{t} - E_{j} \) is the excitation energy due to perturbation. By considering the first non-zero term, we get

\[ a_{0}(t) \approx 1 \quad \text{and} \quad a_{l}(t) = -i \int_{-\infty}^{t} dt' \langle \psi_{l}^0 | \delta \hat{H}(t') | \psi_{0}^0 \rangle e^{i \Delta t \tau} \quad \text{for } l \neq 0 \]  

(3.74)

Combining equations (3.66) and (3.74), we get

\[ a_{l}(t) = -i \int_{-\infty}^{t} \int \frac{d^3 r}{2\pi} \frac{d \omega}{2\pi} \delta v_{ext}(\vec{r}, \omega) \langle \Psi_{l}^0 | \hat{\rho}(\vec{r}) | \Psi_{0}^0 \rangle e^{i \Delta t \tau} \]  

(3.75)

\[ = -i \int_{-\infty}^{t} dt' \int \frac{d^3 r}{2\pi} \frac{d \omega}{2\pi} \delta v_{ext}(\vec{r}, \omega) \langle \Psi_{l}^0 | \hat{\rho}(\vec{r}) | \Psi_{0}^0 \rangle e^{i \Delta t \tau} \]  

(3.76)

\[ = -i \int \frac{d^3 r}{2\pi} \frac{d \omega}{2\pi} \delta v_{ext}(\vec{r}, \omega) \langle \Psi_{l}^0 | \hat{\rho}(\vec{r}) | \Psi_{0}^0 \rangle \frac{e^{i \Delta t \tau}}{\Delta \tau} \]  

(3.77)

Value of square bracket at \( -\infty \) is zero that converts the perturbation adiabatically. So equation (3.77) becomes

44
\[ \Psi(t) = e^{-iE_0t}\Psi_0^0 + \sum_{l \neq 0} a_l(t)e^{-iE_lt}\Psi_l^0 \]  
(3.78)

So, change in electronic density is given by

\[ \delta \rho(\vec{r}, t) = \langle \Psi(t)|\hat{\rho}(\vec{r})|\Psi(t)\rangle - \langle \Psi_0^0|\hat{\rho}(\vec{r})|\Psi_0^0 \rangle \]  
(3.79)

\[ = (e^{-iE_0t}\Psi_0^0 + \sum_{l \neq 0} a_l(t)e^{iE_lt}\Psi_l^0)\hat{\rho}(\vec{r})(e^{-iE_0t}\Psi_0^0 + \sum_{l \neq 0} a_l(t)e^{-iE_lt}\Psi_l^0) - \langle \Psi_0^0|\hat{\rho}(\vec{r})|\Psi_0^0 \rangle \]  
\[ \langle \Psi_0^0|\hat{\rho}(\vec{r})|\Psi_0^0 \rangle \]  
(3.80)

\[ = \sum_{l \neq 0} [a_l(t)e^{-i\Delta_t} \langle \Psi_0^0|\hat{\rho}(\vec{r})|\Psi_l^0 \rangle + a_l(t)e^{i\Delta_t} \langle \Psi_l^0|\hat{\rho}(\vec{r})|\Psi_0^0 \rangle] \]  
(3.81)

Using approximation to first order, the terms containing the product \( a_l(t)a_l(t) \) have

neglected, but both \( l \) and \( \neq 0 \). From equations (3.77) and (3.80), we get

\[ \delta \rho(\vec{r}, t) = \int d^3r' \int \frac{dw}{2\pi} \sum_{l \neq 0} \delta v_{ext}(\vec{r'}, \omega) \langle \Psi_0^0|\hat{\rho}(\vec{r'})|\Psi_0^0 \rangle \frac{e^{-i\vec{w}\cdot \vec{r'}}}{\vec{w}-\Delta_t} \langle \Psi_0^0|\hat{\rho}(\vec{r})|\Psi_0^0 \rangle + \]  
\[ \sum_{l \neq 0} \delta v_{ext}^*(\vec{r'}, \omega) \langle \Psi_0^0|\hat{\rho}(\vec{r'})|\Psi_l^0 \rangle \frac{e^{i\vec{w}\cdot \vec{r'}}}{\vec{w}+\Delta_t} \langle \Psi_l^0|\hat{\rho}(\vec{r})|\Psi_0^0 \rangle \]  
(3.82)

With the substitution of integration variable in the second term, \( \vec{w}^* = \vec{w} - i\eta \) becomes \(-\vec{w} - i\eta = -\vec{w}\). From Fourier transform, \( \delta v_{ext}(\vec{r'}, -\omega) = \delta v_{ext}^*(\vec{r'}, \omega) \). Then

\[ \delta \rho(\vec{r}, t) = \int d^3r' \int \frac{dw}{2\pi} \delta v_{ext}(\vec{r'}, \omega) \ e^{-i\vec{w}\cdot \vec{r'}} \sum_{l \neq 0} \left[ \frac{\langle \Psi_0^0|\hat{\rho}(\vec{r'})|\Psi_0^0 \rangle}{\vec{w}-\Delta_t} \frac{\langle \Psi_0^0|\hat{\rho}(\vec{r})|\Psi_l^0 \rangle}{\vec{w}+\Delta_t} \right] \]  
(3.83)

So, the density response in the form of frequency-domain becomes

\[ \delta \rho(\vec{r}, \omega) = \int d^3r' \delta v_{ext}(\vec{r'}, \omega) \sum_{l \neq 0} \left[ \frac{\langle \Psi_0^0|\hat{\rho}(\vec{r'})|\Psi_0^0 \rangle}{\vec{w}-\Delta_t} \frac{\langle \Psi_0^0|\hat{\rho}(\vec{r})|\Psi_l^0 \rangle}{\vec{w}+\Delta_t} \right] - \]  
\[ \frac{\langle \Psi_0^0|\hat{\rho}(\vec{r'})|\Psi_l^0 \rangle}{\vec{w}+\Delta_t} \frac{\langle \Psi_l^0|\hat{\rho}(\vec{r})|\Psi_0^0 \rangle}{\vec{w}-\Delta_t} \]  
(3.84)

We are looking for response function \( \chi(\vec{r}, \vec{r'}, \omega) \). The external potential change \( \delta v_{ext}(\vec{r'}) \) at \( \vec{r'} \) induces a change in electronic density \( \delta \rho(\vec{r}) \) that sets up response function \( \chi(\vec{r}, \vec{r'}, \omega) \).
\[
\chi(\vec{r}, \vec{r}', \omega) = \sum_{l \neq 0} \left[ \frac{\langle \Psi_0^0 | \hat{\rho}(\vec{r}') | \Psi_0^0 \rangle \langle \Psi_0^0 | \hat{\rho}(\vec{r}) | \Psi_0^0 \rangle}{\omega - \Delta_l} - \frac{\langle \Psi_0^0 | \hat{\rho}(\vec{r}') | \Psi_I^0 \rangle \langle \Psi_I^0 | \hat{\rho}(\vec{r}) | \Psi_0^0 \rangle}{\omega + \Delta_l} \right]
\]

(3.85)

In the above equation (3.85), we don’t know the ground state wave function, excited state wave function, and excitation energies \( \Delta_l \) of the interacting system. So, the response function is hard to calculate. To solve this problem, the calculation should be the type of non-interacting Kohn-Sham system which is possible by replacing external potential \( \delta v_{\text{ext}} \) by effective potential \( \delta v_{\text{eff}} \). The ground state wave function is calculated from the Slater determinant of the occupied Kohn-Sham orbitals \( \phi_i(\vec{r}) \) and excited state is determined by moving electrons from occupied orbitals into unoccupied orbitals [77, 78]. The excitation energies are calculated from orbital energy difference as \( \Delta_I = \epsilon_a - \epsilon_i \) where indices \( a,b \) are used for unoccupied orbitals, and \( I,j \) are used for occupied orbitals. Therefore, the TDDFT method is enough to evaluate to find out the calculation of electronic excited and emission energy/wavelength.

4. MOLECULAR DYNAMICS SIMULATION RESULTS AND DISCUSSION

I prepared configuration files required for simulations using the CHARMM-GUI input generator and visual molecular dynamics (VMD) tool. The missing residues in the mNeptune crystal structure were added to obtain the full protein structure. Necessary mutations were introduced in mNeptune to obtain the fluorescent proteins (mNeptune1, mNeptune2, mNeptune2.5, mCardinal, mCardinal1, mCardinal2, and mKate). These variants are acting as the initial structure of proteins. Normally, the initial structure is an x-ray crystallographic structure including a finite number of water molecules. These structures were solvated using the solvate plugin in the VMD package. The dimension of the simulation box was set up as 80 Å x 80 Å x 80 Å and the box cut off used as 10 Å. The
solvated systems of each variant were electrically neutralized by randomly adding six Na$^+$ and six Cl$^-$ using the VMD auto ionize plugin. Non-bonded cut-off of 12 Å was established for all systems where the Particle Mesh Ewald technique was introduced for long-range interactions. Line search algorithm and conjugate gradient were used for energy minimization. The systems were heated from 20K to 300K using a linear gradient of 20K/ps. The systems were equilibrated at 300K for 15 ps. The time step of 2 fs is used in NVT (constant number, volume, and time) ensemble during equilibration. The temperature of 300K was established for each system using Langevin dynamics. The equilibrated protein systems were subjected to the production run of 100 ns by using Nanoscale Molecular Dynamics (NAMD) simulation package. The MD simulations were done with time steps of 2 fs for hydrogen bond analysis.

I investigated the dynamics of the hydrogen bond network around the chromophore (NRQ) of far RFPs mNeptune1, mNeptune2.5, and mCardinal2 at different temperatures. At a temperature of 300K, chromophore makes a direct hydrogen bond to residues Arginine (Arg, R) and Lysine (Lys, K), R92, and K67 in far RFPs. Chromophore forms both direct hydrogen bond and water-mediated hydrogen bond with the nearby residues. The percentage of frames involved in MD simulation is shown in Table 1, Table 2, and Table 3 for the mNeptune variants. Two charges of Arg residues are detected to be close to the chromophore, R92, and R197. The relative distance of both Arg residues from the chromophore is nearly the same. Also, residue R197 plays the main role in the energetics of the excitation of the chromophore. The R92 is well oriented to interact with the p-system of the chromophore and stable interaction with chromophore at 300K. It is seen that residues R197 and chromophore ring (chromophore moiety including O2, O3,
N2, N3) interact through a water-mediated hydrogen bond. There are also other residues close to the imidazolinone of a chromophore such as Glutamic acid (E) E145 and K67.

![Chromophore and β strand](image)

*Figure 8 Red Fluorescent Protein*

These two residues have strong interaction with R197 and contribute to the flexibility of the chromophore. Notably, the relative orientation of the guanidinium groups of both Arg residues (R92 and R197) concerning the chromophore is preserved throughout the 100 ns of simulation, which reveals a stable and strong interaction and is active in each case. From trajectory visualization, it is concluded that R92 is interacting strongly with the oxygen of the imidazolinone moiety in the case of far RFPs mNeptune1, mNeptune2.5, and mCardinal2. The amine group of residue R197 interacts with the oxygen (OE1 and OE2) of E145, and the interaction remains stable. Also, the residue K67 orients in such a way that it is bonded to O2 and O3 at different times. K67 (NZ) interacts with the stable
combination of R197 - E145 (OE1 and OE2 of E145). Because of the strong interaction of R197 - E145 and K67, the chromophore is unable to make a bond with these nearby residues.

At a low temperature of 77K, the mobility of atoms is slow and form stable direct bond R90: NH2 – NRQ63: O2 and water-mediated hydrogen bond K67: N - NRQ63: O3 throughout the trajectory in far RFPs mNeptune1, mNeptune2.5, and mCardinal2. The interactions among residues K67, E145, and R197 also exist in far RFPs at a low temperature of 77K. With an increase in temperature, the mobility of atoms increased. At 196K, the residues R90 and K67 form a direct hydrogen bond with chromophore at O2 in far RFP mNeptune1 and mNeptune2.5, but not in mCardinal2. It was observed that the water-mediated hydrogen bond is reappeared with chromophore after 80ns of simulation in the case of far RFP mCardinal2. Also, the interaction of R197-E145 in all far RFPs becomes stronger and more stable. At a temperature of 240K, both R92 and K67 form a direct hydrogen bond with NRQ63:O2 in far RFPs mNeptune1, mNeptune2.5, and mCardinal2. At the low temperature of 77K and 196K, the residue R90 makes a direct hydrogen bond with the chromophore and greater than 90% in far RFPs as shown in tables below.

There is also a water-mediated hydrogen bond K67: NZ – NRQ63: O3, which is greater than 60% at the temperature of 300K in mNeptune2.5. There is also a water-mediated hydrogen bond between chromophore and K67 and the percentage of this bond decreases with an increase in temperature in mNeptune1. The percentage of the water-mediated hydrogen bond between the chromophore and K67 decreases with an increase in temperature in the case of mCardinal2, but there is no trend seen in the following tables for
mNeptune1 and mNeptune2.5. After 45 ns of simulation, the direct hydrogen bond K67: NZ – NRQ63: O2 weakens and another strong direct hydrogen bond NRQ63: O2 – R92: NH2 is formed in mNeptune2.5. Above 300K, the residue K67 interacts strongly with NRQ63: O2 rather than with E145, but there is also the presence of a large number of water molecules nearby NRQ in the case of mNeptune2.5 and mCardinal2 at 340K. At this time, the interaction R197 - E145 and NRQ63: O2 – R92: NH2 is seen for all three far RFPs. The concentration of water molecules increases within the chromophore at temperatures 340K and 400K. It is observed that the residues R197, K67, and Q106 interact through a water-mediated hydrogen bond with the chromophore as shown in Figure 9 below.

![Figure 9](image)

*Figure 9 Chromophore-hydrogen bond with nearby amino acid residues*

It is concluded that Arg197 stabilized strongly the anionic form of the chromophore but did not influence the energetic balance of the excited state stabilization. Surprisingly, Arg197 possesses a net +1 charge. Because of position and orientation, Arg197 can interact normally with both rings and the methane-bridge of the chromophore. The amine group of residue R197 interacts with the oxygen (OE1 and OE2) of E145 and
the interaction becomes stable. In some of the far RFPs, water molecules have directly interacted with the chromophore and Arg197 is not able to interact directly. Also, the residue K67 orients in such a way that it is bonded to O2 and O3 in a different frame. K67 (NZ) interacts with the stable combination of R197 - E145 (OE1 and OE2 of E145). Because of the strong interaction of R197 - E145 and K67, the chromophore is unable to make a bond with these nearby residues. For each mutant, bonds with sidechains R92, K67, S28, Q39, and Q106 are reported for each 62-66 sub-state, direct and H2O-mediated, normalized by their respective occupancy. For all entries, 3.5 Å distance and 30° angle cutoff values were used for determining the H-bonds.

4.1 mNEPTUNE1

![Figure 10](image-url)

*Figure 10* Direct hydrogen bond and water-mediated hydrogen bond between extended chromophore (62) of mNeptune1 and nearby amino acid residues

The structure of mNeptune1 is obtained by introducing a point mutation M146T in the crystal structure of mNeptune (PDB ID 3IP2). The most important hydrogen bonds pertinent to the Stokes shift are both direct hydrogen bond and water-mediated hydrogen bond between Phenylalanine (Phe, F) and Glutamine (Gln, Q), F62: O – Q106: NE2 and F62: O - water - Q106: NE2. Here, the oxygen atom in the extended chromophore (F62) is called acylimine oxygen and is responsible to produce the Stokes shift. My simulations at
show that water-mediated hydrogen bond F62: O - water - Q106: NE2 is seen in between 50 ns and 80ns of simulation for this variant. The extended chromophore shows both water-mediated and direct hydrogen bonding F62 – water - Q39 (Figure 10c) and F62 – S66 (Figure 10a), where S stands for Serine amino-acid residue. The Q39 – S66 hydrogen bond plays an important role in orienting the S66 side chain for its interaction with F62. The orientation develops a Q39 water-mediated hydrogen bond with the carbonyl oxygen (Figure 10c). Similarly, there is also a water-mediated H-bond between the carbonyl oxygen and the S66 side chain (not shown in the figure). In other orientations, the N atom of S66 interacts with the oxygen (OE1) of Q106 via a direct hydrogen bond. Similarly, oxygen (OG1) of the same residue interacts directly with the oxygen (OE1) of Q39.

Table 2 H-bonds to chromophore from neighboring sidechains are presented in terms of time fraction as calculated from MD simulations in mNeptune1.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mNeptune1</td>
<td>D</td>
<td>W</td>
<td>D</td>
<td>W</td>
<td>D</td>
<td>W</td>
<td>D</td>
</tr>
<tr>
<td>77k</td>
<td>0</td>
<td>99.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>196k</td>
<td>0</td>
<td>92.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>43.1</td>
</tr>
<tr>
<td>240k</td>
<td>0</td>
<td>85.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>50.6</td>
</tr>
<tr>
<td>300k</td>
<td>0.3</td>
<td>11.8</td>
<td>0</td>
<td>2.9</td>
<td>0</td>
<td>0.1</td>
<td>4.0</td>
</tr>
<tr>
<td>340k</td>
<td>0.2</td>
<td>5.6</td>
<td>0</td>
<td>1.0</td>
<td>25.2</td>
<td>0</td>
<td>51.3</td>
</tr>
</tbody>
</table>

Both structures result in an extended network 62 – 106 – 66 – 39, where a more frequent hydrogen bond between F62 and S66 is seen. The percentage of this bond increases above 300K as shown in Table 2. There is stable interaction between Q39 – S66 and this combination interacting with carbonyl oxygen and the Q106 side chain. Finally, 62 – 66 – 39 is a more stable network seen in the trajectory. At low temperatures (below 300K), the atomic vibration is lower and the mobility of amino acid residues is not enough to form a direct hydrogen bond between them. Therefore, there is no interaction of
acylimine oxygen directly to other residues, but a water-mediated hydrogen bond is seen between acylimine oxygen and Q106, which is decreased with an increase in temperature as in Table 2. Above 300K, the mobility of both amino acid residues and water molecules increases, and the percentage of direct hydrogen bond increases. Because of higher mobility, water molecules reaching to phenol ring also reaches towards the acylimine region. These highly mobile water molecules can make water-mediated hydrogen bonds between the acylimine region and other amino acid residues in a few frames and are difficult to note in our calculation.

Figure 11 Distance time trajectories between the extended chromophore and nearby amino-acid residues of mNeptune1

Figure 11 presents the time series trajectories from MD simulations from 0 to 100 ns. The $\Delta r$ is the distance between the participating atom of different residues side chain and N-acylimine oxygen of residue 62. There is a change in hydrogen bond from
direct hydrogen bond states to the water-mediated bond between time 20 ns to 40 ns as shown above in Figure (11b). The residue pair spends most of their time in water-mediated bonds up to 40 ns, after that, they are in both states in Figure (11c) and Figure (11e).

4.2 mNEPTUNE2.5

\[ \text{Figure 12 Direct hydrogen bond and water-mediated hydrogen bond between extended chromophore of mNeptune2.5 and nearby amino-acid residues} \]

The structure of mNeptune2.5 is obtained from the geometry of mNeptune by introducing point mutations M146T, M11T, S28H, and G41N. Because of these mutations, the equilibrium structure has a significant structural rearrangement that takes place. For example, my preliminary simulations at 300K showed that the water-mediated hydrogen bond that initially exists between F62: O – Q106: NE2 disappears after 30ns, but water-mediated hydrogen bonds are formed between F62: O – HSD28: NE2 and between F62: O – Q39: NE2. Similarly, both direct hydrogen bond and water-mediated hydrogen bond are seen between F62: O – S66 and may have an enhanced Stokes shift at 300K. Also the percentage of the direct hydrogen bond between F62: O – S66 increases with an increase in temperature. The rotation of residue S66 changes the bonding pattern of residues
involved in long network 62 – 66 – 39 - 41 with the carbonyl oxygen and to itself, which also exists at other temperatures. The side chain (both amine and OE1) of Q39 interacts directly with S66: OG and is again involved in a water-mediated hydrogen bond with carbonyl oxygen together with the residues Q39 and HSD28 in the same frame as shown in Figure 9. The direct hydrogen bond between the side chain of N41 and Q39 is stable and is the path of the long network throughout the simulation. This stable bond tends to make the chromophore environment more flexible in comparison to the variant mNeptune1.

**Table 3** H-bonds to chromophore from neighboring sidechains are presented in terms of time fraction as calculated from MD simulations in mNeptune2.5

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mNeptune2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>W</td>
<td>D</td>
<td>W</td>
<td>D</td>
<td>D</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>77K</td>
<td>0</td>
<td>88.2</td>
<td>0</td>
<td>99.9</td>
<td>0</td>
<td>47.2</td>
<td>0</td>
</tr>
<tr>
<td>196K</td>
<td>0</td>
<td>75.9</td>
<td>0</td>
<td>78.1</td>
<td>0</td>
<td>18.4</td>
<td>0</td>
</tr>
<tr>
<td>240K</td>
<td>0.2</td>
<td>75.8</td>
<td>0</td>
<td>78.1</td>
<td>0</td>
<td>65.7</td>
<td>0</td>
</tr>
<tr>
<td>300K</td>
<td>0.1</td>
<td>2.1</td>
<td>0.5</td>
<td>57.4</td>
<td>0.1</td>
<td>18.3</td>
<td>1.6</td>
</tr>
<tr>
<td>340K</td>
<td>0.1</td>
<td>3.0</td>
<td>0</td>
<td>33.0</td>
<td>0</td>
<td>19.2</td>
<td>0</td>
</tr>
</tbody>
</table>

Below 300K, the atomic vibration is not enough to make the direct hydrogen bond between the amino acid residues because of their lower mobility. At 240K, the bond F62: O –water - Q106: NE2 exists until 60ns, but the F62: O – water - HSD28: NE2 is seen throughout the trajectory and also observable at 196K. The percentage of this bond decreases with an increase in temperature as shown in Table 3. Also, the percentage of the direct hydrogen bond between acylimine oxygen and Q106 increases with an increase in temperature. Above 300K, the residues are moving apart and more water molecules are moving around the N-acylimine region and stay in a few frames. The mobility of both atoms and water molecules is higher. Therefore, there is the possibility of having both hydrogen bond states with the acylimine region. The R92 is interacting strongly with the
oxygen of the imidazolinone moiety as in mNeptune1. There is also stable interaction between residues R197 - E145 and to the chromophore as in mNeptune1. The side chain of K67 also interacts with E145 and has both direct hydrogen bond and water-mediated hydrogen bond with chromophore moiety that increases the chromophore flexibility.

Figure 13 Distance time trajectories between the extended chromophore and nearby residues of mNeptune2.5

In the MD time series trajectory, the interaction of F62O-H28 pairs spent most of their time in the water-mediated hydrogen bond, but the residue pair F62O-S66 spent in both states and their population is interconvertible as shown above in Figure (13d) and (13e). Therefore, this pair interaction has greater chances of getting the Stokes shift in comparison to other pair interactions.
4.3 mCARDINAL2

Figure 14 Direct hydrogen bond and water-mediated hydrogen bond between extended chromophore of mCardinal2 and nearby amino-acid residues

The structure of mCardinal2 is obtained from the geometry of mNeptune by introducing point mutations M146T, A104V, I121L, I171H, S28T, G41Q, S143T, N71K, T73P, Q74K, and V218E. Because of a larger number of mutations, the simulation time will have to be extended for the conformation to stabilize and be fully equilibrated. At low temperatures, amino acid residue possesses smaller energy and hence lesser mobility. Therefore the acylimine oxygen has weaker interaction with the nearby residues. But there is a water-mediated hydrogen bond of residues Q106 and Q41 with the acylimine oxygen.
and is tabulated in Table 4. At the temperature of 240K, the vibration of atoms starts to increase and both direct hydrogen bond and water-mediated hydrogen bond between the pair (I) F62O – Q106, (II) F62O – Q41, and (III) F62O – S66 is formed. From Table 4, it is seen that the percentage of the water-mediated hydrogen bond between F62O – Q106, and F62O – Q41 decreases with an increase in temperature. Similarly, the percentage of the direct hydrogen bond between F62O – Q41 decreases with an increase in temperature.

The simulation at 300K shows that there are strong interactions between T28 – Q41 and remain stable for a long time. The orientation of Q41 changes the bonding pattern between the S66 side-chain and acylimine oxygen, which is a water-mediated hydrogen bond

For each mutant, bonds with sidechains R92, K67, T28, Q39, Q41, and Q106 are reported for each 62-66 sub-state, direct and H2O-mediated, normalized by their respective occupancy. For all entries, 3.5 Å distance and 30° angle cutoff values were used for determining the H-bonds.

Table 4 H-bonds to chromophore from neighboring sidechains are presented in terms of time fraction as calculated from MD simulations in mCardinal2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mCardinal2</td>
<td>D</td>
<td>W</td>
<td>D</td>
<td>W</td>
<td>D</td>
<td>W</td>
<td>D</td>
</tr>
<tr>
<td>77k</td>
<td>0</td>
<td>99.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>90.2</td>
<td>0</td>
</tr>
<tr>
<td>196k</td>
<td>0</td>
<td>81.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>86.9</td>
<td>0</td>
</tr>
<tr>
<td>240k</td>
<td>4.2</td>
<td>27.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>34.7</td>
<td>28.1</td>
</tr>
<tr>
<td>300k</td>
<td>0.2</td>
<td>3.9</td>
<td>0</td>
<td>0</td>
<td>6.4</td>
<td>29.9</td>
<td>24.4</td>
</tr>
<tr>
<td>340k</td>
<td>3.6</td>
<td>2.3</td>
<td>0</td>
<td>0</td>
<td>9.7</td>
<td>25.1</td>
<td>14.4</td>
</tr>
</tbody>
</table>

Also, Q41 interacts with acylimine oxygen by both direct hydrogen bond and water-mediated hydrogen bond as shown in Figures (14c), (14d), (14e), and (14f). Therefore, both direct hydrogen bonds and water-mediated hydrogen bonds are seen between F62: O – Q41 and F62: O – S66. Because of the rotations of the side chain of
residues S66 and Q39, there exist the two networks (i) 62 – 41 - 28 and (ii) 62 – water – 41 - 28 and these are helpful in chromophore flexibility. Preliminary work at 300K shows that this variant has two sets of interconverting populations in the ground state and shows the possibility of enhanced Stokes shift [79]. There is a significant rearrangement between the hydrogen bond pattern with switchable but relatively stable conformational sub-states, a hallmark for the correlation with the enhanced Stokes shift. The increase in temperature enhances the atom's vibration. At a temperature of 340K, the residues move apart and water molecules moving rapidly and lacking the hydrogen bond interconversion. At 400K, the water molecules crowded towards the acylimine region without any interactions, and amino acid residues move apart keeping more distance between them. Some water-mediated bond is rarely seen for a short time i.e. F62 – water - 66 and F62 – water- 106. Therefore, above 300K, both hydrogen bond states are seen, which is not observed below 300K.
Figure 15 Distance time trajectories between the extended chromophore and nearby residues of mCardinal2

The amino acid residue pairs (I) F62O – Q106 (II) F62O – 39NE2, and F62O – 39OE1 spend more time in water-mediated hydrogen bonds than direct hydrogen bonds as shown in figure (15a), (15e), and (15f). The distance-time graph in Figures (15b), (15c), (15g), and (15h) shows the interconversion between the direct hydrogen bond states and water-mediated hydrogen bond states and shows the strong propensity to have extended Stokes shift.

5. QUANTUM MECHANICAL CALCULATION

Gaussian is a computational software package initially released in 1970 by John Pople and his co-workers [80]. It is used by physicists, engineers, chemists, and different scientists engaged in diverse scientific research. The name Gaussian comes from the use of Gaussian type orbitals (GTOs) employed in the electronic structure calculation to predict energies, molecular structures, spectroscopic data, and other advanced electronic structure information. It has been continuously updated and I am using Gaussian 09 in High-Performance Computational (HPC) machine at Florida International University (FIU). It can do various types of quantum mechanical (QM) calculations involving diverse methods. I am using Self-Consistent Field (SCF), Local Density Approximation (LDA) of Density
Functional Theory (DFT), Time-dependent DFT, and Multi-Configurational Self-Consistent Field (MCSCF) or Complete Active Space Self-Consistent Field (CASSCF).

The basis set is the set of basis functions that are used to represent the electronic wave function in a calculation. It is developed by John Pople and is well known by the Gaussian set of programs. In this study, I am using a different basis set such as 6-31G (d, p), 6-311G (d, p), and 6-311++ (2d, 2p) [81]. The 6-31G is the Pople’s split-valence double-zeta basis set, where core orbitals have made up of 6 Gaussian types of orbitals (GTOs). Similarly, the valence shell is described by two orbitals: one has 3 GTOs and the other has one GTO. The basis set 6-31G (d, p) is defined as the basis set 6-31G having d polarization functions for non-hydrogen atoms and p polarization functions for hydrogen atoms. The 6-311G is a split-valence triple-zeta basis. It includes one GTO to 6-31G. The 6-311G (d, p) is 6-311G that has a d polarization function for non-hydrogen atoms and p polarization functions for hydrogen atoms. The 6-311G++ (2d, 2p) refers to the 6-311G basis set illustrated by diffuse function, two sets of d function on heavy atoms, supplemented by two sets of p functions on hydrogen atoms. The notation ++ represents the diffuse function of heavy elements, hydrogen atoms, and heavy atoms. The use of a basis set converts the complex physical function into the simpler algebraic equation that is applicable for efficient implementation in computational techniques. Therefore, the search for suitable basis functions and orbitals enables the theoretical observation comparable with both reference data and experimental results.

The structure was performed optimization by using the DFT method at the ground state. The optimized structure was used for excited-state calculation in my study. The keyword during ground state optimization is B3LYP/basis. The B3LYP (Local DFT
and exchange cor-relation) is the implicit density functional, which stands for "Becke, 3-parameter, Lee-Yang–Parr". B3 is Becke’s 3 parameter exchange-correlation functional which uses 3 parameters to mix in the exact Hartree-Fock exchange-correlation and LYP is the Lee-Yang and Parr correlation functional that interprets the dynamic electron correlation. There are the different basis used during optimization and calculation. It depends upon different factors based upon the type of calculation and computational cost. DFT calculations have been widely used to predict and estimate a great variety of material and molecular properties [82]. The calculations are ab initio since only atomic types and their spatial positions are required. All my DFT calculations were performed using the Gaussian 09 Simulation Package [83]. The experimental structure of a molecule was taken from the database provided and optimized by using the first principle DFT to minimize the atomic forces and stress. The structure is moved to a local energy minimum. In first principle DFT, initial structures may be entirely random, where the atoms can be too close to each other (or far apart) that improves the stability of the electronic energy minimization and encourages the formation of connected structures rather than isolated fragments and can space the different species out appropriately. Optimized geometry from the first principle DFT can provide the strain-free lattice constants, atomic coordinates, and ground-state structure. Comparison with experimental physical parameters gives the main circumstantial support in favor of the LDA approximation and exchange-correlation function used in the DFT calculations. Proper exchange terms and optimization constraints usually yield a geometry that shows excellent agreement with the experimental one. If proper optimization is not done then the electronic structure will not correspond to a relaxed and ground-state structure. Therefore, DFT provides a relatively efficient and
unbiased tool to compute the ground state energy of materials and their surfaces. The reliability of such calculations depends on the development of approximations for the exchange-correlation energy functional. Significant theories have been made in recent years in the quality of exchange-correlation functionals and other principles.

When a molecule is described by more than one dominant electronic configuration then single wave function methods fail and one needs to consider multi-configuration methods. CASSCF method treats the multi-configurational character of molecules by including static (non-dynamic) correlation of wave function [84]. It is compromised that dynamical correlation is usually less efficient than other emerging methods, has to be introduced a careful selection of the method especially for wave functions with intermediate static correlation. The CASSCF methods recover changes that occur in the correlation energy in a certain process. This calculation is a combination of an SCF computation with a full configuration interaction (CI) involving several configurations. The full CI is a linear variational approach (approximate wave function) capable of finding an exact numerical solution with a complete basis set. One such method is CASSCF. The active space is a type of molecular orbital, which is partially occupied. It is understood that the electrons come from the highest occupied orbitals in the initial Hartree-Fock determinant and that the remaining orbitals required for the active space come from the initial unoccupied orbitals. The active space is designated by CASSCF \([n, m]\) in the Gaussian program, where \(n\) is the number of valence electrons (pi bond/homo) that are distributed in \(m\) orbitals in all possible ways. A full CI calculation is performed within the active space selected [85]. However, the CASSCF becomes very expensive even for small active spaces. But these are not black-box methods because the specific number
of electrons are included in a specific orbital rather than taking the total electron in the whole orbital. They require expertise and insight into the problem because sometimes CASSCF wave functions are difficult to converge and more time-consuming. Therefore, we are seeking a smaller basis that can preserve the computational cost with great accuracy, which will be an immense achievement in this calculation.

5.1 DFT AND CASSCF CALCULATION ON BENZENE

In the first principle DFT, atoms may be randomly oriented in initial structure. They do not have an adjustable parameter that improves the stability of electronic energy minimization. After energy minimization, optimized geometry from the first principle DFT can provide the strain-free lattice constants, atomic coordinates, and ground-state structure. One such structure is a small planar Benzene (C₆H₆) organic molecule in my study. It is widely used as a fundamental chemical for industrial purposes. The structure of C₆H₆ contains six carbon atoms that form a hexagonal ring, each with a hydrogen atom attached. All of the carbon-carbon (C-C) bonds have exactly equal lengths with an alternate single and double bond as shown in Figure 16. It was easier for us to grab this smaller structure and used it as a trial sample molecule. Also, the computational cost was cheaper, when we calculated the electronic structure. As a simple structure, the learning process became simple, and different basis sets were tested using different methods such as DFT, TDDFT, and CASSCF in Gaussian 09. My final aim was the calculation of Stokes shift of Far-Red Fluorescent Proteins (RFPs) from excitation and emission data, where I used the chromophore structure. The chromophore structure also contains six ring structures connected to five ring structures and other molecules. Therefore, the electronic structure
calculation was helpful for us to estimate the computational time, reliability of methods, and basis sets.

![Benzene hexagonal ring](image)

*Figure 16 Benzene hexagonal ring*

The structure of benzene is taken with bond length C-C and C-H are 1.39Å and 1.07Å, where experimental values for C-C and C-H are 1.3964Å and 1.0831Å respectively [86]. The structure was optimized by the DFT method, using Gaussian 09. The appropriate choice of basis and size of active space have to be studied to find the balance between accuracy and computational cost. Therefore, the optimized structure was used for further calculation in the form of the trial sample to check the precise method and basis set. This calculation notified me to start the electronic structure calculation for the Far RFPs chromophore.

The singlet ground state (S$_0$), lowest triplet state (T$_1$), and first singlet excited state (S$_1$) of benzene are calculated using CASSCF wave functions constructed from GTOs. The keyword and basis set CASSCF (6, 6)/ 6-311G++(2d, 2p) is used to calculate the energies of various wave functions for the S$_0$, T$_1$, and S$_1$ states of benzene are shown in Table 5. The quantity inside the bracket (2d, 2p) is the polarization function or active space that provides flexibility to other rigid basis sets. The ground state, lowest triplet state, and
singlet excited states of benzene were described by $\pi$-space CASSCF wave functions with 6 electrons in 6 orbitals and represented by CASSCF(6,6). As expected from the work of other authors (Peter B Karadokov), $S_0 \rightarrow T_1$ and $S_0 \rightarrow S_1$ excitation energies of benzene agree very well with experimental data and higher-level of theoretical estimates [87]. The unit of energy is in Hartree, which can be converted into an electron volt (eV) unit. My calculation at an active space (6, 6) is not in good agreement with both reference and experimental data as shown below in Table 5.

**Table 5** Singlet ground ($S_0$), lowest triplet ($T_1$), and first singlet excited ($S_1$) of Benzene

<table>
<thead>
<tr>
<th>Energy (au) of</th>
<th>Singlet Ground ($S_0$)</th>
<th>Lowest Triplet ($T_1$)</th>
<th>Singlet Excited ($S_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(6, 6)$</td>
<td>Reference</td>
<td>$(6, 6)$</td>
<td>Reference</td>
</tr>
</tbody>
</table>

1 Hartree = $2 \times 13.6 \, eV = 27.2 \, eV$

$S_0 \rightarrow T_1 = ST\Delta E (6, 6) = 4.776 \, eV; \quad ST\Delta E_{REF} = 3.86 \, eV; \quad ST\Delta E_{EXP} = 3.95 \, eV$

$S_0 \rightarrow S_1 = SS\Delta E (6, 6) = 6.533 \, eV; \quad SS\Delta E_{REF} = 4.96 \, eV; \quad SS\Delta E_{EXP} = 4.90 \, eV$

**Table 6** Singlet ground ($S_0$), lowest triplet ($T_1$), and first singlet excited ($S_1$) of Benzene with more active space

<table>
<thead>
<tr>
<th>Energy (au) of</th>
<th>Singlet Ground ($S_0$)</th>
<th>Lowest Triplet ($T_1$)</th>
<th>Singlet Excited ($S_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(6, 6)$</td>
<td>$(6, 8)$</td>
<td>$(12, 12)$</td>
<td>$(6, 6)$</td>
</tr>
<tr>
<td>-230.5775</td>
<td>-230.8474</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$S_0 \rightarrow T_1 = ST\Delta E (6, 6) = 4.776 \, eV; S_0 \rightarrow T_1 = ST\Delta E (6, 8) = 4.703 \, eV$

$S_0 \rightarrow T_1 = ST\Delta E (12, 12) = 4.05 \, eV$

$S_0 \rightarrow S_1 = SS\Delta E (6, 6) = 6.533 \, eV; S_0 \rightarrow S_1 = SS\Delta E (6, 8) = 6.495 \, eV$

$S_0 \rightarrow S_1 = SS\Delta E (12, 12) = 1.54 \, eV$
The variations of our results at (6, 6) active space with the reference data depend on different factors. I am working on the Gaussian 09 program package, but the reference was used the Dalton package. DFT method is used to optimize the benzene structure. But reference used other methods. Therefore, the active space was increased and the transition was replaced by the polarization function (6, 8), (6, 9), and (12, 12) to match our final result with the experimental data and reference data. The energy from the singlet ground state to the singlet excited state and singlet ground to the lowest triplet was calculated for increased active space as mentioned in Table 6.

The energy (STΔE and SSΔE) found in reference data and experimental results are not matching with the calculated energy for polarization function (6, 6) as in Table 5. Then I was taking double bond electrons and increased active space size one after another. I was able to see the improvement in calculated values of those energies in polarization function (6, 8) over (6, 6) as in Table 6. I tried to calculate the energy for the polarization function (6, 9), and (6, 10) but ground singlet energy and singlet excited energy were overlapped. Therefore, I am unable to explain those energies here. Similarly, the possible polarization function for energy calculation is (6, 9) and (12, 12), but similar results were identified if it is evaluated from the orbital concept. So, I took the polarization function (12, 12) to calculate those energies as mentioned in Table 6.

I also calculated the above energy transition for polarization functions (6, 6) and (12, 12) by taking a smaller basis set [6-311(d,p)]. It is found that the singlet to triplet energy and singlet to singlet excited energy are better than Table 6 and are in close agreement with the experimental value as mentioned in Table 7.
Table 7. The singlet ground ($S_0$), lowest triplet ($T_1$), and first singlet excited ($S_1$) of benzene

<table>
<thead>
<tr>
<th>Energy (au) of</th>
<th>Singlet Ground ($S_0$)</th>
<th>Lowest Triplet ($T_1$)</th>
<th>Singlet Excited ($S_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(12, 12)</td>
<td>(6, 6)</td>
<td>Reference</td>
</tr>
<tr>
<td>$S_0 \rightarrow T_1 = ST\Delta E (6, 6) = 4.83$ eV;</td>
<td>$S_0 \rightarrow T_1 = ST\Delta E (12, 12) = 4.04$ eV;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$ST\Delta E_{REF} = 3.86$ eV; $ST\Delta E_{EXP} = 3.95$ eV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_0 \rightarrow S_1 = SS\Delta E (6, 6) = 7.3$ eV;</td>
<td>$S_0 \rightarrow S_1 = SS\Delta E (12, 12) = 5.2$ eV;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SS\Delta E_{REF} = 4.96$ eV; $SS\Delta E_{EXP} = 4.90$ eV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I calculated singlet and triplet energy for both polarization functions (6, 6) and (12, 12).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The excitation and emission energy for both cases were calculated as shown in table 8. Let $S_0$, $T_{1S}$, $T_1$, and $S_{OT}$ are singlet ground state, triplet from singlet ground state, triplet state, and singlet from triplet state respectively</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singlet Ground = $S_0$; Triplet from $S_0 = T_{1S}$; Triplet = $T_1$; Singlet from $T_1 = S_{OT}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8. Excitation and emission of benzene from singlet energy and triplet energy

<table>
<thead>
<tr>
<th>Excitation (nm)</th>
<th>Emission (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(6, 6)</td>
<td>(12, 12)</td>
</tr>
<tr>
<td>Singlet Ground ($S_0$)</td>
<td>Triplet from $S_0$ ($T_{1S}$)</td>
</tr>
</tbody>
</table>

I have taken the same geometry of benzene as in the initial calculation in Table 5. This structure was optimized in the ground state by using the DFT method in lower basis and lower orbital [6-311G (d, p)]. The optimized structure was used to calculate singlet and triplet energy following the ground state as in Table 8. These energies were used to evaluate the excitation energy as mentioned below. Again geometry from ground state triplet was optimized in the lowest triplet state for further calculation. The lowest triplet
state optimization was done by DFT with the same basis and orbital as in ground-state calculation. The excited state optimize geometry was used to calculate triplet energy and next ground state energy at excited state as mentioned in Table 8. These energies were used to calculate the emission energy as shown below. Let $E_{\text{ex}}$ and $E_{\text{em}}$ be the excitation energy and emission energy of benzene at active space (12, 12). Then,

$E_{\text{ex}} = 4.02 \text{ eV} = 308.42 \text{ nm}$; $E_{\text{em}} = 3.60 \text{ eV} = 344.40 \text{ nm}$; Stokes Shift = 35.98 nm

5.2 TDDFT CALCULATION OF BENZENE

My calculation for active space (12,12) has achieved a good result in comparison with the experimental data, but it requires more than 2 hrs to complete the excitation only for a small molecule like Benzene in the FIU HPC machine. I expect a similar calculation in FPs chromophore that will be difficult to perform the complete calculation including optimization, excitation, and emission now. The main drawback of this calculation is due to the bigger size of the chromophore structure that has more than 50 atoms. It increases the size of active space that requires a higher computational cost to complete the calculation. HPC machine in FIU couldn’t handle bigger active space like (6, 12) to perform electronic structure calculation using CASSCF method. Therefore, it was found necessary to search for a reliable method to execute electronic structure calculation for a Benzene that requires lower computational cost with higher accuracy in results compared to experimental data.

The final step of testing the basis and method was using a 6-31 basis under the TDDFT method. The benzene structure was used to set up excited-state calculation using ground state geometry to get excitation values. Then, the structure was optimized at the excited state and performed excited state energy calculation. The TDDFT method can
calculate both excitation and emission energy/wavelength as mentioned below. By evaluating the Stokes shift of Benzene, I was confirmed that the TDDFT method is enough to proceed with the electronic structure calculation of the FP chromophore.

\[
E_{ex} = 5.510 \text{ eV} = 225.01 \text{ nm}; \quad E_{ems} = 5.204 \text{ eV} = 238.23 \text{ nm}; \quad \text{Stokes shift} = 13 \text{ nm}
\]

The above process is the test calculations to check whether the basis and keyword/method taken for benzene are suitable for the chromophore electronic structure calculations or not.

5.3 QUANTUM MECHANICAL CALCULATION ON CHROMOPHORE

I have tested the benzene as a test sample to search for a suitable basis set and methods to calculate the excited state energy for the chromophore of my interest. During the process, I used different basis sets such as 6-31, 6-311, 631, and 6-311++. The calculation was done using DFT, CASSCF, and TDDFT methods. I was looking for a basis set and methods that could establish precise results comparable to experimental observation and require a lower computational cost. My testing finds that DFT is adequate to optimize the structure at the ground state and TDDFT is fast enough to calculate the excitation energy with great accuracy and lower the computational time.

5.3.1 HYDROGEN CAPPING AND THEIR ELECTRONIC STRUCTURE

The FP contains a large number of atoms and it was difficult to perform electronic structure calculations by taking such a huge structure. From MD trajectory results, it is confirmed that the chromophore is responsible to have extended Stokes shift in FPs. So, I have to isolate a chromophore that contains less than 70 atoms. It is possible to use the TDDFT method to calculate excitation energy and emission energy which is used to evaluate Stokes shift.
I was further looking for an anionic chromophore structure and suitable methods with a reliable basis set. The results from the TDDFT method with a 6-31(d, p) basis are assumed to be successful to get a good matching with experimental data. Therefore, the chromophore was isolated from the protein barrel. Then, multiple structures of chromophore were built by taking different areas and adding two hydrogen caps at C-terminus and N-terminus on each structure. This process creates a different chromophore structure of mNeptune variants as shown in Figure 17. The Hydrogen_Cap, Carbon_Cap, Amine_Cap, and chromophore_no ring are four different structures, each structure has a different number of atoms. Each structure was optimized by using the DFT method at ground state using the same keyword as in Benzene and a 6-31G(d, p) basis. i.e. keyword used in the input is #n b3lyp/6-31G(d, p) opt. The charge and spin used were -1 and 1 during ground state optimization. The optimized geometry was used to perform TDDFT calculation to calculate the excitation energy/wavelength. The keyword used in the input is #n b3lyp/6-31G(d, p) td. The structure was again optimized at an excited state using the TDDFT method with keyword #n b3lyp/6-31G(d, p) td opt. This excited state optimized geometry was used to calculate emission energy/wavelength by using the same keyword.

Table 9. Excitation and emission wavelength of the chromophore

<table>
<thead>
<tr>
<th>Name of Chromophore</th>
<th>TDDFT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E_{ex}$ (nm)</td>
</tr>
<tr>
<td>Hydrogen_Cap</td>
<td>475.01</td>
</tr>
<tr>
<td>Carbon_Cap</td>
<td>523.44</td>
</tr>
<tr>
<td>Amine_Cap</td>
<td>536.17</td>
</tr>
<tr>
<td>Chromophore_no ring</td>
<td>523.95</td>
</tr>
</tbody>
</table>
as it was used in excitation. The calculated excitation and emission values are not matching with the experimental data but lie within the trend. The Stokes shift obtained during that process is too far from the experimental data as shown in Table 9. The Stokes shift for chromophore structures Hydrogen_Cap and Chromophore_no ring are near to experimental observations.

![Figure 17 Chromophore structures based on the different terminus](image)

Again, another set of data was set up for chromophore (NRQ-63) of far RFPs protein (mNeptune variants). For this process, the trajectory data were taken from MD simulation in the interval of every 2.5ns (approximately). Every 10ns contains 5000 frames.
During data extraction, a single frame of protein from the end of every 2.5ns was saved in the form of a dcd file by using save_everyx.TCL script. Then, the trajectory was loaded in VMD to see the whole structure. The chromophore data was selected (not water and resid 62 63) using VMD and the position of atoms was saved in the form of the PDB file in the interval 2.5 ns up to 100ns. These PDB files were loaded in Avogadro software one by one to edit the atoms and add the necessary atoms (hydrogen cap) on the C-terminus and N-terminus. This chromophore structure was named “Carbon_Cap”. Then input file for Gaussian was made by clicking the option “extension” and the input file was saved (dot com). This file was used to calculate the absorption spectra without optimization using time-dependent DFT in the beginning. The keyword used to obtain the wavelength of excitation is \#n b3lyp/6-31G(d, p) td (nstates=3). The same procedures were done for the remaining structures to obtain the excitation wavelength at different time frames. Again, the same structure was taken and DFT optimized the newly added hydrogen (Hydrogen cap) only, where other atoms remained freeze. The optimized hydrogen cap chromophore structure was uploaded in Avogadro software and repeated the same procedure as it was done for the same structure which was completely unoptimized. A large number of data sets were calculated in both structures (unoptimized hydrogen cap and optimized hydrogen cap). The graph was plotted for both sets of data in the form of a histogram as shown in Figure 18. The width of the bin taken during plotting 38. Histogram (a) was plotted using data which is obtained by dividing original data by 1.32. The same procedure was followed to plot histogram (b). The average peak value of absorption is around 600nm, which is in agreement with the experimental data (599nm). In Figure 18 (b) the width of each bin is 38 and the average peak value of excitation is around 600nm.
Figure 18 Histogram Plot of Excited-State of Chromophore (a) Unoptimized Hydrogen Cap (b) Optimized Hydrogen Cap

It means the data obtained with optimized H-cap and unoptimized H-cap has a negligible difference that is evaluated from histogram plot Figure 18.
5.3.2 RESULT ON mCARDINAL2

My work is based on the MD simulation of mNeptune variants, where hydrogen bonds between the extended chromophore and nearby amino-acid residues were observed. The hydrogen bonding to the N-acylimine carbonyl of extended chromophore has been observed in variants mNeptune1, mNeptune2.5, and mCardinal2. I choose mCardinal2 to execute the electronic structure calculation. The N-acylimine was observed to form bonding patterns (both direct hydrogen bond and water-mediated hydrogen bond) with S66 residue i.e F62: O --- S66: N (OG) (see Table 4). I selected the hydrogen bond conformation between the extended chromophore and S66 from simulation 20ns-100ns. The conformation structures are direct hydrogen bond and water-mediated hydrogen bond between residues F62 and S66. Each structure in MD simulations has appeared in a different time/frame. Around 75 structures were taken for each bonding geometry under multiple time frames at a temperature of 300K. Because of the rigidity of the chromophore environment, I could select only 50 frames at 240K. The structures were selected using the VMD and converted the data in the form of PDB. Then, the bonding conformations were isolated and terminated by adding two hydrogen atoms at proper positions which are also called terminus (C and N). The Gaussian input file was generated with the help of Avogadro’s tool. There were two steps to calculate the excited-state energy. In the first step of the calculation, the structure was optimized in the ground state using DFT, by freezing all atoms except two capping hydrogen atoms in C and N terminus. In the next step, all atoms were released and subjected to time-dependent DFT calculation. As soon as the calculation was completed, energy values and wavelength of structures (both direct hydrogen bond and water-mediated hydrogen bond) were collected to their respective
bonding states. These data were collected separately to make precise common intervals for both sets of data that can help to identify the frequencies of each interval. Then average values of each interval were determined. There were three sets of data (average wavelength from each interval, sample size for both bonding structures) for 300K and 240K. The data were selected and plotted in the double histogram (bin width 55nm) as shown in Figure 19. It enables me to compare two kinds of hydrogen bond patterns observed in F62: O --- S66: N (OG) at temperatures of 240K and 300K. The position of the peak found in a histogram is at a longer wavelength for water-mediated hydrogen bond than direct hydrogen bond at 300K as shown in Figure 19 (a). The appearance of two peaks at two different positions, confirms the higher propensity to get extended Stokes shift in mCardinal2 at 300K. The availability of flexibility of chromophore has an immense responsibility to interchange the hydrogen bond states, resulting in extended Stokes shift at that environment. At low temperatures of 240K, the peak position for both direct hydrogen bond and water-mediated hydrogen bond lies in the same wavelength in the histogram plot as shown in Figure 19 (b). This result agrees with the experimental observation of reduced Stokes shift at 240K in mCardinal2.
Figure 19 Double Histogram Plot of Direct Hydrogen Bond and Water-mediated Hydrogen Bond at temperatures (a) 300K and (b) 240K
6. CONCLUSION

I have demonstrated a novel and powerful approach to study the Stokes shift of fluorescent proteins. This approach combines two of the most powerful computational techniques widely used in the field of biophysical research. One is the molecular dynamics simulation based on Newtonian mechanics, the other is electronic structure calculation based on quantum theory. Our MD simulation revealed a strong correlation between the enhanced Stokes shift and the hydrogen-bonding network surrounding the chromophore. I then verified this correlation through QM calculations of the electronic excitations of the far-red chromophore. My research approach can be applied to a variety of fluorescent proteins and their variants, at various temperatures. One can use this approach to study theoretically/artificially mutated fluorescent proteins and to predict their hydrogen bond dynamics and their Stokes shift.

I applied this approach successfully to this study of hydrogen bonding dynamics and the effect on Stokes shift for three variants of the far-red chromophore, at several temperatures. My MD simulation identified the most active hydrogen bonding mechanism that correlated strongly with the Stokes shift. My study revealed that the fluctuation between direct-hydrogen bonding and water-mediated hydrogen bonding of the chromophore’s conformation is most prominent at room temperature, and at lower temperature fluctuation is reduced and chromophore becomes more rigid. This agrees well with the experimental findings. There were experimental researches on other mNeptune variants such as mNeptune681 and mNeptune684, and their hydrogen bonding analysis around the chromophore was found to be responsible for the largest red Stokes shift (80nm). Their emission wavelength lies within the Near Infra-Red (NIR) region which is
used for tissue imaging, tumor diagnosis, and possible treatment of the complex type of cancer. My QM calculation focused on the extended chromophore molecule and used the molecular structure information from MD simulation, therefore combining the most powerful features of these two computational techniques. This allowed me to compute Stokes shift at around 50nm range, again compared favorably with the experimental findings.

The mNeptune variants were undergone MD simulation of 100ns at different temperatures. The results were analyzed by importing their trajectories into the VMD tool. The excitation of amino–acid residues including chromophore increases with an increase in temperature. The trajectory visualization of MD simulation exhibits the occurrences of hydrogen bonding between the extended chromophore and nearby amino acid residues at a temperature of 300K enhanced chromophore flexibility and different hydrogen bond states were observed in my MD simulation. The most prominent hydrogen bond states (F62—S66) were found on mCardinal2 that includes both direct hydrogen bond and watermediate hydrogen bond.

I have performed extensive first-principles electronic structure calculations using FIU’s high-performance cluster. My results on the ground-state structure and excitation energies of Benzene agree well with the experimental values as well as other theoretical studies. The most relevant singlet excitation energy was found to be 5.2eV, only 6% higher than the experimental result. However, a similar calculation on a much larger chromophore molecule would be prohibited expensive. Fortunately, my calculation on Benzene using the TDDFT method demonstrated that the TDDFT method is very efficient to perform large number of computations with acceptable accuracy. The excitation and
emission wavelengths were calculated for all conformations. Those wavelength values were plotted in a histogram. The Stokes shift of mCardinal2 can be found from the peak position difference between direct hydrogen bond states and water-mediated hydrogen bond states. The findings from the histogram plot revealed enhanced Stokes shift in mCardinal2 at 300K, but it is found to be negligible at a low temperature of 240K.

These approximations may not be exact and a better approximation method in the QM technique is desired. There are several ways to get a more reliable excited state energy. The simulation time of 100ns may not be sufficient to analyze the hydrogen bond state. Before entering into the simulation, the system equilibration should be enough to allow the atoms and residues to move in the proper directions for both molecular and electronic interactions. The number of frames used in the TDDFT calculations may not be sufficient. The increase in frame number from MD simulation increases the chances of reducing error in excited-state calculation. It will help to reduce the uncertainties in the final excitation energy distributions. To capture different hydrogen bond configurations under multiple type frames, a strong and automated method should be employed that can improve the TDDFT results. There are other types of possible interaction (non-radiative energy dissipation) within the chromophore environment that has to be studied carefully. Further research on the far RFPs could be done by improving the above drawback to obtain a better result using QM calculations.
REFERENCES


APPENDICES

Appendix A: Chromophore Structure of Far RFP (PDB file is used to show the name of chromophore atoms)

Appendix B: Residue topology and parameter files for chromophore of far RFP

Table 5.1 Residues topology file for anionic chromophore of far RFP

| MASS  | 197 NRC2 | 14.00700 N ! neutral his unprotonated ring nitrogen |
| MASS  | 198 NRC1 | 14.00700 N ! neutral his protonated ring nitrogen |
| MASS  | 199 HAC1 | 1.00800 H ! for alkene; RHC=CR |
| MASS  | 200 HPC | 1.00800 H ! aromatic H |
| MASS  | 201 OC2 | 15.99900 O ! carbonyl oxygen |
| MASS  | 202 OCH | 15.99900 O ! from OH1 |
| !MASS | 203 HCH | 1.00800 H ! polar H |
| MASS  | 204 HAC | 1.00800 H ! nonpolar H |
| MASS  | 205 CA1 | 12.01100 C ! aromatic C |
| MASS  | 206 CA2 | 12.01100 C ! aromatic C |
| MASS  | 207 CA3 | 12.01100 C ! aromatic C |
| MASS  | 208 CPC2 | 12.01100 C ! his C61 carbon |
| MASS  | 209 CPC1 | 12.01100 C ! for alkene; RHC=CR |
| MASS  | 210 CPC1 | 12.01100 C ! his CG and CD2 carbons |
| MASS  | 211 CA4 | 12.01100 C ! aromatic C |
| !MASS | 212 CT3C | 12.01100 C ! aliphatic sp3 C for CH3 |
| MASS  | 213 CTC1 | 12.01100 C ! aliphatic sp3 C for CH |
| MASS  | 492 OM | 15.99900 O ! heme C0/02 oxygen |

DECL -CA
DECL -C
DECL -O
DECL -C3 !Chola
DECL +N1 !Chola
DECL +N
DECL +HN
DECL +CA

DEFAR FIRS NTER LAST CTER
AUTO ANGLES DIHE
RESI NRQ -1.000
GROUP ! Imidazolinone ring
ATOM C1 CPC2 0.50
ATOM N2 NRC2 -0.60
ATOM N3 NRC1 -0.57
ATOM C2 CPC1 0.57
ATOM O2 OC2 -0.57
ATOM CA2 CPC1 0.10
ATOM CB2 CEC1 -0.14
ATOM HB2 HAC1 0.21
ATOM CG2 CA1 -0.09 ! Tyr ring : charges from charmm22
ATOM CD1 CA2 -0.08
ATOM HD1 HPC 0.14
ATOM CD2 CA2 -0.08
ATOM HD2 HPC 0.14
ATOM CE1 CA3 -0.28
ATOM HE1 HPC 0.10
ATOM CE2 CA3 -0.28
ATOM HE2 HPC 0.10
ATOM C2 CA4 0.45
ATOM OH OCH -0.62

! Glycine part from Charmm22
GROUP
ATOM CA3 CT2 -0.18 ! |
ATOM HA31 HB2 0.09 ! |
ATOM HA32 HB2 0.09 ! HA1-CA-HA2
GROUP ! |
ATOM C3 C 0.51 ! |
ATOM O3 O -0.51 ! C=O

! Met part from Charmm22
GROUP
ATOM N1 NH1 -0.16
ATOM CA1 CTCl 0.16 ! atom type changed
GROUP
ATOM CB1 CT2 -0.18
ATOM HB11 HA2 0.09
ATOM HB12 HA2 0.09
GROUP
ATOM CG1 CT2 -0.14
ATOM HG11 HA2 0.09
ATOM HG12 HA2 0.09
ATOM SD S -0.09
ATOM CE CT3 -0.22
ATOM HE11 HA3 0.09
ATOM HE12 HA3 0.09
ATOM HE13 HA3 0.09
BOND CA1 C1 N1 -C C3 +N
BOND N2 CA2 CB2 HB2 CB2 CG2 CD1 HD1 CD1 CE1 CE1 HE1 CZ OH
BOND CZ CE2 CE2 HE2 CD2 HD2 CD2 CG2 CA2 C2
BOND N3 CA3 CA3 HA31 CA3 HA32 CA3 C3 N3 C1 N3 C2
BOND CB1 HB11 CB1 HB12 CB1 CG1 CG1 HG11 CG1 HG12 CA1 CB1
BOND CG1 SD SD CE CE HE11 CE HE12 CE HE13
DOUBLE C1 N2 CA2 CB2 C2 O2 C3 O3 CD1 CG2 CD2 CE2 CZ CE1 CA1 N1

*molecular dioxygen parameter from heme group by chola

RESI O2 0.00 ! O2 ligand for heme
GROUP
ATOM O1 ON 0.021
ATOM O2 OM -0.021
BOND O1 O2

Table 5.2 Parameter file for anionic chromophore of far RFP

*charm parameter file of RFP chromophore (met-tyr-gly)
*
!parameter file

! CHROMOPHORE parameters, protonated form
!
BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type Kb b0

CPC2 CTCl 354.000 1.4900 !ion for RFP Cl-CAl connection
NRC1 CT2 396.000 1.4400 !ion
NRC1 CPC2 400.000 1.3900 !
NRC1 CPC1 400.000 1.4100 !
CPC1 GC2 654.000 1.2400 !ion
NRC2 CPC2 400.000 1.3000 !
CPC1 CPC1 410.000 1.4600 !ion
NRC2 CPC1 400.000 1.4000 !
CPC1 CEC1 500.000 1.3900 !ion
HAC1 CEC1 360.500 1.1000 !
CECl CA1 437.000 1.4100 !ion
CA1 CA2 305.000 1.4800 !ion
HPC CA2 340.000 1.0800 !
CA2 CA3 305.000 1.3500 !ion
HPC CA3 340.000 1.0800 !
CA3 CA4 305.000 1.4500 !ion
OCH CA4 842.000 1.2500 !ion
CTCl NH1 463.000 1.3650 !RFP CA1-N1 connection
CT2 CTCl 222.500 1.5380 !RFP CB1-CA1 connection
**ANGLES**

\[ V(\text{angle}) = K_{\theta}(\theta - \theta_0)^2 \]

\[ V(\text{Urey-Bradley}) = K_{\text{ub}}(S - S_0)^2 \]

\[ K_{\theta}: \text{kcal/mole/rad}^2 \]

\[ \theta_0: \text{degrees} \]

\[ K_{\text{ub}}: \text{kcal/mole/A}^2 \text{ (Urey-Bradley)} \]

\[ S_0: \text{A} \]

<table>
<thead>
<tr>
<th>atom types</th>
<th>K(_{\theta})</th>
<th>(\theta_0)</th>
<th>(K_{\text{ub}})</th>
<th>(S_0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRC2 CPC2 NRC1</td>
<td>130.0000</td>
<td>113.3000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC2 NRC2 CPC1</td>
<td>130.0000</td>
<td>106.6000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC2 NRC1 CPC1</td>
<td>130.0000</td>
<td>107.9000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRC2 CPC1 CPC1</td>
<td>130.0000</td>
<td>108.3000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRC2 CPC1 CEC1</td>
<td>45.8000</td>
<td>129.5000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRC1 CPC1 OC2</td>
<td>50.0000</td>
<td>124.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>NRC1 CPC1 CPC1</td>
<td>130.0000</td>
<td>103.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>OC2 CPC1 CPC1</td>
<td>44.0000</td>
<td>133.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC1 CPC1 CEC1</td>
<td>45.8000</td>
<td>122.7000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC1 CEC1 CA1</td>
<td>130.0000</td>
<td>133.2000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC1 CEC1 HAC1</td>
<td>42.0000</td>
<td>112.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CEC1 CA1 CA2</td>
<td>45.8000</td>
<td>120.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>HAC1 CEC1 CA1</td>
<td>42.0000</td>
<td>115.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CA1 CA2 CA3</td>
<td>40.0000</td>
<td>122.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CA2 CA1 CA2</td>
<td>40.0000</td>
<td>116.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CA2 CA3 CA4</td>
<td>40.0000</td>
<td>122.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CA3 CA4 CA3</td>
<td>40.0000</td>
<td>115.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>HFC CA3 CA4</td>
<td>30.0000</td>
<td>120.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>HFC CA3 CA2</td>
<td>30.0000</td>
<td>120.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>HFC CA2 CA3</td>
<td>30.0000</td>
<td>120.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>HFC CA2 CA1</td>
<td>30.0000</td>
<td>120.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>OCH CA4 CA3</td>
<td>45.2000</td>
<td>120.0000</td>
<td>ALLOW</td>
<td>ARO ALC</td>
</tr>
</tbody>
</table>

!Link to the met(65) fragment

| NH1 CTC1 CPC2 | 50.0000 | 107.0000 | ALLOW | PEP POL ARO ALI |
| NRC2 CPC2 CTC1 | 40.0000 | 125.0000 | ! | |
| NRC1 CPC2 CTC1 | 40.0000 | 121.7000 | !ion | |
| CT2 CT1 CPC2 | 52.0000 | 108.0000 | ALLOW | ALI PEP POL ARO |
| CT2 CTC1 CPC2 | 52.0000 | 108.0000 | ALLOW | ALI PEP POL ARO |
CTC1 NH1 C 50.000 120.0000
NH1 C CTC1 80.000 116.5000
NH1 CTC1 CT2 70.000 113.5000
HA2 CT2 CTC1 33.430 110.1000 22.53 2.17900 !for C36--> HA-->HA2 chol
CT2 CT2 CTC1 58.350 113.50 11.16 2.56100

!Link to the gly(67) fragment
CPC2 NRC1 CT2 40.000 123.3000 !ion
CPC1 NRC1 CT2 40.000 123.8000 !ion
NRC1 CT2 C 50.000 107.0000
NRC1 CT2 HB2 48.000 108.0000 !for C36--> HB-->HB2 chol

! DIHEDRALS
!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!
!atom types    Kchi    n    delta
!
CPC2 NRC2 CPC1 CPC1 14.0000 2 180.000 !
CPC2 NRC1 CPC1 CPC1 14.0000 2 180.000 !
NRC2 CPC2 NRC1 CPC1 14.0000 2 180.000 !
NRC2 CPC1 CPC1 NRC1 4.0000 2 180.000 !
NRC1 CPC2 NRC2 CPC1 4.0000 2 180.000 !
CA1 CA2 CA3 CA4 3.1000 2 180.000 !
CA2 CA1 CA2 CA3 3.1000 2 180.000 !
CA2 CA3 CA4 CA3 3.1000 2 180.000 !
OC2 CA CAC CAC 3.1000 2 180.000 !
CA2 CA3 CA4 OCH 3.1000 2 180.000 !
CA1 CA2 CA3 HFC 4.2000 2 180.000 !
CA2 CA1 CA2 HFC 4.2000 2 180.000 !
CA3 CA4 CA3 HFC 4.2000 2 180.000 !
HFC CA2 CA3 CA4 4.2000 2 180.000 !
HFC CA2 CA3 HFC 2.4000 2 180.000 !
OC2 HCH OC2 CAC CAC 0.9900 2 180.000 !
OC2 HCH OC2 CA4 CA3 0.9900 2 180.000 !
HFC CA3 CA4 OCH 4.2000 2 180.000 !
!
CPC2 NRC2 CPC1 CEC1 3.000 2 180.000 !
NRC2 CPC1 CPC1 CEC1 3.00 2 180.000 !
OC2 CPC1 CPC1 CEC1 2.00 2 180.000 !
CEC1 CA1 CA2 HFC 4.20 2 180.000 !
CEC1 CA1 CA2 CA3 3.10 2 180.000 !
! connection CA-CB

CPC1 CPC1 CEC1 HAC1  3.9000  2  180.00 !
CPC1 CPC1 CEC1 CA1  3.9000  2  180.00 !
NRC2 CPC1 CEC1 HAC1  3.9000  2  180.00 !
NRC2 CPC1 CEC1 CA1  3.9000  2  180.00 !original value 180 Chola

! connection CB-CG2

CPC1 CEC1 CA1 CA2  2.7000  2  180.00 !
HAC1 CEC1 CA1 CA2  2.7000  2  180.00 !
!
CPC2 NRC1 CPC1 OC2 14.0000  2  180.00 !
NRC2 CPC2 NRC1 CT2 14.0000  2  180.00 !
NRC2 CPC1 CEC1 OC2 14.0000  2  180.00 !
CPC1 NRC1 CPC2 CTC1 14.0000  2  180.00 !
OC2 CPC1 NRC1 CT2 14.0000  2  180.00 !
CPC1 NRC2 CPC2 CTC1 14.0000  2  180.00 !
CPC1 NRC2 CPC1 CT2 14.0000  2  180.00 !
CTC1 CPC2 NRC1 CT2 14.0000  2  180.00 !

! Linking the chromophore and the glycine(67) fragment

C  C  CT2  NRC1  0.0000  1  0.00 !
NH1  C  CT2  NRC1  0.6000  1  0.00 !
CPC2 NRC1 CT2 HB2  0.0320  3  0.00 !C36--> HB---->HB2 chola
CPC2 NRC2 CT2 C  0.0320  3  0.00 !
CPC1 NRC1 CT2 HB2  0.0320  3  180.00 !C36--> HB---->HB2 chola
CPC1 NRC1 CT2 C  0.0320  3  180.00 !

! Linking the chromophore and the met(65) fragment

C  NH1  CTC1 CPC2  2.2500  2  180.00 !Taken from X-C-NC2-X Charmm22
NRC2 CPC2 CTC1 CT2  0.1050  3  180.00 !
NRC2 CPC2 CTC1 NH1  0.1050  3  180.00 !
NRC1 CPC2 CTC1 CT2  0.1050  3  0.00 !
NRC1 CPC2 CTC1 NH1  0.1050  3  0.00 !

! connecting N1-Ca1 region due to new type CTC1

O  C  NH1  CTC1  2.5000  2  180.00
CT1  C  NH1  CTC1  2.5000  2  180.00
CT2  CTC1  NH1  C  1.8000  1  0.00
CPC2 CTC1 CT2 HA2  0.2000  3  0.00!C36--> HA---->HA2 chola
CPC2 CTC1 CT2 CT2  0.2000  3  0.00
NH1  CTC1  CT2 HA2  0.2000  3  0.00!C36--> HA---->HA2 chola
NH1  CTC1  CT2 CT2  0.2000  3  0.00

IMPROPER

! \( V(\text{improper}) = K_{\text{psi}}(\text{psi} - \text{psi0})^2 \)

! \( K_{\text{psi}}: \text{kcal/mole/\text{rad}^2} \)
! \( \text{psi0}: \text{degrees} \)
! note that the second column of numbers (0) is ignored

! atom types  \( K_{\text{psi}} \)  \( \text{psi0} \)

93
CPC2 NRC2 NRC1 CTC1  50.0000  0  0.0000  
CPC2 NRC1 NRC2 CTC1  50.0000  0  0.0000  
! 
CPC1 NRC1 CPC1 OC2  50.0000  0  0.0000  
CPC1 CPC1 NRC1 OC2  50.0000  0  0.0000  
! 
NRC1 CPC1 CPC2 CT2  50.0000  0  0.0000  
NRC1 CPC2 CPC1 CT2  50.0000  0  0.0000  
! 
CPC1 NRC2 CPC1 CEC1  50.0000  0  0.0000  
CPC1 CPC1 NRC2 CEC1  50.0000  0  0.0000  
! 
CEC1 CPC1 CA1 HAC1  30.0000  0  0.0000  
CEC1 CA1 CPC1 HAC1  30.0000  0  0.0000  
! 
V(Lennard-Jones) = Ep2,i,j*(Rmin,i,j/r,r,i,j)**12 - 2*(Rmin,i,j/r,r,i,j)**6
! 
epsilon: kcal/mol, Ep2,i,j = sqrt(ep2,i*j)
! Rmin/r = A, Rmin,i,j = Rmin/2,i + Rmin/2,j
! 
@atom ignored epsilon Rmin/2 ignored eps,i,j-1 Rmin/2,i-1
! 
@CAG  5.000000 -0.070000  1.992400  ! ALLOW ARO

NOREPULSE nhxpmd 5 atom ddg shift VALUEN spncharge valuelogo -
CMB  14.0 0.0 0.0 12.0 0.0 2.0 1.0 1.0 1.0 1.0 1.0

CA1  5.000000 -0.070000  1.992400  ! ALLOW ARO
CA2  5.000000 -0.070000  1.992400  ! ALLOW ARO
CA3  5.000000 -0.070000  1.992400  ! ALLOW ARO
CA4  5.000000 -0.070000  1.992400  ! ALLOW ARO
CEC1  0.000000  0.080000  2.089000  ! FOR propane, yin/atom, 12/95
CPC1  0.000000  0.050000  1.300000  ! ALLOW ARO
CPC2  0.000000  0.050000  1.300000  ! ALLOW ARO

!CT3C  0.000000  0.080000  2.060000  0.000000  -0.010000  1.900000  ! ALLOW ALI

MCH  -2.000000  0.046000  0.224500  ! ALLOW PEP POL SUL ARO ALC

HAC  0.000000  0.020000  1.320000  ! ALLOW PEP ALI POL SUL

HAC1  0.000000  0.031000  1.250000  !

HPC  0.000000  0.030000  1.358200  0.000000  -0.030000  1.358200  ! ALLOW ARO

NRC1  0.000000  -0.200000  1.850000  ! ALLOW ARO
NRC2  0.000000  -0.200000  1.850000  ! ALLOW ARO

IQC2  0.000000  -0.120000  1.700000  0.000000  -0.120000  1.400000  ! ALLOW PEP POL
QC2  0.000000  -0.120000  1.700000  ! ALLOW PEP POL,
OC2  0.000000  -0.152100  1.770000  ! ALLOW ALC ARO

CTC1  0.000000  -0.020000  2.275000  0.000000  -0.010000  1.900000

94
Appendix C: Input file of direct hydrogen bond F62:O—S66:N(OG) with hydrogen cap open during DFT optimization at ground state

```
%NProcShared=1
%mem=300MB
#n B3LYP/6-31G(d,p) Opt

Title
-1 1

N -1 -27.38000 20.72300 56.99600
H -1 -27.21300 20.52300 54.03400
C -1 -26.25100 21.07300 55.81300
H -1 -26.20900 20.46300 56.70300
C -1 -24.94900 20.84700 56.50700
H -1 -24.03000 21.25700 55.35000
N -1 -23.17500 21.35100 53.94300
C -1 -24.30100 19.32600 54.71900
C -1 -24.65500 16.66100 53.48700
H -1 -24.81400 19.19000 52.33300
C -1 -24.24100 17.31900 53.51100
H -1 -24.07400 16.74300 52.61200
C -1 -24.06700 16.59000 54.70300
H -1 -23.68500 15.58800 56.67700
C -1 -24.43900 18.62900 55.90600
H -1 -24.11100 19.15500 56.01300
C -1 -24.00700 17.33000 55.09200
H -1 -23.69500 16.80700 56.78400
C -1 -26.25800 22.59300 56.20800
O -1 -26.36000 22.95700 57.41100
C -1 -28.15000 25.24200 54.95900
N -1 -28.40300 25.84600 53.77600
N -1 -29.42000 25.02600 55.64000
C -1 -30.49600 25.61800 54.04700
O -1 -31.65500 25.57400 55.10800
C -1 -29.77100 26.11400 53.62400
C -2 -30.49100 26.71400 52.55300
H -1 -31.53700 26.53900 52.04600
C -1 -30.25700 27.43600 51.35900
C -1 -31.45200 27.74300 50.51300
H -1 -32.41600 27.40300 50.76300
C -1 -28.96200 27.81400 50.80400
H -1 -27.96500 27.78800 51.21900
C -1 -31.13000 28.55400 49.48900
H -1 -32.35100 28.74100 49.95600
C -1 -28.97800 28.47700 49.59500
H -1 -28.01200 28.81500 49.25100
C -1 -30.15200 28.04700 48.00500
O -1 -29.56500 26.51000 47.76700
C -1 -29.61100 24.29800 56.87900
H -1 -30.66000 23.89800 56.79500
H -1 -28.90400 23.53000 56.94300
C -1 -29.58600 25.15000 55.09100
O -1 -30.52200 25.49400 58.73200
N -1 -26.75400 23.36200 55.21300
C -1 -26.76300 24.74400 55.39500
C -1 -25.72600 25.33200 54.47100
H -1 -25.93600 26.41700 54.29400
H -1 -25.92700 24.04800 53.49200
```
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-1</td>
<td>-24.27600</td>
<td>25.18200</td>
</tr>
<tr>
<td>H</td>
<td>-1</td>
<td>-24.00300</td>
<td>25.80100</td>
</tr>
<tr>
<td>H</td>
<td>-1</td>
<td>-24.11400</td>
<td>24.11300</td>
</tr>
<tr>
<td>S</td>
<td>-1</td>
<td>-23.10000</td>
<td>25.73600</td>
</tr>
<tr>
<td>C</td>
<td>-1</td>
<td>-21.66200</td>
<td>25.82000</td>
</tr>
<tr>
<td>H</td>
<td>-1</td>
<td>-20.76300</td>
<td>26.26900</td>
</tr>
<tr>
<td>H</td>
<td>-1</td>
<td>-21.91200</td>
<td>24.78700</td>
</tr>
<tr>
<td>H</td>
<td>-1</td>
<td>-21.95600</td>
<td>26.37900</td>
</tr>
<tr>
<td>N</td>
<td>-1</td>
<td>-28.37500</td>
<td>25.55600</td>
</tr>
<tr>
<td>H</td>
<td>-1</td>
<td>-27.64300</td>
<td>25.28400</td>
</tr>
<tr>
<td>C</td>
<td>-1</td>
<td>-28.04500</td>
<td>26.29800</td>
</tr>
<tr>
<td>H</td>
<td>-1</td>
<td>-28.36800</td>
<td>25.80700</td>
</tr>
<tr>
<td>C</td>
<td>-1</td>
<td>-26.46400</td>
<td>26.35300</td>
</tr>
<tr>
<td>H</td>
<td>-1</td>
<td>-26.04000</td>
<td>26.80900</td>
</tr>
<tr>
<td>H</td>
<td>-1</td>
<td>-26.02900</td>
<td>25.35800</td>
</tr>
<tr>
<td>O</td>
<td>-1</td>
<td>-26.12800</td>
<td>27.15300</td>
</tr>
<tr>
<td>H</td>
<td>-1</td>
<td>-25.98700</td>
<td>26.59000</td>
</tr>
<tr>
<td>C</td>
<td>-1</td>
<td>-28.66500</td>
<td>27.68400</td>
</tr>
<tr>
<td>O</td>
<td>-1</td>
<td>-28.33800</td>
<td>28.48500</td>
</tr>
<tr>
<td>H</td>
<td>0</td>
<td>-29.34566</td>
<td>27.96027</td>
</tr>
<tr>
<td>H</td>
<td>0</td>
<td>-27.80934</td>
<td>19.09490</td>
</tr>
</tbody>
</table>
Appendix D: Input file of direct hydrogen bond F62:O—S66:N(OG) with optimized hydrogen cap to get excitation energy

%NProcShared=4
%nCore=300HW
%n B3LYP/6-31G(d,p) td

<table>
<thead>
<tr>
<th>Title</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>-2.52375</td>
<td>-1.31265</td>
<td>-2.12669</td>
</tr>
<tr>
<td>H</td>
<td>-2.80962</td>
<td>-2.30137</td>
<td>-2.02624</td>
</tr>
<tr>
<td>C</td>
<td>-3.55197</td>
<td>-0.59979</td>
<td>-1.05320</td>
</tr>
<tr>
<td>H</td>
<td>-4.37434</td>
<td>0.09469</td>
<td>-1.40545</td>
</tr>
<tr>
<td>C</td>
<td>-4.07574</td>
<td>-1.70468</td>
<td>-0.01676</td>
</tr>
<tr>
<td>H</td>
<td>-4.36544</td>
<td>-1.27276</td>
<td>0.95797</td>
</tr>
<tr>
<td>H</td>
<td>-3.21421</td>
<td>-2.39968</td>
<td>0.13611</td>
</tr>
<tr>
<td>C</td>
<td>-5.24064</td>
<td>-2.42560</td>
<td>-0.70184</td>
</tr>
<tr>
<td>C</td>
<td>-5.37979</td>
<td>-3.82329</td>
<td>-0.89850</td>
</tr>
<tr>
<td>H</td>
<td>-4.57688</td>
<td>-4.51749</td>
<td>-0.65287</td>
</tr>
<tr>
<td>C</td>
<td>-6.63440</td>
<td>-4.32764</td>
<td>-1.22659</td>
</tr>
<tr>
<td>H</td>
<td>-6.83142</td>
<td>-5.38502</td>
<td>-1.33398</td>
</tr>
<tr>
<td>C</td>
<td>-7.71046</td>
<td>-3.49942</td>
<td>-1.55590</td>
</tr>
<tr>
<td>H</td>
<td>-8.68347</td>
<td>-3.92995</td>
<td>-1.78371</td>
</tr>
<tr>
<td>C</td>
<td>-6.41193</td>
<td>-1.62461</td>
<td>-0.86140</td>
</tr>
<tr>
<td>H</td>
<td>-6.36168</td>
<td>-0.57383</td>
<td>-0.61782</td>
</tr>
<tr>
<td>C</td>
<td>-7.62852</td>
<td>-2.15168</td>
<td>-1.20121</td>
</tr>
<tr>
<td>H</td>
<td>-8.52275</td>
<td>-1.55218</td>
<td>-1.28763</td>
</tr>
<tr>
<td>C</td>
<td>-2.53721</td>
<td>0.29288</td>
<td>-0.27192</td>
</tr>
<tr>
<td>O</td>
<td>-2.64140</td>
<td>1.55218</td>
<td>-0.24353</td>
</tr>
<tr>
<td>C</td>
<td>0.56636</td>
<td>0.35288</td>
<td>-0.32891</td>
</tr>
<tr>
<td>N</td>
<td>1.96906</td>
<td>-0.51636</td>
<td>-0.12771</td>
</tr>
<tr>
<td>N</td>
<td>1.20773</td>
<td>1.07080</td>
<td>-1.56007</td>
</tr>
<tr>
<td>C</td>
<td>2.51503</td>
<td>0.69518</td>
<td>-2.07539</td>
</tr>
<tr>
<td>O</td>
<td>3.01590</td>
<td>1.09362</td>
<td>-3.12326</td>
</tr>
<tr>
<td>C</td>
<td>2.57037</td>
<td>-0.36765</td>
<td>-1.11630</td>
</tr>
<tr>
<td>C</td>
<td>4.20439</td>
<td>-1.04765</td>
<td>-1.31890</td>
</tr>
<tr>
<td>H</td>
<td>4.51255</td>
<td>-0.68712</td>
<td>-2.31169</td>
</tr>
<tr>
<td>C</td>
<td>5.07525</td>
<td>-1.93594</td>
<td>-0.64494</td>
</tr>
<tr>
<td>C</td>
<td>6.25172</td>
<td>-2.44138</td>
<td>-1.41853</td>
</tr>
<tr>
<td>H</td>
<td>6.42206</td>
<td>-2.10560</td>
<td>-2.45376</td>
</tr>
<tr>
<td>C</td>
<td>4.88827</td>
<td>-2.50421</td>
<td>0.68549</td>
</tr>
<tr>
<td>H</td>
<td>4.15941</td>
<td>-2.27181</td>
<td>1.47662</td>
</tr>
<tr>
<td>C</td>
<td>7.23952</td>
<td>-3.11079</td>
<td>-0.88662</td>
</tr>
<tr>
<td>H</td>
<td>8.06428</td>
<td>-3.40735</td>
<td>-1.51838</td>
</tr>
<tr>
<td>C</td>
<td>5.55104</td>
<td>-3.39109</td>
<td>1.11512</td>
</tr>
<tr>
<td>H</td>
<td>5.72863</td>
<td>-3.73063</td>
<td>2.13670</td>
</tr>
<tr>
<td>C</td>
<td>7.04403</td>
<td>-3.82573</td>
<td>0.35308</td>
</tr>
<tr>
<td>O</td>
<td>7.83681</td>
<td>-4.58330</td>
<td>0.98431</td>
</tr>
<tr>
<td>C</td>
<td>0.31934</td>
<td>1.99969</td>
<td>-2.24844</td>
</tr>
<tr>
<td>H</td>
<td>0.56971</td>
<td>1.92529</td>
<td>-3.29579</td>
</tr>
<tr>
<td>H</td>
<td>-0.69690</td>
<td>1.68301</td>
<td>-2.03175</td>
</tr>
<tr>
<td>C</td>
<td>0.48480</td>
<td>3.40537</td>
<td>-1.81202</td>
</tr>
<tr>
<td>O</td>
<td>1.00410</td>
<td>4.23620</td>
<td>-2.45763</td>
</tr>
<tr>
<td>N</td>
<td>-1.31204</td>
<td>-0.28851</td>
<td>-0.19360</td>
</tr>
<tr>
<td>C</td>
<td>-0.31600</td>
<td>0.36456</td>
<td>0.53245</td>
</tr>
<tr>
<td>C</td>
<td>-0.07008</td>
<td>-0.43543</td>
<td>1.78721</td>
</tr>
<tr>
<td>H</td>
<td>0.51053</td>
<td>-0.19669</td>
<td>2.25252</td>
</tr>
<tr>
<td>H</td>
<td>0.02657</td>
<td>-1.45266</td>
<td>1.42693</td>
</tr>
<tr>
<td></td>
<td>-1.11470</td>
<td>-0.25489</td>
<td>2.99223</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>H</td>
<td>-1.08328</td>
<td>0.77320</td>
<td>3.31219</td>
</tr>
<tr>
<td>H</td>
<td>-2.10210</td>
<td>-0.46369</td>
<td>2.42979</td>
</tr>
<tr>
<td>S</td>
<td>-0.86777</td>
<td>-1.36064</td>
<td>4.26114</td>
</tr>
<tr>
<td>C</td>
<td>-1.95186</td>
<td>-0.52583</td>
<td>5.41513</td>
</tr>
<tr>
<td>H</td>
<td>-1.93165</td>
<td>-0.90254</td>
<td>6.45984</td>
</tr>
<tr>
<td>H</td>
<td>-2.59162</td>
<td>-0.61339</td>
<td>5.03368</td>
</tr>
<tr>
<td>H</td>
<td>-1.71931</td>
<td>0.56028</td>
<td>5.42011</td>
</tr>
<tr>
<td>N</td>
<td>0.01740</td>
<td>3.72992</td>
<td>-0.60561</td>
</tr>
<tr>
<td>H</td>
<td>-0.34752</td>
<td>2.95679</td>
<td>-0.09368</td>
</tr>
<tr>
<td>C</td>
<td>-0.04371</td>
<td>5.08410</td>
<td>-0.01852</td>
</tr>
<tr>
<td>H</td>
<td>-0.58406</td>
<td>5.79786</td>
<td>-0.62102</td>
</tr>
<tr>
<td>C</td>
<td>-0.89378</td>
<td>5.02750</td>
<td>1.32436</td>
</tr>
<tr>
<td>H</td>
<td>-0.42784</td>
<td>4.27317</td>
<td>1.99370</td>
</tr>
<tr>
<td>H</td>
<td>-1.97048</td>
<td>4.83359</td>
<td>1.13415</td>
</tr>
<tr>
<td>O</td>
<td>-0.87598</td>
<td>6.30652</td>
<td>1.95531</td>
</tr>
<tr>
<td>H</td>
<td>-1.66391</td>
<td>6.78596</td>
<td>1.71307</td>
</tr>
<tr>
<td>C</td>
<td>1.37000</td>
<td>5.59075</td>
<td>0.21607</td>
</tr>
<tr>
<td>O</td>
<td>2.16337</td>
<td>5.01191</td>
<td>0.94595</td>
</tr>
<tr>
<td>H</td>
<td>1.64960</td>
<td>6.52776</td>
<td>-0.30557</td>
</tr>
<tr>
<td>H</td>
<td>-3.35563</td>
<td>-1.13338</td>
<td>-3.02465</td>
</tr>
</tbody>
</table>
VITA

CHANDRA PRASAD DHAKAL

2015-2021 Graduate Teaching Assistant (Ph.D. Student)
Department of Physics, Florida International University
Miami, FL, 33199, USA

2013-2015 Graduate Teaching/Research Assistant (M.S. Student)
Department of Physics, University of Missouri at Kansas City
Kansas City, Mo, 64111, USA

2003-2005 M. Sc. in Solid State Physics, Tribhuvan University, Kathmandu, Nepal

2000-2003 B. Sc. in Physics, Tribhuvan University, Biratnagar, Nepal

Presentation: