Structural Flexibility and Oxygen Diffusion Pathways in Monomeric Fluorescent Proteins

Chola K. Regmi

FLORIDA INTERNATIONAL UNIVERSITY, cholaregmi@gmail.com

DOI: 10.25148/etd.FI14040821

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

Part of the Biological and Chemical Physics Commons

Recommended Citation


https://digitalcommons.fiu.edu/etd/1122

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.
A dissertation submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

by

Chola K. Regmi

2014
To: Dean Kenneth G. Furton  
College of Arts and Sciences  

This dissertation, written by Chola K. Regmi, and entitled Structural Flexibility and Oxygen Diffusion Pathways in Monomeric Fluorescent Proteins, having been approved in respect to style and intellectual content, is referred to you for judgement.

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

______________________________
Jaroslava Miksovska

______________________________
Xuewen Wang

______________________________
Prem P. Chapagain, Co-Major Professor

______________________________
Bernard S. Gerstman, Co-Major Professor

Date of Defense: March 26, 2014

The dissertation of Chola K. Regmi is approved.

______________________________
Dean Kenneth G. Furton  
College of Arts and Sciences

______________________________
Dean Lakshmi N. Reddi  
University Graduate School

Florida International University, 2014
DEDICATION

I dedicate this dissertation to my beloved parents.
ACKNOWLEDGMENTS

I would like to express great appreciation to my advisors Professor Bernard Gerstman and Professor Prem Chapagain for their invaluable guidance, help, and support during my entire research work and dissertation writing.

I would like to thank my dissertation committee members Professor Xuewen Wang and Professor Jaroslava Miksovska for their helpful suggestions and comments.

I am thankful to the Biophysics Research Group members Yuba, Jeevan, Tim, and Ilan for their help and cooperation.

I would like to thank my parents, brothers and sisters who always supported and encouraged me towards excellence during the long journey of my education. I am also thankful to my parents-in-law for providing support and care during my graduate studies and research.

Finally, I would like to thank my beloved wife Shailaja for her care, support and understanding throughout these years. Words cannot express my love for my son Aarnes who lived with his grandparents during part of my graduate years when Shailaja was also doing her graduate studies at Ithaca College, NY.
Fluorescent proteins are valuable tools as biochemical markers for studying cellular processes. Red fluorescent proteins (RFPs) are highly desirable for *in vivo* applications because they absorb and emit light in the red region of the spectrum where cellular autofluorescence is low. The naturally occurring fluorescent proteins with emission peaks in this region of the spectrum occur in dimeric or tetrameric forms. The development of mutant monomeric variants of RFPs has resulted in several novel FPs known as mFruits. Though oxygen is required for maturation of the chromophore, it is known that photobleaching of FPs is oxygen sensitive, and oxygen-free conditions result in improved photostabilities. Therefore, understanding oxygen diffusion pathways in FPs is important for both photostabilities and maturation of the chromophores. We used molecular dynamics calculations to investigate the protein barrel fluctuations in mCherry, which is one of the most useful monomeric mFruit variants, and its GFP homolog citrine. We employed implicit ligand sampling and locally enhanced sampling to determine oxygen pathways from the bulk solvent into the mCherry chromophore in the interior of
the protein. The pathway contains several oxygen hosting pockets, which were identified by the amino acid residues that form the pocket. We calculated the free-energy of an oxygen molecule at points along the path. We also investigated an RFP variant known to be significantly less photostable than mCherry and find much easier oxygen access in this variant. We showed that oxygen pathways can be blocked or altered, and barrel fluctuations can be reduced by strategic amino acid substitutions. The results provide a better understanding of the mechanism of molecular oxygen access into the fully folded mCherry protein barrel and provide insight into the photobleaching process in these proteins.
# 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. METHODS</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Molecular Mechanics: Potential Energy Functions</td>
<td>5</td>
</tr>
<tr>
<td>2.1.1 Bonded Interactions and Bond Related Interactions</td>
<td>7</td>
</tr>
<tr>
<td>2.1.2 Non-bonded Interactions</td>
<td>9</td>
</tr>
<tr>
<td>2.2 Molecular Mechanics: Time Integration</td>
<td>11</td>
</tr>
<tr>
<td>2.2.1 Verlet Algorithm</td>
<td>12</td>
</tr>
<tr>
<td>2.2.2 Leap-frog Algorithm</td>
<td>13</td>
</tr>
<tr>
<td>2.2.3 Velocity Verlet Algorithm</td>
<td>14</td>
</tr>
<tr>
<td>2.3 Molecular Dynamics: System Setup</td>
<td>14</td>
</tr>
<tr>
<td>2.4 Force Field Parameters for Chromophore</td>
<td>19</td>
</tr>
<tr>
<td>2.5 Implicit Ligand Sampling (ILS)</td>
<td>19</td>
</tr>
<tr>
<td>2.6 Locally Enhanced Sampling (LES)</td>
<td>20</td>
</tr>
<tr>
<td>2.7 Simulation Conditions and System Setup</td>
<td>21</td>
</tr>
<tr>
<td>3. FLUORESCENT PROTEIN BARREL FLEXIBILITY FOR VARIOUS CHROMOPHORE STATES</td>
<td>23</td>
</tr>
<tr>
<td>3.1 Fluorescent Protein Structure</td>
<td>24</td>
</tr>
<tr>
<td>3.2 Different Chemical States of the Chromophore and the Chromophore Maturation Process</td>
<td>25</td>
</tr>
<tr>
<td>3.2.1 Tripeptide, Neutral, and Anionic States of Chromophore</td>
<td>27</td>
</tr>
<tr>
<td>3.3 Root Mean Square Fluctuations (RMSF) of the Protein Barrel and Chromophore Atoms</td>
<td>29</td>
</tr>
<tr>
<td>3.4 Investigations of the Gap Between ( \beta7-\beta10 ) Barrel Strands</td>
<td>31</td>
</tr>
<tr>
<td>3.5 Correlation Between ( \tau ) and ( \phi ) Dihedral Angles</td>
<td>33</td>
</tr>
<tr>
<td>4. MOLECULAR OXYGEN DIFFUSION IN mCHERRY AND CITRINE: IMPLICIT LIGAND SAMPLING</td>
<td>37</td>
</tr>
<tr>
<td>4.1 Implicit Ligand Sampling (ILS) for Molecular Oxygen</td>
<td>39</td>
</tr>
<tr>
<td>4.2 Importance of Sidechains in Controlling Gap Size</td>
<td>42</td>
</tr>
<tr>
<td>4.3 Amino Acid Mutations Made in mCherry to Control Dioxygen Access to the Chromophore</td>
<td>45</td>
</tr>
<tr>
<td>5. MOLECULAR OXYGEN DIFFUSION IN mCHERRY: LOCALLY ENHANCED SAMPLING</td>
<td>49</td>
</tr>
<tr>
<td>5.1 Molecular Oxygen Diffusion Pathways in mCherry</td>
<td>52</td>
</tr>
<tr>
<td>5.2 Free-Energies Along the Pathway Calculated from Implicit Ligand Sampling</td>
<td>60</td>
</tr>
<tr>
<td>5.3 M163Q Mutation and the Enhanced Photostability of mCherry</td>
<td>61</td>
</tr>
<tr>
<td>FIGURES</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 2.1: Bonded and non-bonded interactions</td>
<td>6</td>
</tr>
<tr>
<td>Figure 3.1: Fluorescent protein structure</td>
<td>24</td>
</tr>
<tr>
<td>Figure 3.2: Chromophore structure of mCherry</td>
<td>25</td>
</tr>
<tr>
<td>Figure 3.3: Chemical mechanism of chromophore maturation</td>
<td>26</td>
</tr>
<tr>
<td>Figure 3.4: Different states of chromophore</td>
<td>27</td>
</tr>
<tr>
<td>Figure 3.5: Root mean square fluctuations of Cα atoms of protein barrel</td>
<td>29</td>
</tr>
<tr>
<td>Figure 3.6: Root mean square fluctuations of atoms in imidazole ring and phenyl ring of chromophore</td>
<td>30</td>
</tr>
<tr>
<td>Figure 3.7: The β7-β10 region of mCherry (red) and citrine (yellow)</td>
<td>31</td>
</tr>
<tr>
<td>Figure 3.8: Gap fluctuations between β7-β10 strands in different chromophore states</td>
<td>32</td>
</tr>
<tr>
<td>Figure 3.9: Chromophore dihedral angles τ and φ</td>
<td>34</td>
</tr>
<tr>
<td>Figure 3.10: Probability distribution for chromophore dihedral angles τ and φ from MD simulations</td>
<td>36</td>
</tr>
<tr>
<td>Figure 4.1: Free energy isosurfaces for molecular dioxygen</td>
<td>40</td>
</tr>
<tr>
<td>Figure 4.2: Free energy values of dioxygen at locations along the pathways</td>
<td>41</td>
</tr>
<tr>
<td>Figure 4.3: Fluctuations in the β7-β8 gap in mCherry</td>
<td>43</td>
</tr>
<tr>
<td>Figure 4.4: Free-energy plot (in kcal/mol) of the β7-β8 strands</td>
<td>44</td>
</tr>
<tr>
<td>Figure 4.5: β7-β10 gap for mCherry compared to its mutant</td>
<td>46</td>
</tr>
<tr>
<td>Figure 5.1: Various oxygen-hosting pockets in the mCherry protein barrel</td>
<td>53</td>
</tr>
<tr>
<td>Figure 5.2: Trajectories of molecular oxygen showing its distance from the chromophore</td>
<td>55</td>
</tr>
</tbody>
</table>
Figure 5.3: (a) Volumetric trace of the oxygen diffusion channel from one side of the protein to the other. (b) Amino acid residues involved in the oxygen diffusion channel.

Figure 5.4: Free-energy values for the oxygen molecule at locations along the curved pathway.

Figure 5.5: Distribution and schematic of side chain dihedral angles for the gateway residue GLN163 in mCherry versus MET163 in mCherry-Q163M.
# 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>EGFP</td>
<td>Enhanced Green Fluorescent Protein</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>LJ</td>
<td>Lennard-Jones</td>
</tr>
<tr>
<td>PME</td>
<td>Particle Mesh Ewald</td>
</tr>
<tr>
<td>PBC</td>
<td>Periodic Boundary Condition</td>
</tr>
<tr>
<td>VDW</td>
<td>Van der Waals</td>
</tr>
<tr>
<td>LES</td>
<td>Locally Enhanced Sampling</td>
</tr>
<tr>
<td>RMSF</td>
<td>Root Mean Square Fluctuations</td>
</tr>
<tr>
<td>KFP</td>
<td>Kindling Fluorescent Protein</td>
</tr>
<tr>
<td>FRET</td>
<td>Forster Resonance Energy Transfer</td>
</tr>
<tr>
<td>ILS</td>
<td>Implicit Ligand Sampling</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

The development of fluorescent proteins (FPs) has revolutionized cell biology research (1-3). Fluorescent proteins are very useful tools for bio-imaging of molecules, cells, and the bodies of living animals. A green fluorescent protein (GFP) was discovered in the 1960s (4) from the jellyfish *Aequorea victoria* and the GFP was used as a cell marker in the early 1990s when scientists were able to expressed its gene in other cells (3, 5). Now, these FPs have a wide range of applications in cell biology, such as protein labeling, promoter tracking, timing, Forster resonance energy transfer (FRET), and biosensors (1, 6-8).

After the successful expression of the gene for GFP, researchers focused on new and improved variants that are brighter, more photostable, monomeric, faster maturing, and covering a broad spectral range of visible light (8). The S65T (serine to threonine) amino acid mutation in the chromophore of the wild type GFP changed the double excitation peak to a single excitation peak and generated a much brighter variant called the enhanced green fluorescent protein (EGFP) (9). A Y66H mutation on the second residue (tyrosine) of the chromophore of the wild type GFP yielded a blue fluorescent protein (BFP) variant, and Y66W yielded a cyan fluorescent protein (CFP) variant (10). A T203Y mutation in the GFP protein barrel yielded a yellow fluorescent protein (YFP) (11). However, orange and red spectra were not achieved from mutations in the wild type of GFP. The first discovery of a red fluorescent protein (RFP) was from the non-bioluminescent Anthozoa species in 1999 (12), which added more colors to the palettes in cell biology. Recently, Kumagai et al. (6) reported a new class of fluorescent protein,
UnaG from a Japanese freshwater eel species, which is the first fluorescent protein from vertebrates. The fluorescent mechanism is not same as in GFP, which fluoresces through a chromophore, whereas UnaG fluoresces when it binds with the small molecule bilirubin. The applications of UnaG are not yet fully tested.

In the present work, we focus on the monomeric variants of red fluorescent proteins (mRFPs). They are especially valuable as markers for in vivo applications in mammalian cell imaging (13) because they absorb and emit light in the red region of the spectrum where cellular autofluorescence is low (14). As monomers, the mRFPs are fairly light-weight and therefore do not perturb the processes of the molecules in the cells that are being investigated. In contrast, the naturally occurring fluorescent proteins with emission peaks in this region of the spectrum occur in dimeric or tetrameric forms (12, 15), which tend to oligomerize (16, 17) and render them unsuitable for fusion tagging (18) because they cause significant perturbations to normal cellular activities. The development of mutant monomeric variants of RFPs to avoid these issues has resulted in several novel monomeric FPs known as mFruits (19). Some of the most promising mFruits are mCherry, mOrange and mStrawberry (20), whose names reflect the wavelengths of their corresponding emission spectra.

Applications of mRFPs would be expanded through the development of variants with higher photostability. Better shielding of the chromophore from the environment to reduce the access of molecular oxygen to the chromophore has been shown to significantly increase the photostabilities of both GFPs and RFPs (21). The Q64H and F99Y mutations introduced in mOrange resulted in the significantly more photostable mOrange2 variant, possibly because these mutations help to block chromophore
detrimental oxidation by rearranging the structure of the protein-chromophore environment (22). This suggests that irreversible photobleaching can occur as a result of the diffusion of molecular oxygen through the protein barrel surface (16) into the protein interior, in addition to the irreversible photobleaching from transient dark states produced by photoisomerization or excited state proton transfer. Though oxygen access to the mature chromophore can ruin fluorescence, oxygen access is necessary in earlier stages for chromophore formation and maturation. Therefore, understanding of the details of oxygen diffusion pathways in FPs is important from the perspectives of both photostability and chromophore maturation.

The competing effects of oxygen on mFRPs, required for beneficial maturation of the chromophore but detrimental by causing photobleaching, poses limitations to the next generation of single-molecule spectroscopy and low-copy fluorescence microscopy experiments. Therefore, improving the photostability of the mFruits is highly desirable. Increasing evidence suggests that protein flexibility plays a major role in gas access into many proteins (23-28), and dynamic fluctuations in the size of transient cavities due to residues’ thermal fluctuations are the determining factor in the pathways of gas diffusion (29-32). For example, oxygen diffusion in myoglobin’s distal pocket has been extensively studied, both experimentally and by simulations, in light of the influence of different protein conformations or mutations (24, 33). The interaction between the chromophore and the surrounding protein has important implications for both parts of the protein (34). The electronic molecular orbitals of the chromophore that are responsible for its spectral properties may be modified by the surrounding protein and affect the
spectral properties and the lifetime of the fluorescence (35). Also, the structural fluctuations of the protein may be modified by interactions with the chromophore.

We used molecular dynamics (MD) computer simulations to investigate the dynamic structural behavior of the fluorescent protein barrel in mCherry and citrine, and to determine the role that flexibility of the protein barrel plays in molecular oxygen diffusion through the barrel to the chromophore. Chapter 2 describes the molecular dynamics computer simulation techniques. Chapter 3 describes our results concerning the structural flexibility of the protein barrel in different chromophore maturation states. Chapter 4 describes the oxygen diffusion pathways in mCherry and citrine that we uncovered using implicit ligand sampling computational techniques. Chapter 5 describes the oxygen diffusion pathways using the locally enhanced sampling computational method.
2. METHODS

2.1 MOLECULAR MECHANICS: POTENTIAL ENERGY FUNCTIONS

The understanding of the relationship between molecular structure, dynamics, and function in biological macromolecules is greatly facilitated by Molecular Dynamics (MD) simulations (36). Quantum mechanical (QM) treatment of a system gives more detailed properties of a molecular system, and is possibly more accurate than MD, but QM calculations are expensive in terms of computation and not feasible for the systems size of fluorescent proteins. MD simulation, which uses basically classical calculations, is a widely used technique to study the macromolecular properties. Though the results are not as detailed as QM calculations, it is possible to investigate important dynamics of a macromolecule within a reasonable time of computation. Potential energy functions, also called FORCE FIELDS, used in MD calculations are approximations to the exact potential energies experienced by the atoms. The accuracy of the force fields employed compared to the real potentials are critically important in determining the accuracy of the model (37) in simulating to the true behavior of a system.

Numerical values for the parameters used in the force fields are derived using quantum mechanical calculations or obtained by fitting expressions to reproduce experimental results (38) such as high resolution crystal structures.

A well-used MD computational simulation package is CHARMM (37). The CHARMM potential energy function is given in equation 2.1. Bonded interactions between atoms (figure 2.1) are modeled as harmonic potential functions except for the dihedral energy term. Non-bond interactions are also included.
\[ U(R) = \sum_{\text{bonds}} K_b \left( b - b_0 \right)^2 + \sum_{\text{angles}} K_\theta \left( \theta - \theta_0 \right)^2 + \sum_{\text{UB}} K_{UB} \left( S - S_0 \right)^2 + \sum_{\text{dihedrals}} K_\varphi \left( 1 + \cos(n\varphi - \delta) \right) + \sum_{\text{impropers}} K_\omega \left( \omega - \omega_0 \right)^2 + \]

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

\[ \text{residuesUCMAP}(\varphi, \psi) \quad (2.1) \]

where \( K_b \) is the force constant for bond stretching, \( K_\theta \) is for bond angle changes, \( K_{UB} \) is for the non-bonded Urey-Bradley energy also related to bond-angle bending, \( K_\varphi \) is for dihedral angle rotations, and \( K_\omega \) is for improper dihedral angle changes. Similarly \( b_0, \theta_0, S_0, \text{ and } \omega_0 \) are the respective equilibrium values. All of these terms are explained in more detail below. 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.

Figure 2.1 Bonded and non-bonded interactions
2.1.1 BONDED INTERACTIONS AND BOND RELATED INTERACTIONS

Bond Stretching Energy

The first term in equation 2.1, \( \sum_{bonds} K_b (b - b_0)^2 \) gives the bond stretching energy between two covalently bonded atoms. The strength of the bond is determined by the force constant \( K_b \). The value of the equilibrium bond length \( b_0 \) and the force constant \( K_b \) are specific to the identities of the atoms in the pair. Figure 2.1 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.

Angle Bending Energy

The angle bending energy is the energy required to bend an angle formed by two bonds; for example, in figure 2.1 a bond angle can be defined by the pairs i-j and j-k. The second term in equation 2.1, \( \sum_{angles} K_\theta (\theta - \theta_0)^2 \) represents the angle bending potential, which also has an harmonic form. \( K_\theta \) and \( \theta_0 \) are the force constant and equilibrium bond angle, respectively. The bond stretching and angle bending potentials provide the energy necessary to cause a deviation from the equilibrium geometry of the system, which is presumed to have the lowest energy.

Urey-Bradley Energy

The third term in equation 2.1, \( \sum_{UB} K_{UB} (S - S_0)^2 \) is called the Urey-Bradley energy term. This is an energy term that was added in the development of the CHARMM force field to attain better agreement with experimental data on vibrational frequencies. This potential is also represented as a harmonic potential, with force constant \( K_{UB} \). The equilibrium distance \( S_0 \) is the distance between the first and third atom in a system of two
bonds, such as between atoms i and k in figure 2.1. It is not included (some $K_{UB}=0$) for all i, j, k bond angles; $K_{UB}\neq0$ only when the identity of the atoms i, j, k requires it, along with the improper energy, to optimize the fit to vibrational spectra and out of plane motions.

**Torsional Energy (bond rotation)**

The fourth term in equation 2.1, $\sum_{\text{dihedrals}} K_{\phi} (1 + \cos(n \phi - \delta))$ is called the torsional energy of the system, which arises from the rotation about a bond. This potential function comes to play when a rotation occurs around the middle bond of four covalent atoms. For the four bonded atoms i-j-k-l as shown in figure 2.1, the angle between the plane i-j-k and j-k-l is called the torsion angle. This potential function is periodic in nature; hence the cosine function. In equation 2.1, $K_{\phi}$ is the constant which gives the strength of the potential and the barrier height resisting full rotation, $\phi$ is the torsion angle, $\delta$ is an angular phase factor that reflects the most favorable torsional angle, and $n$ is the multiplicity representing the number of potential energy barriers experienced in one full rotation of the angle $\phi$.

**Improper Energy**

The fifth term in equation 2.1, $\sum_{\text{impropers}} K_{\omega} (\omega - \omega_0)^2$ denotes the improper dihedral angle energy. This energy term is used to maintain planarity or chirality of the system composed of four atoms, i-j-k-l. The improper angle $\omega$ is often defined identically to the torsional angle $\phi$ for the four atoms, but the form of the potential and the intent are different. The torsional potential described above represents rotations around a bond that may have multiple minimum, whereas the improper term represents a tendency that some systems have to remain planar. Unlike the torsional term, the improper potential term is also defined as an harmonic potential form with force constant $K_{\omega}$.
CMAP Correction

The last term in equation 2.1, is $\sum_{\text{residues}} U_{\text{CMAP}}(\varphi, \psi)$ called CMAP (correction map) term which is a numerical value of energy added recently to the CHARMM force field to improve the ability to match backbone dihedral angles $\varphi$ and $\psi$ to experimentally observed secondary structure. The latest version of CHARMM, CHARMM36, includes this phi-psi backbone dihedral CMAP correction term, which was not included in earlier versions, such as CHARMM22, and also includes side-chain dihedral angles optimization. This optimization balances the sampling between $\alpha$-helices and $\beta$ regions and corrected the $\alpha$-helical bias of CHARMM22 force fields (39).

2.1.2 NON-BONDED INTERACTIONS

The sixth term in equation 2.1, $\Sigma_{\text{non-bonded}} \left\{ \epsilon \left( \frac{\sigma}{r_{ij}} \right)^{12} - 2 \left( \frac{\sigma}{r_{ij}} \right)^{6} \right\} + \frac{q_i q_j}{4 \pi \varepsilon_0 \varepsilon r_{ij}}$ which contains the van der Waals (VDW) interactions term and the electrostatic interaction term, represent the non-bonded interactions in the force field. The van der Waals interactions are represented by a 6-12 Lennard-Jones (LJ) potential and the electrostatic interaction is represented by the Coulomb potential. All atom pairs contribute to the non-bonded energy in the potential function via the VDW interaction. The repulsive $1/r^{12}$ part of the van der Waals interactions can be explained on the basis of the Pauli exclusion principle, and the attractive $1/r^{6}$ part is due to induced dipole-dipole interactions. The electrostatic interaction only occurs between atoms with a net charge, $q_i$. In equation 2.1, $\sigma$ is the collision diameter between two atoms in the LJ term, $\epsilon$ is the LJ potential well depth, and $r_{ij}$ is the intra-atomic distance. In the CHARMM force field the hydrogen bond interactions are accounted through electrostatic and van der Waals...
interactions and therefore there is no explicit energy term for the hydrogen bond interaction in equation 2.1. There are $N^2/2$ distinct atom pairs to compute the non-bonded interactions in a system of $N$ atoms. This large number of interactions makes computations of non-bonded interactions in MD simulations more expensive in terms of computational time than bonded interactions. In general, distance cut-offs are used to limit the number of non-bonded interactions that are calculated; atoms separated by a distance greater than this cut-off have their non-bonded interactions set equal to zero without a calculation, thus saving computational time. Since the non-bonded interactions decrease in magnitude as the separation distance $r_{ij}$ increases, a distance cut-off is set large enough, e.g. 12Å, so that atoms separated by this distance have non-bonded interactions that are assumed to be weak enough to ignore.

The empirical potential functions used in classical MD simulations are a compromise between accuracy and efficiency. There are several limitations in the potential functions that give inaccurate calculations of energies and forces in the biomolecular system. One of the major limitations is the setting of fixed charges on the atoms throughout the simulations, which cannot incorporate changes in atomic charges due to reaction mechanisms in a system. In systems where changes in charge are known to be important, more detailed computational approaches are use such as hybrid simulations combining Quantum Mechanical with Molecular Mechanical (QM/MM) models. In the systems that we explore, this is not necessary and our work employs only MM techniques. A crucial aspect of MM techniques is the correct algorithms for the discrete steps through time.
2.2 MOLECULAR MECHANICS: TIME INTEGRATION

Newton’s Second Law of Motion is the fundamental principle in MD simulations. Integration with respect to time of the equations of motion gives the time trajectories of the atoms. Discrete time steps are used in MM simulations. The simulations provide the atoms’ positions, velocities and accelerations at each time. The acceleration of each particle at each time step is necessary to calculate the next position and velocity of the particles. The acceleration of any particle at a time $t$ can be calculated using $a = \frac{F}{m}$, where $F$ is the force acting on the particle at time $t$ and $m$ is the mass of the particle. The gradient of the potential functions expressed in equation 2.1 gives the force acting on the particle at a given time through the relation, $F = -\frac{du}{dr}$. If an initial configuration of a protein is given that contains the $(x, y, z)$ coordinates of each atom, then $U$ can be calculated from equation 2.1 for each particle and then the force experienced by each particle. If the initial velocities of each particle are also provided, then we can determine the future evolution of their positions and velocities with time. Thus, MD simulations are deterministic. In actual MD simulations, velocities are sometimes given random changes, which removes determinacy from the simulations. However, if the random velocity increments are assigned to each particle by following a specific physical distribution, such as the Maxwell-Boltzmann velocity distribution, the dynamics of a protein should be similar from one simulation to the next if both simulations are initialized with the same spatial configuration.

In MD simulations of biomolecules, the initial positions of the atoms can be obtained from x-ray crystallographic structures or NMR structures, and initial velocities
are randomly chosen using a Maxwell-Boltzmann or Gaussian distributions for the desired temperature. Due to the complex nature of the biomolecules, there is no analytical solution to a Hamiltonian incorporating the potential function of equation 2.1. Instead, numerical algorithms with time increments are used in MD simulations. These integration algorithms are based on the assumption that positions, velocities, and accelerations can be expanded in terms of Taylor series expansions. Below are some commonly used algorithms for performing the numerical, step-by-step time integration.

2.2.1 VERLET ALGORITHM

Let $\Delta t$ be the small time step. Then the Taylor series expansion of position as a function of time can be written as:

$$r(t + \Delta t) = r(t) + \frac{dr}{dt} \Delta t + \frac{1}{2} \frac{d^2r}{dt^2} (\Delta t)^2 + \cdots$$

Similarly,

$$r(t - \Delta t) = r(t) - \frac{dr}{dt} \Delta t + \frac{1}{2} \frac{d^2r}{dt^2} (\Delta t)^2 + \cdots$$

After addition of the above two equations, we get

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t)(\Delta t)^2$$

in which the first order term in $\Delta t$ does not appear. Equation (2.2) gives the position of a particle at a time $t + \Delta t$, by using the position and acceleration at the current time, $t$, and the position at the previous time, $t - \Delta t$. The advantage of the Verlet algorithm is that it is correct to second order in time $(\Delta t^2)$ but does not require a first order term in $\Delta t$ that depends on velocity explicitly to calculate the new position of the particle.
2.2.2 LEAP-FROG ALGORITHM

In the Leap-Frog algorithm, the position of the particle at time $t + \Delta t$ is calculated by using the relation

$$ r(t + \Delta t) = r(t) + v\left(t + \frac{\Delta t}{2}\right) \Delta t , \tag{2.3} $$

where the definition for the velocity is

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

The expansion of the velocity in the time step $\frac{\Delta t}{2}$ gives

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

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

Subtraction of these two equations gives

$$ v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + a(t) \Delta t , \tag{2.4} $$

where $a(t) = \frac{dv(t)}{dt}$. Equation (2.4) is used to calculate the velocity at time $\left(t + \frac{\Delta t}{2}\right)$ to be inserted in equation 2.3. With this velocity, the position at $(t + \Delta t)$ can be calculated using equation (2.3). The velocities at time $t$ can be approximated by the relationship

$$ v(t) = \frac{1}{2} \left[v\left(t + \frac{\Delta t}{2}\right) + v\left(t - \frac{\Delta t}{2}\right)\right] $$

The advantage of the Leap-Frog method is that $r(t + \Delta t)$ is calculated more accurately by using $v(t + \Delta t/2)$ and $a(t)$ than if $v(t)$ is used with $a(t)$. 

13
2.2.3 VELOCITY VERLET ALGORITHM

In the velocity Verlet algorithm, position and velocity at time \( t + \Delta t \) can be calculated by using the following relations.

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

In this algorithm, the acceleration at \( t \) and at \((t + \Delta t)\) are averaged to calculate positions and velocities. This algorithm is more accurate than the others.

2.3 MOLECULAR DYNAMICS: SYSTEM SETUP

In previous sections, we discussed the potential energy function and integration algorithms that are used with Newton’s equations of motions to create a trajectory of the motions of the atoms in the molecular system under investigation. As explained above, the computations require an initial \((t=0)\) configuration of the protein molecule containing the \((x, y, z)\) positions of all the atoms, as well as the initial velocity of each atom. In this section, we summarize the initialization of the MD simulation. The description provided is based on the CHARMM simulation package.

Initial Structure

In molecular dynamics we need an initial structure of the system of study to begin simulations. For the simulation of biomolecules, the initial conformation of a molecule can be obtained on-line from the Protein Data Bank (PDB). The structures contained in the PDB are obtained from x-ray crystallography or NMR methods and deposited in the PDB. The structural files obtained from the PDB are in a special format known as a pdb file. In order to use the pdb file for a specific protein in a MD simulation, a careful study
of the pdb file is required first. Some of the residues or atoms in the protein may be missing in the pdb file. Also, the x-ray structure of a protein contains none of the hydrogen atoms, which means that the hydrogen atoms are not present in the pdb file. The software package MODELLER (40) uses educated guesses from similar sequences in other proteins to add missing residues in the system. The HBUILD command in CHARMM adds the necessary hydrogen atoms to the molecular system.

Once the structure of the molecular system is complete, it may be necessary to change the notation for the names of some atoms to make the pdb file compatible with MD packages. The MMTSB software toolset (41) is a good package for the system setup for use with CHARMM.

Solvation

Solvation is the process at the beginning of the computational simulation in which biomolecules such as proteins and nucleic acids are surrounded by water to mimic the natural environment of biomolecules. The size of the water solvation box should be sufficiently large to have several layers of molecules outside every part of the protein, including when the protein changes its shape during the MD simulation. This may require a solvation box that is more than twice the length of the initial protein configuration in each direction (x, y, z), which can greatly increase the computational time. An alternative method that allows the use of a solvation box that is only a little bigger than the protein is to use periodic boundary conditions (PBC). In this approach, if residues approach the edge of the solvation box, they experience the solvation of water molecules located at the opposite face of the box. In order to prevent residues at one edge of the solvation box from experiencing aphysical interactions with residues on the other side of
the protein, the size of the solvation water box should be at least the length of the longest axis of the protein plus twice the cutoff distance for non-bonded interactions. This size also usually prevents water molecules at one edge of the box from experiencing interactions with residues from the other side of the protein unless the protein becomes especially elongated.

Neutralization

Some amino acids have a positive or negative electric charge. Macromolecules may have an overall net positive or negative charge. The Particle Mesh Ewald (PME) (42) method is the technique used to calculate long range electrostatic interactions when PBC is used. To use PME, the system should be overall neutral. In the neutralization step, counter ions are added to make the system neutral. Energy minimization is required after these steps.

Energy Minimization

After creating a complete pdb file and solvating the protein, it will be possible to start an MD simulation. However, the MD simulation may quickly crash after only a few time steps because the initial positions of the atoms may be physically unrealistic. The positions of the atoms based upon x-ray crystallography, MODELLER insertion of amino acids, and solvation involve guessing. If the initial guesses place atoms too close by even a fraction of an angstrom, the potential energy in some locations in the molecule calculated using equation 2.1 may result in forces and accelerations that are unphysically large. This will lead to unphysically large velocities and gigantic changes in positions that makes the MD simulations numerically unstable, which, for example, might break bonds that, in actuality, do not break. MD simulation packages have built-in tests to determine if
energies and energy gradients are unphysically large, and to stop (crash) the calculations if the test criteria are violated. To avoid these problems in the MD simulations, MD packages contain routines that move atoms by small amounts relative to their initial positions in order to minimize the molecule’s energy. This initial energy minimized configuration may be a local minimum and not the global energy minimum structure. However, this initial energy-minimized configuration is physically realistic and allows the full MD simulation to be performed. The energy minimization steps follow similar procedures as followed during the full MD simulations. However, during energy minimization, the MD program allows unphysically large energies and forces in order to make small rearrangements of the atoms. The PME criteria and potential energy cutoff parameters used during energy minimization should be exactly the same as for the planned MD simulation. There is no definitive number of steps that must be used during energy minimization of a system. Once the change in energy from one time step to the next time step becomes very small (<0.001 kcal/mol), the system is close to convergence to a local energy minimum and the minimization procedure can be terminated.

**Heating and Equilibration**

Once we have a neutral, solvated biomolecular system in its minimum energy configuration we are ready for the next steps before performing the actual MD simulation. If the atoms are assigned velocities according to a Maxwell-Boltzmann distribution corresponding to biologically relevant temperatures (~300K), some atoms may be given velocities that are large enough to cause unphysical motions of the atoms that lead to unrealistic changes in configuration. To avoid this, the velocities of the atoms are initially given small values corresponding to low temperature (<50K). The system is
then allowed to perform a certain number of time steps (e.g. 500 or 1000) in order to gently allow the atoms to interact and reach new equilibrium positions. The velocities of the atoms are then given increments corresponding to a small temperature increase (10K or 20K), and the atoms are then allowed to relax again. This process of adding velocity increments to heat the system is repeated until the actual simulation temperature is reached. This incremental heating prevents structural distortion and instability of the system due to sudden large changes of velocity at high temperature. When the system reaches the final temperature, equilibrate is performed until the system properties like pressure, temperature, and energy become stable with time.

**Production Run**

The MD simulation that is used to investigate the actual dynamics of the protein is called the “production run”. This is the final step of the MD simulation. The production run is performed for a length of time that is set to the expected time for a protein to exhibit an interesting change in configuration. The time evolution of the configuration of the system is called a trajectory which contains the (x, y, z) position and velocity of each atom at each time step. This trajectory can be used to analyze the behavior of the protein. Unfortunately, many structural changes such as folding occur on millisecond or longer time scales that are too long to be feasibly accessible with current computer power. For example, a microsecond of dynamics for a protein with 100 residues may take months to simulate. This limits the type of investigations that can currently be performed with MD simulations.
2.4 FORCE FIELD PARAMETERS FOR CHROMOPHORE

Equation 2.1 for the potential energy requires parameters (force constants and equilibrium bond distances and angles) for every pair of atoms. Most of these parameters are the same for a specified atomic pair in any protein, and the values of these parameters are assigned within the MD package. However, if bonding patterns of a group of atoms is not commonly found, then numerical values for the parameters are required that are specific to the special arrangement found only in one protein, or a small number of closely related proteins. The numerical values for the parameters for the chromophore for GFP-like fluorescent proteins have been determined (parameterized) by Reuter et al. (43) by using the CHARMM force field parameterization protocol. Parameters for citrine and mCherry chromophores in their neutral and anionic form were adopted from reference 34. For the acylimine part in mCherry, parameters were taken from the CHARMM27 force field.

2.5 IMPLICIT LIGAND SAMPLING (ILS)

Implicit ligand sampling (ILS) (44) is a method to find pathways for the diffusion of small gas molecules inside the protein by using MD trajectories that were performed without the gas molecules. ILS is a computational method that computes the potential of mean force (PMF) corresponding to the placement of a given small ligand such as O₂ and CO, everywhere inside the protein. The calculated PMF describes the Gibb’s free-energy cost of having a particle located at a given position, averaged over all orientations of the ligand at a given position. This free-energy can be determined for many locations in a single time frame of the trajectory, or averaged over all the frames in the trajectory.
Calculated PMF values also show the area accessible to the ligand for various values of the associated free-energy cost.

The PMF can be estimated by the relation derived in ref. (45);

\[ G(r) = -k_B T \ln \sum_{i=1}^{N} \sum_{j=1}^{C} e^{\Delta E_{i,j}(r)/k_B T} \]

Where \( G(r) \) is the ligand free-energy when placed at a position \( r \), \( C \) is the number of conformations of the ligand, \( N \) is the number of MD trajectory frames, and \( \Delta E_{i,j}(r) \) is the position dependent interaction energy of the ligand when added at position \( r \) to the given frames.

The ILS calculation method can be implemented in the VMD package (46), which uses an MD trajectory as an input file. The protein structure file (psf) should be in X-PLOR file format while using VMD to do ILS.

### 2.6 Locally Enhanced Sampling (LES)

In addition to ILS calculations to uncover oxygen diffusion pathways, a locally enhanced sampling (LES) technique (5) was employed. ILS maps out free-energy pathways by placing an oxygen molecule in various static locations in the frames of an MD simulation performed without oxygen. Since the oxygen is placed at fixed locations, the ILS maps out free-energy pathways but does not allow the oxygen to move. Therefore, it is not possible to tell if molecular gates, such as swinging sidechains, may greatly slow the diffusion to the point of being too slow to be of biological relevance. In order to determine if there are biologically relevant pathways for oxygen diffusion, it is necessary to include the oxygen molecule in the MD simulation. However, the important
pathway may not be where one might expect from examining a frame with a static structure of the protein. Therefore, it is necessary to place many oxygen molecules around the outside of the protein so that, as in a real solution, oxygen molecules can probe the protein at many different locations.

In order to include many oxygen molecules without necessitating many different, time consuming MD simulations, the LES technique is used. In this method, a single oxygen molecule is first placed in the solvated box at a location outside the protein barrel. When the MD production run commences, multiple copies of the oxygen molecule are created, and each oxygen molecule is given a different velocity so that they head in different directions. Within a few time steps, there are oxygen molecules at many locations around the protein surface, interacting with the protein and probing for pathways. Importantly, in this method, one copy of the O₂ molecule cannot see or interact with another copy, so they can occupy the same space or pass through each other. This allows a single MD simulation to include multiple oxygen molecules. The psfgen plugin implemented in VMD was used to generate a total of 15 non-interacting copies of the O₂ molecule for enhanced searching for diffusion pathways. Simulations were performed with no biasing force or potential that might facilitate oxygen entry into the interior. The NAMD package (47) was used to implement the LES technique in MD simulations.

2.7 SIMULATION CONDITIONS AND SYSTEM SETUP

CHARMM and NAMD packages were used for the MD simulations. The MMTSB toolset (41) was used for system setup for the CHARMM package, and VMD was used for the NAMD package. Initial protein structures were obtained from the
protein data bank (mCherry PDB code: 2H5Q, and citrine PDB code: 1HUY). Missing residues were added by using MODELLER (40). The all-atom CHARMM27 force field was used for simulations in both of the packages. The initial structures of mCherry and citrine were separately solvated with TIP3P water molecules with a box cut-off of 10 Å in each of the simulation setups.
3. FLUORESCENT PROTEIN BARREL FLEXIBILITY FOR VARIOUS CHROMOPHORE STATES

In this chapter, we present the results of our work in determining how the flexibility of the protein barrel of two fluorescent proteins (FPs) depends on the chemical state of the chromophore. The reason why this is important is because oxygen must have access to the chromophore in order for the chromophore to undergo the chemical changes that allow it to mature from a tripeptide to the final structure found in FPs. However, after maturation, the protein barrel must act as a shield to prevent oxygen access to the mature chromophore in order to maintain the fluorescent properties. We performed our investigations on one of the most useful monomeric variants of red fluorescent proteins (RFP) mCherry (PDB code: 2H5Q), and also on a yellow variant (YFP) of green fluorescent protein (GFP) citrine (PDB code: 1HUY). Although citrine and mCherry belong to different FP families and the photobleaching mechanisms may be different, we compare the barrel structural integrity of these proteins because of two main reasons. First, citrine is a GFP homologue of mCherry with a similar barrel structure. Second, citrine is the most useful of the FP among all YFPs due to its reduced halide sensitivity and improved photostability (48, 49). Other GFPs variants are very sensitive to halides due to easy ion access via a solvent channel or cavity formed close to the dimer interface (50, 51) from which other YFPs are derived. In citrine, this cavity is filled by the mutation Gln69Met which prevents the access to the ion (48). A similar effect is desired in mCherry to increase its photostability. We performed MD simulations using these proteins and analyzed the structural dynamics of barrels and chromophores (52).
3.1 FLUORESCENT PROTEIN STRUCTURE

Figure 3.1 Fluorescent protein structure. The chromophore and two connecting residues are shown and the central \( \alpha \)-helix is colored blue.

Green fluorescent protein (GFP) was first discovered in the early 1960s (4) and its first protein data bank crystal structures (1EMA (11) and 1GFL (53)) were obtained in 1996 by two independent groups. In general, the monomeric variants of fluorescent proteins (FPs) are composed of 11 \( \beta \)-strands which form a cylindrical barrel-like structure (Fig.3.1). Figure 3.1 shows a cartoon-ribbon representation of a fluorescent protein structure. The length and diameter of the cylindrical barrel is \( \sim 40 \) Å and \( \sim 30 \) Å, respectively. A central \( \alpha \)-helix which runs down the geometric center of the \( \beta \)-barrel holds the chemically modified tripeptide called the chromophore. Figure 3.2 shows a detailed view of the structure of a chromophore. These FPs are formed by \( \sim 220 \) to 240 amino acids. Loops and short \( \alpha \)-helices on the top and bottom of the barrel seal the ends of the barrel which protects the chromophore from quenching by restricting oxygen access from the surrounding solvent, and also restricts access of other denaturants (53).
3.2 DIFFERENT CHEMICAL STATES OF THE CHROMOPHORE AND THE CHROMOPHORE MATURATION PROCESS

In all fluorescent proteins, the chromophore is formed by chemical modification of three amino acids residues at the positions 65-67 (this numbering is based on the wild type GFP). The wild type GFP has Ser-Tyr-Gly as the tripeptide building block for the chromophore. Mutation at position 66 by an aromatic amino acid shifts the emission spectra towards cyne/blue (10). The amino acid residue at position 67 (GLY) is the only conserved amino acid in all fluorescent proteins, and different amino acids can be found at position 65 in different color variants of FPs (54).
The chemical mechanisms of chromophore formation in FPs are cyclization, dehydration, and oxidation. The process starts by nucleophilic attacks on the carboxyl carbon atom at position 65 by the amide nitrogen of glycine at position 67, followed by dehydration, and results in formation of an imidazolin-5-one heterocyclic ring system. Green fluorescence emission occurs when oxidation of the tyrosine C$_\alpha$-C$_\beta$ carbon bond by molecular oxygen extends conjugation of the imidazoline ring system to include the tyrosine phenyl ring. Red fluorescent protein chromophore formation involves a second oxidation step involving the C$_\alpha$ and amide nitrogen at position 65 that further increases the extended $\pi$-bonding electron system to include the carboxyl group of the residue at position 64 (55). A detailed schematic representation of the chromophore maturation steps in GFP and GFP-like fluorescent proteins are shown in fig 3.3 which is adapted from reference (56).
3.2.1 TRIPLETTIDE, NEUTRAL AND ANIONIC STATES OF CHROMOPHORE

The chromophores of citrine and mCherry are formed by the chemical modification of Gly-Tyr-Gly and Met-Tyr-Gly at the position 65, 66, and 67, respectively. The uncyclized three amino acid sequences are referred to as the tripeptide state. The cyclized form of the chromophore which is protonated at its tyrosine phenyl oxygen is called the neutral state, and the deprotonated tyrosine phenyl oxygen form is known as the anionic state. All three chromophore chemical states for each protein are shown in figure 3.4.

Figure 3.4 Different states of chromophore in a) citrine b) mCherry
We used molecular dynamics (MD) simulation to investigate the structural flexibility of the β-barrel and chromophore in the different maturation states of the chromophore in citrine and mCherry to determine why oxygen can enter mCherry more easily than citrine. MD simulations of 50 ns were performed and the last 40 ns of the MD trajectories are used to analyze the structural dynamics of the protein barrel and the chromophore. In addition, for the deprotonated form of the chromophore at the phenolic oxygen, the protonated Glu222 in citrine and the equivalent residue Glu215 in mCherry were used in their anionic forms.

**Methods: MD Setup**

The VMD package was used to setup the protein system for the MD simulations. The initial structures of citrine and mCherry including crystallographic water molecules were solvated by using the solvate plugin in VMD. Using a box cutoff of 10Å, the dimensions of the simulation box were 75.9Å x 73.2Å x 74.3Å and 83.3Å x 75.6Å x 63.4Å for citrine and mCherry, respectively. The solvated system was electrically neutralized by randomly adding eight Na⁺ ions for citrine and six Na⁺ ions for mCherry in the bulk water using the VMD autoionize plugin. For each system, the Particle Mesh Ewald method was used to treat long-range interactions with a 12Å nonbonded cutoff. Energy minimization was performed using the conjugate gradient and line search algorithm. The system was then heated with a linear gradient of 20 K/ps from 20 to 300 K. At 300 K, the system was equilibrated for 15 ps with a 2 fs integration time step in the NVT (constant number, volume, and temperature) ensemble. Langevin dynamics was used to maintain the temperature at 300 K. The last 40 ns of the NVT MD simulations with 2 fs time steps was used for analysis.
3.3 ROOT MEAN SQUARE FLUCTUATIONS (RMSF) OF THE PROTEIN BARREL AND CHROMOPHORE ATOMS

Figure 3.5 displays the root mean square fluctuations (RMSF) of the C\(\alpha\) atoms of the protein barrel for the three different states of the chromophore and shows that the overall flexibility of the C\(\alpha\) atoms are almost same in all three states. Loop regions show higher flexibility for the atoms as compared to other regions. In some regions of the barrel, the flexibility of the tripeptide form of the chromophore is less as compared to the neutral and anionic forms of the chromophore. This indicates that the citrine barrel is already rigid even before the chromophore matures. The RMSF analysis of mCherry shows that more flexibility of the protein barrel with the tripeptide state of the chromophore than with the neutral and anionic form of the chromophores. As with citrine, loop regions are more floppy than other parts of the barrel.

![Figure 3.5 Root mean square fluctuations of C\(\alpha\) atoms of protein barrel](image1)

Figure 3.6 shows RMSF values for the fluctuations of atoms in the imidazole ring and phenyl ring of the chromophore. The position of these atoms in the chromophore structure is shown in fig. 3.2. The RMSF values show that the atoms in the imidazole ring
Figure 3.6 Root mean square fluctuations of atoms in imidazole ring and phenyl ring of chromophore. Position of atoms in chromophore structure is shown in fig. 3.2 and phenyl ring in the chromophore are more flexible in the neutral chromophore than in the anionic chromophore in both citrine and mCherry.

The peaks in Fig 3.6 correspond to the atoms in the chromophore which can form hydrogen bonds to the immediate chromophore environment. Flexibility of atoms is observed more in the phenyl ring as compared to the imidazole ring. The phenyl ring is far from the chromophore connection point to the main chain helix. This connection point anchors the chromophore inside the protein barrel. The neutral chromophore, with the protonated phenolic oxygen, makes a weak hydrogen bond with nearby protein residues, which allows increased flexibility of this ring. The rigidity of the chromophore inside the protein barrel is related to the quantum yields because a more rigid structure increases the quantum yield by damping non-radiative modes of de-excitation. The average RMSF of all atoms in the chromophore in the anionic form is 0.46 Å in citrine and 0.50 Å in mCherry and is consistent with the experimental observed quantum yield of 0.76 in citrine (48) and 0.22 in mCherry (19).
3.4 INVESTIGATIONS OF THE GAP BETWEEN THE $\beta$7-$\beta$10 BARREL STRANDS

Figure 3.7 The $\beta$7-$\beta$10 region of mCherry (red) and citrine (yellow). a) Superposition of the ribbon structures of mCherry and citrine (52), b) Superposition of space filling structures of mCherry yellow citrine.

Dimeric and tetrameric FPs are formed by joining monomers in the $\beta$7-$\beta$10 regions of the protein barrels. Therefore, the $\beta$7-$\beta$10 region is an obvious place to look for structural weakness in the monomeric forms. Figure 3.7 compares the structures of mCherry and citrine. Figure 3.7a is a superposition of the ribbon structures of the mCherry (red) and its GFP homologue citrine (yellow). Figure 3.7b displays a space-filling model of the $\beta$7-$\beta$10 region and shows that the gap between $\beta$7 and $\beta$10 is smaller in citrine. We investigated the dynamics of this gap in citrine and mCherry using all three states of chromophore.
Figure 3.8 Gap fluctuations between $\beta_7$-$\beta_{10}$ strands in different chromophore states a) citrine b) mCherry

Figure 3.8 displays time series of the separation between strands $\beta_7$ and $\beta_{10}$. The plots represent the separation $\Delta r$ between a pair of residues, one residue on $\beta_7$ and the other on $\beta_{10}$. In order to characterize the size of the gap, an atom is chosen on each residue that is closest to the other residue across the gap. For citrine, $\Delta r$ is the distance between the $C_\alpha$ of residues Ser147 and Gln204, and for mCherry, $\Delta r$ is the distance between the $C_\alpha$ of residues Ala145 and Lys198.

In citrine, the $\beta_7$-$\beta_{10}$ gap fluctuations is almost the same in all states of chromophore. The average gap distance from the 40 ns MD simulations of tripeptide, neutral and anionic states of chromophore are 4.70 Å, 4.67Å, and 4.70Å, respectively. However, mCherry, which has a slightly bigger $\beta_7$-$\beta_{10}$ gap than citrine, shows a larger
gap opening when the chromophore is in the tripeptide state with an average value of 7.8 Å. In the anionic form of the chromophore the average gap distance of 4.67 Å in mCherry is comparable to the gap in citrine. The neutral form of the chromophore in mCherry has a 5.4 Å average gap that is slightly bigger than citrine.

These analyses show that gap fluctuations between β7-β10 is almost the same for all three different states of the chromophore in citrine, but in mCherry the gap flexibility decreases when the chromophore matures from the tripeptide to the neutral and anionic forms. The ability to form a hydrogen bond by the phenolic oxygen of the chromophore with a nearby residue in the β7 strand has some effect on gap flexibility. Hydrogen bond calculation shows that in citrine the hydrogen bond between the chromophore phenolic oxygen to His148 occurs in 38.8 % and 52.4 % of the MD frames when the chromophore is in the neutral and anionic forms, respectively. The equivalent residue Ser146 in mCherry makes a hydrogen bond with the phenolic oxygen of the chromophore only 11.6 % and 62 % of the time when the chromophores is in the neutral and anionic forms.

QM/MM energy calculations of the excited state of the chromophore in the wild-type wt-GFP shown that the His148 hydrogen bond with the chromophore phenolate oxygen is sufficient to stabilize the chromophore (57). This shows that the hydrogen bond is important for the stability of the chromophore in fluorescent proteins.

3.5 CORRELATION BETWEEN τ AND φ DIHEDRAL ANGLES

Conformational and configurational isomerization can take place when two adjacent bonds twist concerted (58). This mechanism is called Hula Twist (HT). If the direction of rotation of the τ and φ dihedral angles are in same direction then the correlation of these two angles is positive and called a positive Hula Twist, and if they
move in opposite direction then the correlation between them is negative and is called negatively Hula Twist (59). The twist around the two rings of the chromophore has been proposed as a radiationless relaxation mechanism of the chromophore from its excited state (7). An NMR study of the fluorescent protein Dronpa also suggested that the fluorescence of the protein depends on the degree of flexibility of the chromophore (60).

Figure 3.9 Chromophore dihedral angles $\tau$ (N2-CA2-CB2-CG2) and $\phi$ (CA2-CB2-CG2-CD1).

We analyzed the dynamics of two dihedral angles of the chromophore: the tau ($\tau$) dihedral angle formed by the atoms N2-CA2-CB2-CG2, and the phi ($\phi$) dihedral angle formed by the atoms CA2-CB2-CG2-CD1. These two dihedral angles are adjacent to each other as shown in fig 3.9. There is a double bond between atoms CA2 and CB2 and a single bond between CB2 and CG2 in all GFP and GFP-like fluorescent proteins.

Table 3.1 Dihedral angle (deg) parameters of citrine and mCherry

<table>
<thead>
<tr>
<th>Protein</th>
<th>$\tau_{max}$</th>
<th>$\tau_{min}$</th>
<th>$\tau_{avg}$</th>
<th>$\phi_{max}$</th>
<th>$\phi_{min}$</th>
<th>$\phi_{avg}$</th>
<th>$\tau_{range}$</th>
<th>$\phi_{range}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>citrine</td>
<td>anionic</td>
<td>44.7</td>
<td>-23.4</td>
<td>4.7</td>
<td>37.5</td>
<td>6.0</td>
<td>68.1</td>
<td>65.1</td>
</tr>
<tr>
<td></td>
<td>neutral</td>
<td>22.5</td>
<td>-27.5</td>
<td>-0.5</td>
<td>36.9</td>
<td>-36.3</td>
<td>50.0</td>
<td>73.2</td>
</tr>
<tr>
<td>mCherry</td>
<td>anionic</td>
<td>40.8</td>
<td>-26.7</td>
<td>6.2</td>
<td>36.3</td>
<td>-38.4</td>
<td>67.5</td>
<td>74.7</td>
</tr>
<tr>
<td></td>
<td>neutral</td>
<td>26.6</td>
<td>-36.0</td>
<td>0.6</td>
<td>45.8</td>
<td>-52.0</td>
<td>62.6</td>
<td>97.8</td>
</tr>
</tbody>
</table>
Table 3.1 summarizes parameters from the MD trajectories for the \( \tau \) and \( \phi \) dihedral angles in citrine and mCherry in their neutral and anionic forms of the chromophores. The average \( \tau \)-dihedral angle in both fluorescent proteins in their neutral form is around 0\(^\circ\) but the anionic forms twist this angle to \( \sim 5\)^\circ. The \( \phi \)-dihedral angle analysis shows this angle deviates from 0\(^\circ\) in all situations. This shows that the single bond \( \phi \)-dihedral angle conformations are out of planar in these fluorescent proteins. The \( \tau_{\text{range}} \) of dihedral conformations in anionic citrine is 18.1\(^\circ\) larger than in the neutral form of citrine, but the \( \phi_{\text{range}} \) is only 8.1\(^\circ\) with the neutral form having the larger range. Similarly, mCherry also has the same trend but the differences in range are 4.9\(^\circ\) and 23.1\(^\circ\). This analysis shows that the neutral form of the mCherry chromophore has more rotational freedom in its \( \phi \)-dihedral angle than all other cases. Figure 3.10 shows the probability plots of \( \tau - \phi \) dihedral angles in the neutral and anionic form of the chromophores in citrine and mCherry. These two angles are found to be negatively correlated with a correlation magnitude of 0.53 and 0.49 for anionic and neutral chromophores in citrine, and 0.49 and 0.07 in mCherry.

In conclusion, analysis shows that citrine has a more rigid protein barrel than mCherry. A series of mutations on the tetrameric form of the DsRed fluorescent protein was performed to engineer the monomeric form of mCherry. In this process, mCherry may have lost structural rigidity as compared to citrine which is also an engineered variant from GFP. Flexibility of the chromophore is found to be least in the anionic form of the chromophore as compared to the neutral form in both citrine and mCherry. The hydrogen bond between the chromophore and the barrel residue His148 in citrine and Ser146 in mCherry is found to be critical for stabilizing the chromophore by reducing the
flexibility. Dihedral angle correlations between the $\tau$-dihedral and $\phi$-dihedral angles of the chromophore also increase for the chromophore in the anionic form in both of the variants. Cis-trans isomerization of the chromophore have not observed during the MD simulations in any of the cases. The gap between the $\beta_7$-$\beta_{10}$ strands seems to be related to the ability to form hydrogen bond between the phenolate oxygen of the chromophore and His148 (in citrine) and Ser146 (in mCherry).

Figure 3.10 Probability distribution for chromophore dihedral angles $\tau$ (N2-CA2-CB2-CG2) and $\phi$ (CA2-CB2-CG2-CD1) from MD simulation.
4. MOLECULAR OXYGEN DIFFUSION IN mCHERRY AND CITRINE: IMPLICIT LIGAND SAMPLING

Recent developments in efficient computational sampling methods have allowed thorough scanning of the possible pathways for gas diffusion in the interior of proteins (45, 61-64). For example, such computational investigations have proved useful in understanding gas diffusion in may protein systems such as molecular dioxygen pathways via dynamic oxygen access channels in flavoproteins (65-67), ammonia transport in carbamoyl phosphate synthetase (68-70), and gas diffusion and channeling in hemoglobin (45, 71). We employ implicit ligand sampling (ILS) to determine oxygen pathways from the bulk solvent into the mCherry active site. Using these results as a guide, we show that the barrel fluctuations and the oxygen pathways can be altered or blocked with strategic amino acid substitutions.

Following the folding of the protein, the chromophore formation involves cyclization of tripeptide and oxidation, which requires molecular oxygen (72). Therefore, the maturation times can depend on the accessibility of molecular oxygen. For example, it is shown that a water-filled pore was essential for fast maturation of TurboGFP chromophore (73). The pore that leads from outside of the barrel to the inside chromophore possibly facilitates molecular oxygen entry. Upon comparing the crystal structures of GFPs and mFruits, structural differences in the beta barrels are observed in Figure 3.7. The tetramer subunit interactions present in the naturally occurring red fluorescent protein DsRed are not present in the mFruit monomeric forms and therefore the latter is expected to have less structural integrity. The crystal structures show larger openings in the mFruits’ protein structure, which may be transiently increased further by
more pronounced thermal fluctuations. These larger openings may allow oxygen to pass more easily to the chromophore which may have implications to both chromophore maturation speed as well as photobleaching due to oxidation.

Figure 3.7a is a superposition of ribbon structures of the RFP mCherry (PDB code 2H5Q) and its GFP homologue citrine (PDB code 1HUY). Figure 3.7b displays a space-filling model of the β7-β10 region and shows that the gap between β7 and β10 is smaller in citrine. We show that differences in this region create pathways in mCherry for dioxygen diffusion through the barrel to the chromophore.

**Methods: MD setup**

The MMTSB toolset was used to set-up the system for simulations. The initial structures of mCherry and citrine were separately solvated in octahedral boxes under periodic boundary conditions with TIP3P water molecules with a box cut-off of 10 Å. For mCherry, 11,319 water molecules were used, and for citrine, 9,290 water molecules were used. All water molecules overlapping with the protein were removed. The particle mesh Ewald method\(^{(42)}\) was used to treat long range interactions with a 9-Å non-bonded cutoff. Energy minimization was performed using the adopted basis Newton–Raphson (ABNR) method. Each system was then neutralized by adding sodium counter ions: six sodium ions for RFP and eight sodium ions for YFP. Water molecules that overlapped with the sodium ions were removed and ABNR energy minimization was performed again. The systems were then heated with a linear gradient of 50 K/ps from 50K to 300K. At 300 K, the systems were equilibrated for 2 ns with a 2 fs integration time step in the NVT (constant number, volume, and temperature) ensemble. The SHAKE algorithm was
used to constrain the bonds connected with hydrogen atoms. This was followed by a 10 ns NVT dynamics simulation with 2 fs time steps for each protein that was used for analysis.

4.1 IMPLICIT LIGAND SAMPLING (ILS) FOR MOLECULAR OXYGEN

We applied PMF/ILS calculations to the frames from our MD simulations to determine locations in a protein that are especially important for blocking, or facilitating oxygen passage, and to quantify the differences at these locations between FP variants. A total of 5,000 protein conformations from a 10-ns MD trajectory were used for ligand sampling. Therefore, the free-energy value at each of the locations is the average obtained from ILS performed every 2 ps for the 10-ns MD simulation trajectory. For the free energy calculation at each location, 20 different rotational orientations of molecular dioxygen were sampled at each grid position with a volume element size of 1 Å³. The free-energy is compared to a dioxygen molecule placed outside the protein in the surrounding water, where the free-energy is defined to be zero. In the figures, all locations with a free-energy below -2.0 kcal/mol are colored red, and all locations with a free-energy above +10.0 kcal/mol are colored blue. The values of the free energy as a function of reaction coordinate were calculated for specific positions separated by approximately 1Å distance along the pathways, extending from outside the protein in the solvent, into the protein and leading to the chromophore. The pathways were determined from visual inspection as well as from the 3D grid data of free-energy values from ILS simulations.
**Dioxygen Access Routes to the Chromophore in mCherry**

The large and fluctuating gap between strands $\beta7$ and $\beta10$ displayed in Fig. 3.8 for mCherry makes this region a prime candidate for dioxygen access in mCherry.

![Figure 4.1 Free-energy isosurfaces for molecular dioxygen in (a) citrine and (b) mCherry.](image)

The free-energy slice shown includes the $\beta7$-$\beta10$ region as well as the $\beta7$-$\beta8$ region. The color red represents low free-energy ($< -2 \text{kcal/mole}$) locations, blue represents high free-energy ($> +10 \text{kcal/mol}$) locations, and white represents locations for which the oxygen has intermediate free-energy. Neither the $\beta7$-$\beta10$ region nor the $\beta7$-$\beta8$ region in citrine offers low free-energy routes for dioxygen entry, whereas in mCherry both regions display gaps representing low free-energy access routes.
To determine which regions or pathways allow the molecular oxygen to enter the protein barrel, we calculated the free-energy of placing molecular oxygen in and around the entire protein barrel using ILS, which uses potential of mean force (PMF) calculations to determine the free-energy of placing a small molecule such as dioxygen at a specific location in a protein. Figure 4.1a displays ensemble-averaged free-energy diagrams for a dioxygen molecule if it is placed in, or around, mCherry and compare that with citrine in Fig 4.1b. We calculated the free-energy of the dioxygen using the ILS routine implemented in the VMD molecular dynamics package (46). Fig. 4.1a, display a slice that includes the β7 region. The color red represents low free-energy (< -2 kcal/mole) locations, white represents intermediate free energy locations, and blue represents high free-energy (> +10 kcal/mol) locations.

![Diagram of free-energy values of dioxygen at locations along the pathways shown in Fig. 4.1 leading from the solvent outside the protein (9 Å) into the chromophore (0 Å). The mCherry has two easy routes that can be accessed by entering through either the β7-β10 gap (R1) or the β7-β8 gap (R2). The routes for citrine through either the β7-β10 (Y1) region or β7-β8 (Y2) region are blocked by a high free-energy barrier.](image)
In order for the chromophore to have access to molecular oxygen, a pathway without substantial free-energy barriers must exist from the region outside the protein, through the protein barrel, to the chromophore location in the interior. Fig. 4.1a shows that mCherry displays two low free-energy routes through the barrel: one through the $\beta_7$-$\beta_{10}$ gap (R1) and the other through the $\beta_7$-$\beta_8$ gap (R2). These two entry routes for dioxygen lead all the way to the chromophore. In contrast, citrine has no easy pathway, including the $\beta_7$-$\beta_{10}$ region (Y1) or the $\beta_7$-$\beta_8$ region (Y2) both of which involve high free-energy (blue) barriers.

Figure 4.2 quantifies the value of the free-energy along the pathways shown in Fig. 4.1. The reaction coordinate is the position of the oxygen molecule along the route. The oxygen molecule is placed at steps along the path that are separated by 1 Å. The coordinate 0 represents a location near the chromophore, and 9 Å represents a location outside the protein in the solvent. It is seen in this figure that both routes in citrine (Y1, Y2) face large free energy barriers due to the protein barrel, whereas there is no substantial barrier for either of the pathways in mCherry (R1, R2). The identification of these pathways is used later to guide mutations of key residues in order to create barriers in mCherry to block these routes and prevent dioxygen access to the chromophore.

### 4.2 IMPORTANCE OF SIDECHAINS IN CONTROLLING GAP SIZE

The information presented in the previous sections show that the $\beta_7$-$\beta_{10}$ gap and the $\beta_7$-$\beta_8$ gap in mCherry are large enough to allow dioxygen to pass through the protein barrel to the chromophore. An aim of this work is to determine if site-specific amino acid mutations can alter these routes. In order to ascertain more details of the structural
fluctuations in the barrel, we determined if the large fluctuations in the mCherry $\beta_7-\beta_{10}$ gap is due to motion of strand $\beta_7$ or strand $\beta_{10}$, and similarly for the $\beta_7-\beta_8$ gap.

![Figure 4.3 Fluctuations in the $\beta_7-\beta_8$ gap in mCherry determined by the distance between C$_\alpha$ atoms of $\beta_7$-Ala145 and $\beta_8$-Gln163 (darker line) compared to the separation determined by the distance between C$_\alpha$ on $\beta_7$-Ala145 and the N on the sidechain of $\beta_8$-Gln163 (lighter line). The sidechain of Gln163 narrows the gap significantly more than the backbone.](image)

In comparing the time series of the fluctuations in the size of the mCherry $\beta_7-\beta_{10}$ gap and the $\beta_7-\beta_8$ gap, we found that the openings and closings of the gaps were out of phase with each other. When the $\beta_7-\beta_{10}$ gap is large, the $\beta_7-\beta_8$ is small, and vice versa, implying that the fluctuations in the $\beta_7-\beta_{10}$ gap and the $\beta_7-\beta_8$ gap are mostly due to movement of $\beta_7$. In addition, to provide more information for guiding the mutations, we
wish to determine why Figs. 4.1 and 4.2 both show that the dioxygen pathway through the β7-β8 gap (R2) is not as easy as the pathway through the β7-β10 gap (R1) even though the backbone separations are similar.

Figure 4.3 show that amino acid sidechains play an important role in closing the β7-β8 gap. Figure 4.3 displays the results of 10 ns MD simulations for the separation between stands β7 and β8 determined in two different ways. For both time series, the separation is measured between Ala145 on β7 and Gln163 on β8. One plot shows the time series of fluctuations in the separation between the amino acids’ Cα atoms. The other time series displays the fluctuating distance between the Cα of Ala145 (on β7) and the N on the sidechain of Gln163 (on β8). Figure 4.3 shows that the sidechain of Gln163 narrows the gap significantly more than the backbone.

Figure 4.4 Free-energy plot (in kcal/mol) of the β7-β8 strands. The horizontal axis is the separation between the Cα on β7- Ala145 and the N on the sidechain of β8-Gln163, the vertical axis is the separation between the Cα of β7- Ala145 and the Cα of β8-Gln163.
There are two distinct islands of low free-energy, showing that the β7 strand spends most time in these two distinct positions. Additionally, when the Cα-Cα separation is large, the sidechain undergoes larger fluctuations, which restricts the gap from opening widely.

Figure 4.4 further quantifies the importance of sidechains for determining gap sizes. In Fig. 4.4, we present a contour plot of the free energy of Ala145 on β7 and Gln163 on β8 as a function of their separation, measured in the same two ways that are used in Fig. 4.3. The vertical axis is the separation between the Cα of residue β7-Ala145 and the Cα of residue β8-Gln163 (dark line in Fig. 4.3) and the horizontal axis is the separation between the Cα of β7-Ala145 and the N on the sidechain of residue β8-Gln163 (light line in Fig. 4.3). Figure 4.4 shows that there are two islands of low free energy, which implies that β7 is stable at two distinct separations from β8, with the larger separation meaning that β7 is closer to β10. Another important feature is that when β7 is further from β8, the range of fluctuations in the sidechain of β8 are also larger. This allows the large side chain of β8 to partially close the gap even when β7 is far away.

4.3 AMINO ACID MUTATIONS MADE IN mChERRY TO CONTROL DIOXYGEN ACCESS TO THE CHROMOPHORE

Figures 4.1 and 4.2 show that the easiest pathway for oxygen access in mCherry is through the gap between β7 and β10, and Figs. 4.3 and 4.4 show that sidechains play a role in closing the β7-β8 gap. Therefore, our aim in making amino acid mutations is to decrease the β7-β10 gap without substantially increasing the size of the gap between β7 and β8.
On comparing amino acid properties of the residues in the β7, β8, and β10 strands of mCherry and citrine, it is seen that there are more charged residues in mCherry as compared to just two in β8 of citrine. The citrine residues in the region of interest are mostly polar (Ser, Tyr, Asn, His) or hydrophobic (Ala, Val, Phe, Ile, Leu).

Figure 4.5 β7-β10 gap for mCherry compared to its mutant (a) Results from the 10 ns MD simulation of the β7-β10 gap for mCherry compared to its mutant (purple). The β7-β10 gap in mut-RFP is greatly reduced compared to mCherry. (b) Free-energy isosurface for molecular dioxygen in mut-RFP. As compared to the isosurface displayed in Fig. 4.1b for mCherry, mut-RFP isosurface shows significantly less favorable pathways for oxygen entry with high free-energy barriers (blue).
We first attempted a few mutations in mCherry such that the charged amino acids are replaced by polar or hydrophobic residues, similar to the pattern in citrine. This change, however, either made the fluctuations worse or the β7-β10 gap increased even further. Since the β7-β10 gap fluctuates the most, a possible strategy to reduce this fluctuation is to create appropriate ionic interactions. In this region of mCherry, the inter-strand charged amino acid residues participate in inter-strand ionic interactions and give rise to salt-bridges. For example, the attractive interaction between Glu144(-) in β7 and 164Arg(+) in β8 swings β7 towards β8 and helps to open the gap between β7-β10. To reduce the barrel flexibility in this region, two amino acid replacements were made one in β7 and one in β8. Polar aminoacid 143Trp was replaced in β7 by a positively charged 143Lys(+), and the 164Arg(+) in β8 was replaced by 164Glu(-). The mutations introduce two possible electrostatic interactions that might close the β7-β10 gap. The attraction between the mutated β7 143Lys(+) and β10 200ASP(-) pulls β7 towards β10, and the repulsion between β7 Glu144(-) and the mutated β8 164Glu(-) pushes β7 away from β8 and towards β10. Since the location of the gap is close to the original tetramer-breaking mutations, the barrel folding in monomeric form can be sensitive to new mutations such as the one presented here. A new set of mutations must still allow the barrel to fold and chromophore to mature. If this is achieved, the marked reduction in the barrel fluctuations that limit the oxygen entry may result in a more photostable FP.

The success of the mutations in closing the β7-β10 gap in the mutated RFP (mut-RFP) is shown in Fig.4.5. The two curves display time series for mCherry and mut-RFP for the β7-β10 gap. The β7-β10 gap in Fig. 4.5 for the mut-RFP starts out at approximately 7.5Å, which is as large as it ever gets for mCherry. This is expected
because the initial positions for the atoms in our proposed mut-RFP are given by the PDB file for mCherry. Within a short time, Fig. 4.5a shows that new interactions in mut-RFP greatly reduce the β7-β10 gap compared to mCherry which significantly reduces the ease of oxygen permeability as displayed by the free-energy isosurface in Fig. 4.5b.
5. MOLECULAR OXYGEN DIFFUSION IN mCHERRY: LOCALLY ENHANCED SAMPLING

Recent investigations have shown that protein flexibility plays a major role in gas access into many proteins (23-28). Conformational flexibility of the side chains of the residues involved in forming transient cavities or pathways can alter the sizes of the bottlenecks for gas diffusion, as well as changing the gating mechanism at the protein surface (29-32, 74). In FPs, in addition to affecting the structures of both the chromophore and the protein barrel (34), the chromophore-barrel interaction can also affect the fluctuations of the barrel, which in turn can modify the spectral properties and lifetime of the fluorescence (35). It is shown in a recent important work on cyan fluorescent protein that the reduction in the flexibility of a beta strand in the barrel has led to a dramatic improvement in fluorescence quantum yield (75).

In an important work (76), Roy et al. investigated the diffusion pathways of oxygen in the phototoxic KillerRed protein. In this protein, reactive oxygen is generated from molecular oxygen that diffuses into the interior of the protein. They were able to identify the pores and channels for the oxygen to escape through the protein barrel to the bulk solvent. This study also suggested that the ease of molecular oxygen diffusion through a channel is the cause of the high susceptibility for photobleaching (76). In our work (ref. (52)), oxygen diffusion pathways in mCherry were investigated by implicit ligand sampling techniques which we explained in earlier chapter 4. In that study, an immature tripeptide form of the chromophore was used, and crystallographic water molecules were not included in order to quicken barrel fluctuations so that they could be observed in shorter simulation time scales. To better understand the diffusion process in
a more realistic setting, we performed molecular dynamics (MD) simulations with explicit molecular oxygen in the system. We use force field parameters for a mature chromophore and also include the crystallographic water in the simulations. The results of these computations describe a pathway that allows oxygen molecules to enter from the solvent and travel through the protein. The pathway contains several oxygen hosting pockets, which are identified by the amino acid residues that form the pocket. We calculate the free-energy of an oxygen molecule at any point along the path. The results provide a better understanding of the mechanism of molecular oxygen access into the fully folded mCherry protein barrel and provide insight into the photobleaching process in these proteins.

Methods: MD setup

The VMD package was used to setup the system for simulations. The initial structure of mCherry with crystallographic water molecules and one molecular oxygen was solvated by using the solvate plugin in VMD. Using a box cutoff of 10Å, the dimensions of the simulation box were 83.3Å x 75.6Å x 63.4Å. The solvated system was electrically neutralized by adding six Na⁺ ions randomly in the bulk water using the VMD autoionize plugin. The final system contained a total of 37,276 atoms. All water molecules overlapping with the protein were removed. The particle mesh Ewald method was used to treat long-range interactions with a 12Å nonbonded cutoff. Energy minimization was performed using the conjugate gradient and line search algorithm. The system was then heated with a linear gradient of 20 K/ps from 20 to 300 K. At 300 K, the system was equilibrated for 15 ps with a 2 fs integration time step in the NVT (constant number, volume, and temperature) ensemble. Langevin dynamics was used to maintain
the temperature at 300 K. An 80 ns NVT dynamics simulation with 2 fs time steps was used for analysis.

**Structural Features of mCherry**

The first monomeric variant of the red fluorescent protein, mRFP1 was derived from the Discosoma sp. fluorescent protein DsRed (19, 77). Development of the monomeric RFP overcame the problems of tetramerization and slow maturation of the parent protein DsRed. However, mRFP1 suffered from lower quantum yield and lower photostability, possibly due to the compromised barrel structure caused by the mutations introduced to break the tetramer interactions at the interface. The monomeric variant mCherry is one of the next-generation monomeric RFPs derived from mRFP1 and has significantly improved photophysical properties (19). Among the mutations introduced to obtain mCherry, Q66M enhanced the maturation, V7I and M182K enhanced the folding, and M163Q removed an unwanted absorbance peak, in addition to significantly enhancing the mCherry photostability (22). As discussed later, our simulation results show that the M163Q mutation in mCherry significantly reduces molecular oxygen entry into the barrel, which might help explain the role of molecular oxygen in permanent photobleaching of FPs and improving the photostability in mCherry.

As with other mFruit variants, the barrel structural integrity is compromised, especially in the β7 and β10 region, due to missing tetrameric interactions present in the naturally occurring DsRed. Transient thermal fluctuations (52) can allow easier oxygen access to the chromophore. This may help chromophore maturation but can cause fluorescence quenching or faster photobleaching due to oxidation. In cyan fluorescent protein, β7 flexibility has been attributed to cause collisional fluorescent quenching due
to the collision of the Ile146 side chain with the chromophore. In a recent work, structure guided amino acid replacements to reduce β7 flexibility have led to a significantly brighter and highly photostable fluorescent protein mTurquoise2, with the highest quantum yield (93%) among monomeric fluorescent proteins (75).

In ref. (52), the barrel structure of mCherry is compared with that of citrine. Implicit oxygen ligand sampling showed that the β7- β10 gap in mCherry provides an easy path for oxygen entry. In addition, static structural comparison also shows differences at the top and the bottom of the barrels. For example, the top of the citrine barrel contains an extra α-helix which is not present in mCherry. In this region of mCherry, there is a random coil (or loop segment), and therefore, larger structural fluctuation can be expected. In the earlier study with implicit ligand sampling, no clear entry path for oxygen was observed from the top of the barrel. However, this does not conclusively prove that an oxygen pathway does not exist because implicit ligand sampling may not capture a dynamic pathway that opens and closes but with gatekeeping amino acids that open for a very short time. Moreover, the presence of an actual oxygen molecule can slightly modify the environment in a way that might allow the oxygen to enter the protein barrel. The explicit oxygen simulations carried out in the present investigations are able to capture these possibilities.

5.1 MOLECULAR OXYGEN DIFFUSION PATHWAYS IN mCHERRY

For enhanced search statistics, our explicit oxygen calculation employed 15 copies of oxygen in our NAMD LES calculations. The oxygen molecules do not interact with each other, but interact with the rest of the system. With these 15 copies of noninteracting oxygen molecules in the system, an 80-ns production run was performed.
The simulations reveal several different types of events of molecular oxygen entry into the protein barrel from various locations, which are pictured in Figure 5.1 and described below. Several protein pockets were observed far from the chromophore where the molecular oxygen can enter and remain for an extended time. Some of these pockets are dead-ends with no access to the chromophore, but some of the pockets join to pathways that ultimately lead to the chromophore (78).

Figure 5.1 Various oxygen-hosting pockets in the mCherry protein barrel. (a-d) Pockets A, A’, B, and C are part of the same oxygen diffusion channel. Pocket C is in the vicinity of the chromophore. Other pockets exist (X and D) but do not connect to the channel. (e) Pocket X close to the middle of the barrel but off to the side (f) pocket D near the top of the barrel.
Diffusion of Oxygen into Pockets Far from the Chromophore

In our simulations, the first oxygen entry occurred at the bottom of the barrel at approximately 4 ns into the production run. This oxygen-hosting pocket is formed by the side chains of PHE56, ILE60, PHE129, and MET136. This pocket is labeled as pocket A and displays it in Figure 5.1a. Oxygen enters this pocket through residues HIS25, PRO55, and MET136, which act as gateway residues for multiple entries and exits of the oxygen molecules. It is found that the oxygen can also enter pocket A through residues PRO134 and TYR173. Entry through this gate took 5 ns to reach pocket A.

We focus our attention on the gate composed of residues HIS25, PRO55, and MET136. The first oxygen that enters at 4 ns escapes back to the solvent after only 200 ps in the pocket. However, another O₂ molecule enters again through this gate at 10 ns. This time, the oxygen stays in the pocket for approximately 1 ns and then moves further into the protein to a second pocket (pocket B) formed by residues GLN64, ARG95, MET97, and VAL105 as shown in Fig.5.1c. Multiple back and forth transitions of the oxygen between pockets A and B were observed. These transitions are shown in Figure 5.2, which displays the distance of the oxygen molecule from a reference point on the chromophore. We chose nitrogen N3 of the imidazoline ring as the reference point because this part of the chromophore is least flexible. The oscillations of the oxygen (red) between pockets A and B occurs rapidly and shows no intermediate pocket. During further simulations, we observed a different oxygen molecule enter pocket A and also make back and forth transitions between pockets A and B, but via an intermediate pocket (pocket A’ of Figure 1b) formed by residues LEU61, MET97, LEU124, T127, and GLY126. The MET97 side chain forms a part of both pockets A’ and B, so the movement
of this side chain allows the transition between these pockets. Figure 5.2 shows a second trajectory (green) for the oxygen movement that occurs via the intermediary pocket (pocket A’). Ultimately, this oxygen molecule makes the transition to pocket C (Figure 5.1d) and then escapes the protein barrel (Supplementary Movie S1 in ref. 78) as described later. In the 7 ns time window shown in Figure 5.2, only the second trajectory (green) includes an oxygen transition to pocket C, whereas the first trajectory (red) shows oscillations between pockets A and B only. Supplementary Movie S1 shows approximately the last 4 ns of the green trajectory.

![Figure 5.2 Trajectories of molecular oxygen showing its distance from the chromophore for a 7-ns window. One trajectory (red) involves oscillations between two pockets in the protein (pockets A and B of Figure 5.1). The other trajectory (green) involves oscillation between the same two pockets, but also includes oxygen movement in an intermediate pocket, pocket A’ of Figure 5.1b, and to pocket C close to the chromophore. The starting point Δt=0 corresponds to 52 ns into the simulation for the green trajectory and 66 ns for the red trajectory.](image)

55
An interesting pocket (pocket X of Figure 5.1e) observed during the simulation is formed by residues LEU46, VAL48, TYR208, THR209, and VAL211, and lies in the barrel interior but off to a side, close to the $\beta 3$ and $\beta 11$ strands of the barrel. This mostly hydrophobic pocket hosts the oxygen for a very long time. Oxygen is observed to enter this pocket at 34 ns and remain there for the rest of the 80-ns simulation. Although the oxygen in this pocket is far from the chromophore, 18Å from the phenolate oxygen, and is not part of any oxygen diffusion channel, we cannot rule out the possibility that an oxygen in this pocket can affect the chromophore, especially since it stays in the pocket for tens of nanoseconds. Another pocket (pocket D, Figure 5.1f) was observed near the top of the barrel and is formed by residues ILE8, PRO37, ALA71, TYR72 and PHE118, with the side chain of TYR72 blocking the oxygen from diffusing further into the barrel. All of these residues are conserved in the mFruits. An oxygen that enters pocket D escapes back to the solvent in less than a nanosecond via the same route that it entered. This pocket may be important because oxygen enters and leaves this pocket multiple times.

We also observed a transient cavity at the barrel surface that is formed by the side chains of LEU54, PHE56, TRP58, and LEU61. These amino acids belong to the central $\alpha$-helix inside the barrel. The oxygen enters this cavity at 9 ns and escapes back to the solvent at 10 ns. No pathways to the chromophore or to other pockets are found, so an oxygen molecule that enters this cavity will quickly escape to the solvent. The side chain of LEU61 separates the chromophore from this cavity and therefore this residue plays a crucial role in preventing the oxygen from reaching the chromophore. Comparison of amino acid sequences shows that these four amino acids are conserved throughout the
mFruit family, as well as the parent DsRed. Except for LEU54, the other three amino acid residues are also conserved in GFP.

**Diffusion of Oxygen into Pockets in the Vicinity of the Chromophore**

The $\beta_7$-$\beta_{10}$ region of the barrel is important for oxygen entry because the $\beta_7$-$\beta_{10}$ gap undergoes thermal fluctuations that repeatedly produce a large opening. The amino acid residues near this gap, TRP143 of $\beta_7$, GLN163 of $\beta_8$, and LEU199 of $\beta_{10}$, form a small gateway pocket just inside the surface. This gateway pocket is important because it provides access both to the chromophore and to other pockets further inside the protein, but it is transient, depending on the fluctuations of $\beta_7$ and $\beta_{10}$. Later in this article, we focus our attention on residue 163 because the side chain of GLN163 is found to be responsible for hindering the diffusion of molecular oxygen further into the barrel. This provides a possible molecular basis for the improved photostability in mCherry, as compared to its predecessor variant, which has MET163 at that position.

In order to investigate the $\beta_7$-$\beta_{10}$ gateway pocket in more detail, we cut-out uninteresting computational time during which the oxygen molecule moved around in the solvent outside the protein. The initial entry of an oxygen molecule from the solvent into this gateway pocket required 71 ns. In order to avoid the computational wait-time for the oxygen in the solvent to get to the gate point between $\beta_7$-$\beta_{10}$, 20 new simulations were run. Instead of LES with 15 different oxygen molecules, we used 20 independent simulations, each with just one oxygen molecule placed at the $\beta_7$-$\beta_{10}$ gate. This expedited the search for pathways into the protein through the $\beta_7$-$\beta_{10}$ gate. A simulation was terminated if the oxygen molecule escaped the protein and went out to the solvent. We considered the oxygen molecule to have completely escaped and terminated the
simulation run if the oxygen’s distance from the chromophore’s phenolate oxygen exceeded 15Å. This distance ensures that a simulation will not be terminated if the oxygen has not truly escaped the \( \beta7-\beta10 \) gate and an oxygen that remains in the vicinity just outside the pocket will be given time to reenter. In nine runs, the oxygen molecule escaped immediately (within picoseconds). In the other 11 of these 20 simulations, the oxygen molecule entered completely into the gateway pocket. For 10 of these 11 simulations, the oxygen molecule remained in the pocket for 13 ns on average, before escaping back into the solvent. Only in one simulation did the oxygen molecule manage to diffuse further into another pocket, pocket C shown in Figure 5.1d and Supplementary Movie S2 (ref. 78).

Pocket C is formed by several amino acids including ARG95, MET97, TRP143, GLN163, and VAL177. This pocket is especially important because it is very close to the chromophore (just below the chromophore). As we show later, a mutation of GLN163 allows easier entry for the molecular oxygen from the gateway pocket to pocket C.

**Oxygen Diffusion Channels Connecting Multiple Pockets**

Analysis of the trajectories of individual oxygen molecules showed two vulnerable regions of the protein through which molecular oxygen can enter and ultimately reach pocket C, which is directly underneath the chromophore. As described above, one of these regions on the barrel close to pocket C is a gap on the surface between strands \( \beta7 \) and \( \beta10 \). The other entry region is near pocket A and further away from the chromophore. Residues HIS25, PRO55, MET136, PRO134 and TYR173 act as the gateway residues in this region (near pocket A).
Figure 5.3 (a) Volumetric trace of the oxygen diffusion channel from one side of the protein to the other. (b) Amino acid residues involved in the oxygen diffusion channel. Common amino acid residues shared by different pockets indicate that the side chains of these residues separate the pockets. In addition to the amino acids involved in defining the different pockets, the gateway residues are also shown in light colored boxes just outside pocket A and pocket C.

A complete passage of an oxygen molecule that enters from one side of the protein and leaves through the other side (solvent → pocket A ↔ B → pocket C → β7-β10 gateway pocket → solvent) was observed from 51 ns to 63 ns of simulation. During this time, we also observed the diffusion of oxygen between pocket A ↔ pocket A’. The reverse pathway was observed at 71 ns in which an oxygen molecule entered through the gap between β7-β10 and traveled a path from solvent → β7-β10 gateway pocket → pocket C → pocket B → pocket A → solvent. The oxygen molecule that entered the barrel at 71 ns through β7-β10 ultimately escaped from the bottom of the barrel at 73 ns.
This pathway is a complete channel from the β7-β10 gateway near pocket C into the barrel interior and then exiting near pocket A. Figure 5.3 summarizes the pathways and the residues involved in forming the channel for oxygen diffusion from one side of the protein to the other.

5.2 FREE-ENERGIES ALONG THE PATHWAY CALCULATED FROM IMPLICIT LIGAND SAMPLING

![Free-energy values for oxygen molecule](image)

Figure 5.4 Free-energy values for the oxygen molecule at locations along the curved pathway (channel) of Figure 5.3 connecting the solvent and the protein interior. The reaction coordinate is the distance of the oxygen molecule with respect to the chromophore.

Implicit ligand sampling was performed to calculate the free-energy of an oxygen molecule at positions along the channel displayed in Figure 5.3. Figure 5.4 displays the free-energy along a reaction coordinate that represents the oxygen diffusion pathway (channel) which allows an oxygen molecule in the solvent to enter the protein and...
approach the chromophore. The free-energy curve in Figure 5.4 is based upon reaction coordinate points that are at 1 Å separations along the channel. There are clear free-energy minima for pockets A, B, and C. The barrier between pocket A and pocket B is relatively small, which indicates that it is relatively easy for the oxygen to move back and forth between these pockets. This is consistent with the short oscillation times displayed in the red curve of Figure 5.2. In contrast, the barrier between pocket B and C is much higher. The high free-energy barrier makes it more difficult for the oxygen molecule to travel between these pockets. During the 80 ns LES simulation, only a total of three $B \leftrightarrow C$ barrier crossing events: two from pocket B to pocket C and one from pocket C to pocket B were observed.

5.3 M163Q MUTATION AND THE ENHANCED PHOTOSTABILITY OF mCHERRY

During the directed evolution of mCherry from mRFP1, the mutation M163Q was experimentally determined to be solely responsible for enhanced photostability in mCherry (22). In order to understand the role of molecular oxygen in fluorescent protein photostability, we performed simulations with the reverse mutation Q163M in mCherry and investigated the oxygen diffusion pathways in the mutant mCherry-Q163M. We performed both LES as well as a number of independent simulations. A 30-ns LES simulation with 15 copies of molecular oxygen was performed. As with the mCherry simulations, the 15 copies were placed just outside the protein barrel. Within the 30-ns simulation, a total of five molecular oxygen entries were observed into the barrel, three through the gateway pocket near the β7-β10 gap, and two from the bottom of the barrel. (In contrast, there was no molecular oxygen entry into the mCherry barrel through the
gateway pocket near the β7-β10 gap until 71 ns.) The first close approach of molecular oxygen to the chromophore in mCherry-Q163M occurs fairly quickly, at 11 ns, and this oxygen enters through the gateway pocket near the β7-β10 gap. The distance of closest approach between the phenolic oxygen in the chromophore and the molecular oxygen was 2.5 Å. The oxygen molecule then quickly transitions further into pocket C.

In two other oxygen entries (at 14 ns and 26 ns) through this gateway pocket in mCherry-Q163M, the molecular oxygen continued diffusing further into the protein but through a new route, not observed in mCherry. Rather than diffusing into pocket C, these oxygen molecules slide through a barrel-chromophore interface and ultimately reach the back of the chromophore, making contact with the MET part of the chromophore. These oxygen molecules remain in this region for the rest of the 30-ns simulations. Molecular oxygen entries from the bottom of the protein barrel were observed at 17 ns and 23 ns following a similar route as in mCherry (i.e. pocket A ↔ pocket B ↔ pocket C). Thus, the LES simulations show that the same oxygen diffusion channels in mCherry were also observed in mCherry-Q163M. However, the rate of entry as well as the number of oxygen molecules was found to be significantly higher in mCherry-Q163M as compared to mCherry, implying a correlation between reduced oxygen permeability and the improved photostability in mCherry.

In order to further investigate differences in the mechanism of oxygen diffusion, as was done with mCherry, for mCherry-Q163M 20 independent simulations were performed in which an oxygen molecule was placed just outside the protein barrel but near the chromophore (close to the gateway pocket). The initial location of the molecular oxygen is comparable to the location of the oxygen used in the mCherry simulations: the
distance between the chromophore phenolic oxygen and the molecular oxygen was 2.69 Å in mCherry versus 2.70 Å in mCherry-Q163M. In the 20 runs with mCherry-Q163M, molecular oxygen was observed to enter the gateway pocket in 14 runs. In 5 of these runs, molecular oxygen diffused further into pocket C. In contrast, simulations in mCherry showed the diffusion of molecular oxygen into pocket C in only one run. Also, in mCherry-Q163M, we observed molecular oxygen making back and forth transitions between the gateway pocket and pocket C, but did not observe this in mCherry.

Figure 5.5 Distribution and schematic of side chain dihedral angles for the gateway residue GLN163 in mCherry versus MET163 in mCherry-Q163M. The MET163 is much more flexible and is more likely to allow diffusion of molecular oxygen into the protein compared to the more rigid GLN163.
In order to understand the structural basis for the differences in the oxygen diffusion between gateway pocket ↔ pocket C, we inspected the dynamics of the amino acid side chain involved in the interface between these pockets. The crucial side chain at the interface is residue 163, which is GLN163 in mCherry versus MET163 in mCherry-Q163M. Although the sizes of the GLN and MET side chains are quite similar, we found their dynamics to be different. In Figure 5.5 we plot the distribution of the dihedral angles (C-C-C-N and C-C-S-C) of GLN163 in mCherry and MET163 in mCherry-Q163M. The distributions of dihedral angles are obtained from a 25-ns window within their respective LES simulations. The dihedral angle distributions clearly show that the GLN163 side chain in mCherry is more rigid than the MET163 side chain in mCherry-Q163M.

The side chain of MET163 can flip between two structures with very different peak dihedral angles of -72° and +72°. In addition, there is a probability for the MET163 side chain to assume a dihedral angle of 180°, which is not observed for GLN163. Although a combination of chromophore positions as well as the position of nearby atoms plays a role in the gateway pocket ↔ pocket C oxygen transitions, a number of such oxygen transitions was observed when the dihedral angle was close to 180°. We also observed that the mutation at 163 causes the rearrangement of nearby residues, such as a shift in the dihedral angle distribution of MET97, which might alter other oxygen diffusion pathways.
6. CONCLUSIONS

Access for molecular oxygen to get inside the protein barrel is required for chromophore maturation in fluorescent proteins. However, oxygen access can also cause irreversible photobleaching and significantly reduce the photostability of an FP. In this work, we performed molecular dynamics simulations to investigate the diffusion of molecular oxygen into the protein barrel of the monomeric RFP variant mCherry. A clear channel for oxygen diffusion into the protein was described, and the free-energy of an oxygen molecule at any point along the path has been calculated. The pathway contains several oxygen hosting pockets, which are identified by the amino acid residues that form each pocket. One end of the channel is accessed from the solvent through the floppy $\beta_7$-$\beta_{10}$ gap, which leads immediately to a gateway pocket that provides some access to the chromophore. Diffusion of molecular oxygen deeper into the protein, providing better access to the chromophore, is hindered by GLN163, but our calculations show that it becomes easier upon the mutation Q163M, which is consistent with experimental observations of significantly lower photostability for mCherry-Q163M as compared to that of mCherry. Another entrance from the solvent is at the bottom of the protein barrel and leads through other pockets to join the same pocket next to the chromophore. The oxygen access to the chromophore through regions close to $\beta_7$ not only leads to collisional quenching but also affects the protein’s fluorescence lifetime. Such computational identifications of oxygen diffusion pathways can be helpful in guiding mutagenesis efforts to design fluorescent proteins with improved photophysical properties. We have also shown that specific point mutations can alter the oxygen pathways in the RFPs.
Blocking or altering these pathways through the barrel can have an effect on FP maturation as well as on its photostability. For example, easy oxygen access may significantly reduce the photostability whereas it may be useful for chromophore maturation, especially at low oxygen conditions. The computational approach can provide important insights for guiding efficient mutagenesis experiments to improve the maturation speed and photostability of mFruits.
REFERENCES


19. Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from


backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. Journal of chemical
23341755; PubMed Central PMCID: PMC3549273.

40. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY,
et al. Comparative protein structure modeling using Modeller. Current protocols in
bioinformatics / editorial board, Andreas D Baxevanis [et al]. 2006;Chapter 5:Unit 5 6.

41. Feig M, Karanicolas J, Brooks CL, 3rd. MMTSB Tool Set: enhanced sampling
and multiscale modeling methods for applications in structural biology. Journal of
PubMed PMID: 15099834.


the Neutral and Deprotonated Forms of the Chromophore. Molecular Dynamics
Simulations of the Wild Type and S65T Mutant. The Journal of Physical Chemistry B.

44. Cohen J, Olsen KW, Schulten K. Finding gas migration pathways in proteins
using implicit ligand sampling. Methods in enzymology. 2008;437:439-57. doi:

O2, CO, NO, and Xe inside myoglobin. Biophys J. 2006;91(5):1844-57. Epub
2006/06/06. doi: S0006-3495(06)71896-3 [pii] 10.1529/biophysj.106.085746. PubMed
PMID: 16751246; PubMed Central PMCID: PMCPMC1544290.

46. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. Journal of

molecular dynamics with NAMD. Journal of computational chemistry.
16222654; PubMed Central PMCID: PMCPMC2486339.

48. Griesbeck O, Baird GS, Campbell RE, Zacharias DA, Tsien RY. Reducing the
environmental sensitivity of yellow fluorescent protein. Mechanism and applications.
PubMed PMID: 11387331.


APPENDICES

Appendix A: Chromophore Structure of citrine and mCherry (Atoms name are according to pdb file)
## Appendix B: Residue topology and parameter files for the chromophore of mCherry

### Table 5.1 Residue topology file for the chromophore of mCherry (anionic)

<table>
<thead>
<tr>
<th>Mass</th>
<th>Symbol</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>197</td>
<td>NRc2</td>
<td>14.00700</td>
<td>N ! neutral his unprotonated ring nitrogen</td>
</tr>
<tr>
<td>198</td>
<td>NRc1</td>
<td>14.00700</td>
<td>N ! neutral his protonated ring nitrogen</td>
</tr>
<tr>
<td>199</td>
<td>HAc1</td>
<td>1.00800</td>
<td>H ! for alkene; RHC=CR</td>
</tr>
<tr>
<td>200</td>
<td>HPc</td>
<td>1.00800</td>
<td>aromatic H</td>
</tr>
<tr>
<td>201</td>
<td>Oc2</td>
<td>15.99900</td>
<td>O ! carbonyl oxygen</td>
</tr>
<tr>
<td>202</td>
<td>OcH</td>
<td>15.99900</td>
<td>O ! from OH1</td>
</tr>
<tr>
<td>203</td>
<td>HcH</td>
<td>1.00800</td>
<td>polar H</td>
</tr>
<tr>
<td>204</td>
<td>HAc</td>
<td>1.00800</td>
<td>nonpolar H</td>
</tr>
<tr>
<td>205</td>
<td>CA1</td>
<td>12.01100</td>
<td>C ! aromatic C</td>
</tr>
<tr>
<td>206</td>
<td>CA2</td>
<td>12.01100</td>
<td>C ! aromatic C</td>
</tr>
<tr>
<td>207</td>
<td>CA3</td>
<td>12.01100</td>
<td>C ! aromatic C</td>
</tr>
<tr>
<td>208</td>
<td>CPc2</td>
<td>12.01100</td>
<td>his CE1 carbon</td>
</tr>
<tr>
<td>209</td>
<td>CEc1</td>
<td>12.01100</td>
<td>for alkene; RHC=CR</td>
</tr>
<tr>
<td>210</td>
<td>CPc1</td>
<td>12.01100</td>
<td>his CG and CD2 carbons</td>
</tr>
<tr>
<td>211</td>
<td>CA4</td>
<td>12.01100</td>
<td>aromatic C</td>
</tr>
<tr>
<td>212</td>
<td>CT3c</td>
<td>12.01100</td>
<td>aliphatic sp3 C for CH3</td>
</tr>
<tr>
<td>213</td>
<td>CTc1</td>
<td>12.01100</td>
<td>aliphatic sp3 C for CH</td>
</tr>
</tbody>
</table>

**Declaration and Auto Definitions**

**Declaration**
- CA
- C
- O
- CA
- C
- N1
- N
- CA

**Auto Definitions**
- FIRS NTER LAST CTER
- ANGLES DIHEDRALS

**Residue**
- CH6 -1.000

**Group**
- Imidazolinone ring

<table>
<thead>
<tr>
<th>Atom</th>
<th>Symbol</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>CPC2</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>N2</td>
<td>NRc2</td>
<td>-0.60</td>
<td></td>
</tr>
<tr>
<td>N3</td>
<td>NRc1</td>
<td>-0.57</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>CPC1</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>O2</td>
<td>Oc2</td>
<td>-0.57</td>
<td></td>
</tr>
<tr>
<td>CA2</td>
<td>CPC1</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>CB2</td>
<td>CEc1</td>
<td>-0.14</td>
<td></td>
</tr>
<tr>
<td>HB2</td>
<td>HAc1</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>CG2</td>
<td>CA1</td>
<td>-0.09</td>
<td>Tyr ring : charges from charmm22</td>
</tr>
<tr>
<td>CD1</td>
<td>CA2</td>
<td>-0.08</td>
<td></td>
</tr>
<tr>
<td>HD1</td>
<td>HPc</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>CD2</td>
<td>CA2</td>
<td>-0.08</td>
<td></td>
</tr>
<tr>
<td>HD2</td>
<td>HPc</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>CE1</td>
<td>CA3</td>
<td>-0.28</td>
<td></td>
</tr>
<tr>
<td>HE1</td>
<td>HPc</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>CE2</td>
<td>CA3</td>
<td>-0.28</td>
<td></td>
</tr>
<tr>
<td>HE2</td>
<td>HPc</td>
<td>0.10</td>
<td></td>
</tr>
</tbody>
</table>
ATOM CZ CA4 0.45
ATOM OH OcH -0.62

!Glycine part from Charmm22
GROUP
ATOM CA3 CT2 -0.18 !|
ATOM HA31 HB 0.09 !|
ATOM HA32 HB 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 CTc1 0.16 !atom type changed
GROUP
ATOM CB1 CT2 -0.18
ATOM HB11 HA 0.09
ATOM HB12 HA 0.09
GROUP
ATOM CG1 CT2 -0.14
ATOM HG11 HA 0.09
ATOM HG12 HA 0.09
ATOM SD S -0.09
ATOM CE CT3 -0.22
ATOM HE11 HA 0.09
ATOM HE12 HA 0.09
ATOM HE13 HA 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 C2 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

Table 5.2 Parameter file for the chromophore of mCherry (anionic)

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

! GFP Chromophore parameters, deprotonated form
!
BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
<table>
<thead>
<tr>
<th>atom type</th>
<th>Kb</th>
<th>b0</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPc2 CTc1</td>
<td>354.000</td>
<td>1.4900</td>
</tr>
<tr>
<td>NRc1 CT2</td>
<td>396.000</td>
<td>1.4400</td>
</tr>
<tr>
<td>NRc1 CPC2</td>
<td>400.000</td>
<td>1.3900</td>
</tr>
<tr>
<td>NRc1 CPC1</td>
<td>400.000</td>
<td>1.4100</td>
</tr>
<tr>
<td>CPC1 Oc2</td>
<td>854.000</td>
<td>1.2400</td>
</tr>
<tr>
<td>NRc2 CPC2</td>
<td>400.000</td>
<td>1.3000</td>
</tr>
<tr>
<td>CPC1 CPC1</td>
<td>410.000</td>
<td>1.4600</td>
</tr>
<tr>
<td>NRc2 CPC1</td>
<td>400.000</td>
<td>1.4000</td>
</tr>
<tr>
<td>CPC1 CEc1</td>
<td>500.000</td>
<td>1.3900</td>
</tr>
<tr>
<td>HAc1 CEc1</td>
<td>360.500</td>
<td>1.1000</td>
</tr>
<tr>
<td>CEc1 CA1</td>
<td>437.000</td>
<td>1.4100</td>
</tr>
<tr>
<td>CA1 CA2</td>
<td>305.000</td>
<td>1.4300</td>
</tr>
<tr>
<td>HPc CA2</td>
<td>340.000</td>
<td>1.0800</td>
</tr>
<tr>
<td>CA2 CA3</td>
<td>305.000</td>
<td>1.3500</td>
</tr>
<tr>
<td>HPc CA3</td>
<td>340.000</td>
<td>1.0800</td>
</tr>
<tr>
<td>CA3 CA4</td>
<td>305.000</td>
<td>1.4550</td>
</tr>
<tr>
<td>OcH CA4</td>
<td>842.000</td>
<td>1.2500</td>
</tr>
<tr>
<td>CTc1 NH1</td>
<td>463.000</td>
<td>1.3650</td>
</tr>
<tr>
<td>CT2 CTc1</td>
<td>222.500</td>
<td>1.5380</td>
</tr>
</tbody>
</table>

ANGLES

!V(angle) = Ktheta(Theta - Theta0)**2

!V(Urey-Bradley) = Kub(S - S0)**2

!Ktheta: kcal/mole/rad**2

!Theta0: degrees

!Kub: kcal/mole/A**2 (Urey-Bradley)

!S0: A

<table>
<thead>
<tr>
<th>atom types</th>
<th>Ktheta</th>
<th>Theta0</th>
<th>Kub</th>
<th>S0</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRc2 CPC2</td>
<td>130.000</td>
<td>113.3000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC2 NRc1</td>
<td>130.000</td>
<td>106.6000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC2 NRc1 CPC1</td>
<td>130.000</td>
<td>107.9000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>NRc2 CPC1 CPC1</td>
<td>130.000</td>
<td>108.3000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>NRc2 CPC1 CEc1</td>
<td>45.800</td>
<td>129.5000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>NRc1 CPC1 Oc2</td>
<td>50.000</td>
<td>124.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>NRc1 CPC1 CPC1</td>
<td>130.000</td>
<td>103.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>Oc2 CPC1 CPC1</td>
<td>44.000</td>
<td>133.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC1 CPC1 CEc1</td>
<td>45.800</td>
<td>122.7000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC1 CEc1 CA1</td>
<td>130.000</td>
<td>133.2000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC1 CEc1 HAc1</td>
<td>42.000</td>
<td>112.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CEc1 CA1 CA2</td>
<td>45.800</td>
<td>120.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>HAc1 CEc1 CA1</td>
<td>42.000</td>
<td>115.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CA1 CA2 CA3</td>
<td>40.000</td>
<td>122.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CA2 CA1 CA2</td>
<td>40.000</td>
<td>116.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CA2 CA3 CA4</td>
<td>40.000</td>
<td>122.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CA3 CA4 CA3</td>
<td>40.000</td>
<td>115.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>HPc CA3 CA4</td>
<td>30.000</td>
<td>120.0000</td>
<td>!</td>
<td></td>
</tr>
</tbody>
</table>
HPC CA3 CA2  30.000  120.0000 !
HPC CA2 CA3  30.000  120.0000 !
HPC CA2 CA1  30.000  120.0000 !

OcH CA4 CA3  45.200  120.0000 ! ALLOW ARO ALC !

!Link to the met(65) fragment

NH1 CTc1 CPC2  50.000  107.0000 ! ALLOW PEP POL ARO ALI
NRc2 CPC2 CTc1  40.000  125.0000 !
NRc1 CPC2 CTc1  40.000  121.7000 !ion
CT2 CT1 CPC2  52.000  108.0000 ! ALLOW ALI PEP POL ARO
CT2 CTc1 CPC2  52.000  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
HA CT2 CTc1  33.430  110.1000  22.53  2.17900
CT2 CT2 CTc1  58.350  113.50  11.16  2.56100

!Link to the gly(67) fragment

CPc2 NRc2 CT2  40.000  128.3000 !ion
CPc1 NRc1 CT2  40.000  123.8000 !ion
NRc1 CT2 C  50.000  107.0000
NRc1 CT2 HB  48.000  108.0000

! DIHEDRALS

!V(dihedral) = Kchi(1 + cos(n(chi) - delta))

!Kchi: kcal/mole
!n: multiplicity
!delta: degrees

!atom types

<table>
<thead>
<tr>
<th></th>
<th>Kchi</th>
<th>n</th>
<th>delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPc2 NRc2 CPC1 CPc1</td>
<td>14.0000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>CPc2 NRc1 CPC1 CPc1</td>
<td>14.0000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>NRc2 CPC2 NRc1 CPc1</td>
<td>14.0000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>NRc2 CPC1 CPC1 NRc1</td>
<td>4.0000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>NRc1 CPC2 NRc2 CPC1</td>
<td>4.0000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>CA1 CA2 CA3 CA4</td>
<td>3.1000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>CA2 CA1 CA2 CA3</td>
<td>3.1000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>CA2 CA3 CA4 CA3</td>
<td>3.1000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>Oc2 CAC CAC CAC</td>
<td>3.1000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>CA2 CA3 CA4 OcH</td>
<td>3.1000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>CA1 CA2 CA3 HPC</td>
<td>4.2000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>CA2 CA1 CA2 HPC</td>
<td>4.2000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>CA3 CA4 CA3 HPC</td>
<td>4.2000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>HPC CA2 CA3 CA4</td>
<td>4.2000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>HPC CA2 CA3 HPC</td>
<td>2.4000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>!OcH Oc2 CAC CAC</td>
<td>0.9900</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>!OcH OcH CA4 CA3</td>
<td>0.9900</td>
<td>2</td>
<td>180.00 !</td>
</tr>
<tr>
<td>HPC CA3 CA4 OcH</td>
<td>4.2000</td>
<td>2</td>
<td>180.00 !</td>
</tr>
</tbody>
</table>

CPC2 NRc2 CPC1 CEc1  3.000  2   180.00 !
NRc1 CPc1 CPc1 CEc1 3.00 2 180.00 !
Oc2 CPc1 CPc1 CEc1 2.00 2 180.00 !
CEcl CA1 CA2 HPc 4.20 2 180.00 !
CEcl CA1 CA2 CA3 3.10 2 180.00 !
!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 !
!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 CPc1 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 CPc1 NRc1 CT2 14.0000 2 180.00 !
CTc1 CPc2 NRc1 CT2 14.0000 2 180.00 !

! Linking the chromophore and the glycine(67) fragment
O  C  CT2  NRc1  0.0000 1  0.00 !
NH1  C  CT2  NRc1  0.6000 1  0.00 !
CPc2 NRc1 CT2 HB  0.0320 3  0.00 !
CPc2 NRc1 CT2 C  0.0320 3  0.00 !
CPc1 NRc1 CT2 HB  0.0320 3 180.00 !
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
Charmmm22
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 HA  0.2000 3  0.00
CPc2 CTc1 CT2 CT2  0.2000 3  0.00
NH1 CTc1 CT2 HA  0.2000 3  0.00
NH1 CTc1 CT2 CT2  0.2000 3  0.00

IMPROPER
!
!V(improper) = Kpsi(ψi - ψi0)**2
!
!Kpsi: kcal/mole/rad**2
psi0: degrees
note that the second column of numbers (0) is ignored

atom types          Kpsi                  psi0

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) = Eps,i,j[(Rmin,i,j/ri,j)**12 - 2(Rmin,i,j/ri,j)**6]
epsilon: kcal/mole, Eps,i,j = sqrt(eps,i * eps,j)
Rmin/2: A, Rmin,i,j = Rmin/2,i + Rmin/2,j
atom ignored    epsilon Rmin/2 ignored   eps,1-4 Rmin/2,1-4

CAc  5.000000 -0.070000  1.992400 ! ALLOW ARO

NONBONDED nbxmod 5 atom cdiel shift vatom vdistance vswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 el4fac 1.0 wmin 1.5

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 -0.068000  2.090000 ! ! for propene, yin/adm
jr., 12/95
CPc1 0.000000 -0.050000  1.800000 ! ALLOW ARO
CPc2 0.000000 -0.050000  1.800000 ! ALLOW ARO
HAc 0.000000 -0.022000  1.320000 ! ALLOW PEP ALI POL SUL
HAc1 0.000000 -0.031000  1.250000
HPc  0.000000 -0.030000  1.358200
NRc1 0.000000 -0.200000  1.850000 ! ALLOW ARO
NRc2 0.000000 -0.200000  1.850000 ! ALLOW ARO
Oc2  0.000000 -0.120000  1.700000 ! ALLOW PEP POL,
OcH  0.000000 -0.152100  1.770000 ! ALLOW ALC ARO
CTc1 0.000000 -0.020000  2.275000
Appendix C: Residue topology and parameter files for the chromophore of citrine

Table 5.3 Residue topology file for the chromophore of citrine (anionic)

<table>
<thead>
<tr>
<th>MASS</th>
<th>197 NRc2</th>
<th>14.00700 N</th>
<th>neutral his unprotonated ring nitrogen</th>
</tr>
</thead>
<tbody>
<tr>
<td>MASS</td>
<td>198 NRc1</td>
<td>14.00700 N</td>
<td>neutral his protonated ring nitrogen</td>
</tr>
<tr>
<td>MASS</td>
<td>199 HAc1</td>
<td>1.00800 H</td>
<td>for alkene; RHC=CR</td>
</tr>
<tr>
<td>MASS</td>
<td>200 HAc2</td>
<td>1.00800 H</td>
<td>aromatic H</td>
</tr>
<tr>
<td>MASS</td>
<td>201 Oc2</td>
<td>15.99900 O</td>
<td>carbonyl oxygen</td>
</tr>
<tr>
<td>MASS</td>
<td>202 OcH</td>
<td>15.99900 O</td>
<td>from OH1</td>
</tr>
<tr>
<td>MASS</td>
<td>204 HAc</td>
<td>1.00800 H</td>
<td>nonpolar H</td>
</tr>
<tr>
<td>MASS</td>
<td>205 CA1</td>
<td>12.01100 C</td>
<td>aromatic C</td>
</tr>
<tr>
<td>MASS</td>
<td>206 CA2</td>
<td>12.01100 C</td>
<td>aromatic C</td>
</tr>
<tr>
<td>MASS</td>
<td>207 CA3</td>
<td>12.01100 C</td>
<td>aromatic C</td>
</tr>
<tr>
<td>MASS</td>
<td>208 CPC2</td>
<td>12.01100 C</td>
<td>his CE1 carbon</td>
</tr>
<tr>
<td>MASS</td>
<td>209 CPC1</td>
<td>12.01100 C</td>
<td>for alkene; RHC=CR</td>
</tr>
<tr>
<td>MASS</td>
<td>210 CPC1</td>
<td>12.01100 C</td>
<td>his CG and CD2 carbons</td>
</tr>
<tr>
<td>MASS</td>
<td>211 CA4</td>
<td>12.01100 C</td>
<td>aromatic C</td>
</tr>
<tr>
<td>MASS</td>
<td>213 CTc1</td>
<td>12.01100 C</td>
<td>aliphatic sp3 C for CH</td>
</tr>
</tbody>
</table>

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

DEFA FIRS NTER LAST CTER
AUTO ANGLES DIHE
RESI CRO -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 CZ CA4 0.45
ATOM OH OCH -0.62
Glycine (67) part from Charmm22

GROUP
ATOM CA3 CT2 -0.18 ! | 
ATOM HA31 HB  0.09 ! | 
ATOM HA32 HB  0.09 ! HA1-CA-HA2
GROUP
ATOM C3 C  0.51 ! | 
ATOM O3 O  -0.51 ! C=O

Gly (65) part from charmm22

GROUP
ATOM N1 NH1 -0.47 !atom type changed
ATOM HN1 H  0.31
ATOM CA1 CTc1 -0.02
ATOM HA11 HB  0.09
ATOM HA12 HB  0.09

BOND CA1 C1 N1 -C +N CA1 N1 N1 HN1 CA1 HA11 CA1 HA12
BOND N2 CA2 CB2 HB2 CB2 CG2 CD1 HD1 CD1 CE1 CE1 HE1 CZ OH
BOND C2 CE2 CE2 HE2 CD2 HD2 CD2 CG2 CA2 C2
BOND N3 CA3 CA3 HA31 CA3 HA32 CA3 C3 N3 C1 N3 C2
DOUBLE C1 N2 CA2 CB2 C2 O2 C3 O3 CD1 CG2 CD2 CE2 CZ CE1

Table 5.4 Parameter file for the chromophore of citrine (anionic)

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

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

CPC2 CTc1  354.000 1.4900 !ion for RFP C1-CA1 connection
NRc1 CT2  396.000 1.4400 !ion
NRc1 CPC2 400.000 1.3900 !
NRc1 CPC1 400.000 1.4100 !
CPC1 OC2  854.000 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 !
CEc1 CA1  437.000 1.4100 !ion

84
<table>
<thead>
<tr>
<th>Atom 1</th>
<th>Atom 2</th>
<th>Distance</th>
<th>Angle</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA1</td>
<td>CA2</td>
<td>305.000</td>
<td>1.4300</td>
<td>ion</td>
</tr>
<tr>
<td>HpC</td>
<td>CA2</td>
<td>340.000</td>
<td>1.0800</td>
<td></td>
</tr>
<tr>
<td>CA2</td>
<td>CA3</td>
<td>305.000</td>
<td>1.3500</td>
<td>ion</td>
</tr>
<tr>
<td>HpC</td>
<td>CA3</td>
<td>340.000</td>
<td>1.0800</td>
<td></td>
</tr>
<tr>
<td>CA3</td>
<td>CA4</td>
<td>305.000</td>
<td>1.4550</td>
<td>ion</td>
</tr>
<tr>
<td>OcH</td>
<td>CA4</td>
<td>842.000</td>
<td>1.2500</td>
<td>ion</td>
</tr>
<tr>
<td>NH1</td>
<td>CTc1</td>
<td>320.000</td>
<td>1.4300</td>
<td></td>
</tr>
<tr>
<td>HB</td>
<td>CTc1</td>
<td>330.000</td>
<td>1.0800</td>
<td></td>
</tr>
</tbody>
</table>

**ANGLES**

\[ V(\text{angle}) = K_{\theta}(\Theta - \Theta_0)^{**2} \]

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

\( K_{\theta}: \text{kcal/mole/rad}^{**2} \)

\( \Theta_0: \text{degrees} \)

\( K_{UB}: \text{kcal/mole/A}^{**2} \) (Urey-Bradley)

\( S_0: \text{A} \)

<table>
<thead>
<tr>
<th>Atom 1</th>
<th>Atom 2</th>
<th>Atom 3</th>
<th>Distance</th>
<th>Angle</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRc2</td>
<td>CPc2</td>
<td>NRc1</td>
<td>130.000</td>
<td>113.300</td>
<td>ion</td>
</tr>
<tr>
<td>CPc2</td>
<td>NRc2</td>
<td>CPc1</td>
<td>130.000</td>
<td>106.600</td>
<td>ion</td>
</tr>
<tr>
<td>CPc2</td>
<td>NRc1</td>
<td>CPc1</td>
<td>130.000</td>
<td>107.900</td>
<td></td>
</tr>
<tr>
<td>NRc2</td>
<td>CPc1</td>
<td>CPc1</td>
<td>130.000</td>
<td>108.300</td>
<td></td>
</tr>
<tr>
<td>NRc2</td>
<td>CPc1</td>
<td>CEc1</td>
<td>45.800</td>
<td>129.500</td>
<td></td>
</tr>
<tr>
<td>NRc1</td>
<td>CPc1</td>
<td>Oc2</td>
<td>50.000</td>
<td>124.000</td>
<td>ion</td>
</tr>
<tr>
<td>NRc1</td>
<td>CPc1</td>
<td>CPc1</td>
<td>130.000</td>
<td>103.000</td>
<td></td>
</tr>
<tr>
<td>Oc2</td>
<td>CPc1</td>
<td>CPc1</td>
<td>44.000</td>
<td>133.000</td>
<td>ion</td>
</tr>
<tr>
<td>CPc1</td>
<td>CPc1</td>
<td>CEc1</td>
<td>45.800</td>
<td>122.700</td>
<td></td>
</tr>
<tr>
<td>CPc1</td>
<td>CEc1</td>
<td>CA1</td>
<td>130.000</td>
<td>133.200</td>
<td>ion</td>
</tr>
<tr>
<td>CPc1</td>
<td>CEc1</td>
<td>HAcl</td>
<td>42.000</td>
<td>112.000</td>
<td></td>
</tr>
<tr>
<td>CEc1</td>
<td>CA1</td>
<td>CA2</td>
<td>45.800</td>
<td>120.000</td>
<td>ion</td>
</tr>
<tr>
<td>HAcl</td>
<td>CEc1</td>
<td>CA1</td>
<td>42.000</td>
<td>115.000</td>
<td>ion</td>
</tr>
<tr>
<td>CA1</td>
<td>CA2</td>
<td>CA3</td>
<td>40.000</td>
<td>122.000</td>
<td>ion</td>
</tr>
<tr>
<td>CA2</td>
<td>CA1</td>
<td>CA2</td>
<td>40.000</td>
<td>116.000</td>
<td>ion</td>
</tr>
<tr>
<td>CA2</td>
<td>CA3</td>
<td>CA4</td>
<td>40.000</td>
<td>122.000</td>
<td>ion</td>
</tr>
<tr>
<td>CA3</td>
<td>CA4</td>
<td>CA3</td>
<td>40.000</td>
<td>115.000</td>
<td>ion</td>
</tr>
<tr>
<td>HpC</td>
<td>CA3</td>
<td>CA4</td>
<td>30.000</td>
<td>120.000</td>
<td></td>
</tr>
<tr>
<td>HpC</td>
<td>CA3</td>
<td>CA2</td>
<td>30.000</td>
<td>120.000</td>
<td></td>
</tr>
<tr>
<td>HpC</td>
<td>CA2</td>
<td>CA3</td>
<td>30.000</td>
<td>120.000</td>
<td></td>
</tr>
<tr>
<td>HpC</td>
<td>CA2</td>
<td>CA1</td>
<td>30.000</td>
<td>120.000</td>
<td></td>
</tr>
<tr>
<td>OcH</td>
<td>CA4</td>
<td>CA3</td>
<td>45.200</td>
<td>120.000</td>
<td>ALLOW ARO ALC</td>
</tr>
</tbody>
</table>

!Link to the gly(65) fragment

<table>
<thead>
<tr>
<th>Atom 1</th>
<th>Atom 2</th>
<th>Distance</th>
<th>Angle</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH1</td>
<td>CTc1</td>
<td>CPc2</td>
<td>50.000</td>
<td>107.000</td>
</tr>
<tr>
<td>HB</td>
<td>CTc1</td>
<td>CPc2</td>
<td>50.000</td>
<td>109.500</td>
</tr>
<tr>
<td>NRc2</td>
<td>CPc2</td>
<td>CTc1</td>
<td>40.000</td>
<td>125.000</td>
</tr>
<tr>
<td>NRc1</td>
<td>CPc2</td>
<td>CTc1</td>
<td>40.000</td>
<td>121.700</td>
</tr>
<tr>
<td>CTc1</td>
<td>NH1</td>
<td>C</td>
<td>50.000</td>
<td>120.000</td>
</tr>
<tr>
<td>NH1</td>
<td>CTc1</td>
<td>HB</td>
<td>48.000</td>
<td>108.000</td>
</tr>
<tr>
<td>H</td>
<td>NH1</td>
<td>CTc1</td>
<td>35.000</td>
<td>117.000</td>
</tr>
</tbody>
</table>
**Link to the gly(67) fragment**

<table>
<thead>
<tr>
<th>Atom Types</th>
<th>Kchi</th>
<th>n</th>
<th>delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPc2 NRc2 CPc1 CPc1</td>
<td>14.0000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CPc2 NRc1 CPc1 CPc1</td>
<td>14.0000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>NRc2 CPc2 NRc1 CPc1</td>
<td>14.0000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>NRc2 CPc1 CPc1 NRc1</td>
<td>4.0000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>NRc1 CPc2 NRc2 CPc1</td>
<td>4.0000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CA1 CA2 CA3 CA4</td>
<td>3.1000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CA2 CA1 CA2 CA3</td>
<td>3.1000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CA2 CA3 CA4 CA3</td>
<td>3.1000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>Oc2 CAC CAC CAC</td>
<td>3.1000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CA2 CA3 CA4 OcH</td>
<td>3.1000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CA1 CA2 CA3 Hpc</td>
<td>4.2000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CA2 CA1 CA2 Hpc</td>
<td>4.2000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CA3 CA4 CA3 Hpc</td>
<td>4.2000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>Hpc CA2 CA3 CA4</td>
<td>4.2000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>Hpc CA2 CA3 Hpc</td>
<td>4.2000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>Hpc CA3 CA4 OcH</td>
<td>4.2000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CPc2 NRc2 CPc1 CEc1</td>
<td>3.000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>NRc1 CPc1 CPc1 CEc1</td>
<td>3.00</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>Oc2 CPc1 CPc1 CEc1</td>
<td>2.00</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CEc1 CA1 CA2 Hpc</td>
<td>4.20</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CEc1 CA1 CA2 CA3</td>
<td>3.10</td>
<td>2</td>
<td>180.00</td>
</tr>
</tbody>
</table>

**Connection CA-CB**

<table>
<thead>
<tr>
<th>Atom Types</th>
<th>Kchi</th>
<th>n</th>
<th>delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPc1 CPc1 CEc1 HAc1</td>
<td>3.9000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CPc1 CPc1 CEc1 CA1</td>
<td>3.9000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>NRc2 CPc1 CEc1 HAc1</td>
<td>3.9000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>NRc2 CPc1 CEc1 CA1</td>
<td>3.9000</td>
<td>2</td>
<td>180.00</td>
</tr>
</tbody>
</table>

**Connection CB-CG2**

<table>
<thead>
<tr>
<th>Atom Types</th>
<th>Kchi</th>
<th>n</th>
<th>delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPc1 CEc1 CA1 CA2</td>
<td>2.7000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>HAc1 CEc1 CA1 CA2</td>
<td>2.7000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>CPc2 NRc1 CPc1 Oc2</td>
<td>14.0000</td>
<td>2</td>
<td>180.00</td>
</tr>
<tr>
<td>NRc2 CPc2 NRc1 CT2</td>
<td>14.0000</td>
<td>2</td>
<td>180.00</td>
</tr>
</tbody>
</table>
NRc2 CPc1 CPc1 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 CPc1 NRc1 CT2  14.0000  2  180.00 !
CTc1 CPc2 NRc1 CT2  14.0000  2  180.00 !

! Linking the chromophore and the glycine(67) fragment
O C CT2 NRc1  0.0000  1  0.00 !
NH1 C CT2 NRc1  0.6000  1  0.00 !
CPc2 NRc1 CT2 HB  0.0320  3  0.00 !
CPc2 NRc1 CT2 C  0.0320  3  0.00 !
CPc1 NRc1 CT2 HB  0.0320  3  180.00 !
CPc1 NRc1 CT2 C  0.0320  3  180.00 !

! Linking the chromophore and the gly(65) fragment
C NH1 CTc1 CPc2  0.2000  1  180.00 !
NRc2 CPc2 CTc1 NH1  0.1050  3  180.00 !
NRc1 CPc2 CTc1 NH1  0.1050  3  0.00 !
H NH1 CTc1 CPc2  0.0000  1  0.00 !
NRc2 CPc2 CTc1 HB  0.1050  3  180.00 !
NRc1 CPc2 CTc1 HB  0.1050  3  0.00 !
CT1 C NH1 CTc1  1.6000  1  0.00 ! ALLOW PEP
HB CTc1 NH1 C  0.0000  1  0.00 ! ALLOW PEP
O C NH1 CTc1  2.5000  2  180.00 ! ALLOW PEP
HB CTc1 NH1 H  0.0000  1  0.00 ! ALLOW PEP

IMPROPER

!V(improper) = Kpsi(psi - psi0)**2

!Kpsi: kcal/mole/rad**2
!psi0: degrees
!note that the second column of numbers (0) is ignored

!atom types           Kpsi                   psi0

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
Appendix D: Residue topology and parameter files for the chromophore of citrine

Table 5.5 Residue topology file for the chromophore of citrine (neutral)
Table 5.6 Parameter file for the chromophore of citrine (neutral)

*Charmm parameter file of citrine Chromophore (gly-tyr-gly)
* !parameter file

! GFP Chromophore parameters, protonated form ! BONDS
!
!V(bond) = Kb(b - b0)**2 !
!Kb: kCal/mole/A**2 !b0: A
<table>
<thead>
<tr>
<th>atom type</th>
<th>Kb</th>
<th>b0</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPC2 CTC1</td>
<td>320.000</td>
<td>1.4900</td>
</tr>
<tr>
<td>NRC1 CT2</td>
<td>352.000</td>
<td>1.4500</td>
</tr>
<tr>
<td>NRC1 CPC2</td>
<td>400.000</td>
<td>1.3900</td>
</tr>
<tr>
<td>NRC1 CPC1</td>
<td>400.000</td>
<td>1.4100</td>
</tr>
<tr>
<td>CPC1 OC2</td>
<td>807.000</td>
<td>1.2200</td>
</tr>
<tr>
<td>NRC2 CPC2</td>
<td>400.000</td>
<td>1.3000</td>
</tr>
<tr>
<td>CPC1 CPC1</td>
<td>410.000</td>
<td>1.4900</td>
</tr>
<tr>
<td>NRC2 CPC1</td>
<td>400.000</td>
<td>1.4000</td>
</tr>
<tr>
<td>CPC1 CEC1</td>
<td>560.000</td>
<td>1.3600</td>
</tr>
<tr>
<td>HAC1 CEC1</td>
<td>360.500</td>
<td>1.1000</td>
</tr>
<tr>
<td>CEC1 CA1</td>
<td>370.000</td>
<td>1.4500</td>
</tr>
<tr>
<td>CA1 CA2</td>
<td>305.000</td>
<td>1.3750</td>
</tr>
<tr>
<td>HPC CA2</td>
<td>340.000</td>
<td>1.0800</td>
</tr>
<tr>
<td>CA2 CA3</td>
<td>305.000</td>
<td>1.3750</td>
</tr>
<tr>
<td>HPC CA3</td>
<td>340.000</td>
<td>1.0800</td>
</tr>
<tr>
<td>CA3 CA4</td>
<td>305.000</td>
<td>1.3750</td>
</tr>
<tr>
<td>OCH CA4</td>
<td>334.300</td>
<td>1.4110</td>
</tr>
<tr>
<td>OCH HCH</td>
<td>545.000</td>
<td>0.9600</td>
</tr>
<tr>
<td>NH1 CTC1</td>
<td>320.000</td>
<td>1.4300</td>
</tr>
<tr>
<td>HB CTC1</td>
<td>330.000</td>
<td>1.0800</td>
</tr>
</tbody>
</table>

**ANGLES**

! V(angle) = Ktheta(Theta - Theta0)**2

! V(Urey-Bradley) = Kub(S - S0)**2

! Ktheta: kCal/mole/rad**2
! Theta0: degrees
! Kub: kCal/mole/A**2 (Urey-Bradley)
! S0: A

<table>
<thead>
<tr>
<th>atom types</th>
<th>Ktheta</th>
<th>Theta0</th>
<th>Kub</th>
<th>S0</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRC2 CPC2</td>
<td>130.000</td>
<td>114.0000</td>
<td>ion</td>
<td></td>
</tr>
<tr>
<td>CPC2 NRC2</td>
<td>130.000</td>
<td>106.0000</td>
<td>ion</td>
<td></td>
</tr>
<tr>
<td>CPC2 NRC1</td>
<td>130.000</td>
<td>107.9000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>NRC2 CPC1</td>
<td>130.000</td>
<td>108.3000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>NRC2 CPC1</td>
<td>45.800</td>
<td>129.5000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>NRC1 CPC1</td>
<td>42.000</td>
<td>126.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>NRC1 CPC1</td>
<td>130.000</td>
<td>103.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>OC2 CPC1</td>
<td>38.000</td>
<td>132.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>CPC1 CPC1</td>
<td>45.800</td>
<td>122.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>CPC1 CEC1</td>
<td>130.000</td>
<td>130.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>CPC1 CEC1</td>
<td>42.000</td>
<td>114.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>CEC1 CA1</td>
<td>42.000</td>
<td>116.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>HAC1 CEC1</td>
<td>42.000</td>
<td>116.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>Atom 1</td>
<td>Atom 2</td>
<td>Atom 3</td>
<td>Atom 4</td>
<td>KChi</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>CA1</td>
<td>CA2</td>
<td>CA3</td>
<td>CA4</td>
<td>40.000</td>
</tr>
<tr>
<td>CA2</td>
<td>CA1</td>
<td>CA2</td>
<td>40.000</td>
<td>120.000</td>
</tr>
<tr>
<td>CA2</td>
<td>CA3</td>
<td>CA4</td>
<td>40.000</td>
<td>120.000</td>
</tr>
<tr>
<td>CA3</td>
<td>CA4</td>
<td>CA3</td>
<td>40.000</td>
<td>120.000</td>
</tr>
<tr>
<td></td>
<td>HPC</td>
<td>CA4</td>
<td>30.000</td>
<td>120.000</td>
</tr>
<tr>
<td></td>
<td>HPC</td>
<td>CA3</td>
<td>30.000</td>
<td>120.000</td>
</tr>
<tr>
<td></td>
<td>HPC</td>
<td>CA2</td>
<td>30.000</td>
<td>120.000</td>
</tr>
<tr>
<td></td>
<td>HPC</td>
<td>CA2</td>
<td>30.000</td>
<td>120.000</td>
</tr>
<tr>
<td></td>
<td>OCH</td>
<td>CA4</td>
<td>45.200</td>
<td>120.000</td>
</tr>
<tr>
<td></td>
<td>HCH</td>
<td>OCH</td>
<td>65.000</td>
<td>108.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NH1</td>
<td>CTC1</td>
<td>CPC2</td>
<td>50.000</td>
</tr>
<tr>
<td></td>
<td>HB</td>
<td>CTC1</td>
<td>CPC2</td>
<td>50.000</td>
</tr>
<tr>
<td></td>
<td>NRC2</td>
<td>CPC2</td>
<td>CTC1</td>
<td>40.000</td>
</tr>
<tr>
<td></td>
<td>NRC1</td>
<td>CPC2</td>
<td>CTC1</td>
<td>35.000</td>
</tr>
<tr>
<td></td>
<td>CTC1</td>
<td>NH1</td>
<td>C</td>
<td>50.000</td>
</tr>
<tr>
<td></td>
<td>NH1</td>
<td>CTC1</td>
<td>HB</td>
<td>48.000</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>NH1</td>
<td>CTC1</td>
<td>35.000</td>
</tr>
<tr>
<td></td>
<td>HB</td>
<td>CTC1</td>
<td>HB</td>
<td>36.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CPC2</td>
<td>NRC1</td>
<td>CT2</td>
<td>36.000</td>
</tr>
<tr>
<td></td>
<td>CPC1</td>
<td>NRC1</td>
<td>CT2</td>
<td>32.000</td>
</tr>
<tr>
<td></td>
<td>NRC1</td>
<td>CT2</td>
<td>C</td>
<td>50.000</td>
</tr>
<tr>
<td></td>
<td>NRC1</td>
<td>CT2</td>
<td>HB</td>
<td>48.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CPC2</td>
<td>NRC2</td>
<td>CPC1</td>
<td>CPC1</td>
</tr>
<tr>
<td></td>
<td>CPC2</td>
<td>NRC1</td>
<td>CPC1</td>
<td>CPC1</td>
</tr>
<tr>
<td></td>
<td>NRC2</td>
<td>CPC2</td>
<td>NRC1</td>
<td>CPC1</td>
</tr>
<tr>
<td></td>
<td>NRC2</td>
<td>CPC1</td>
<td>CPC1</td>
<td>NRC1</td>
</tr>
<tr>
<td></td>
<td>NRC1</td>
<td>CPC2</td>
<td>NRC2</td>
<td>CPC1</td>
</tr>
<tr>
<td></td>
<td>CA1</td>
<td>CA2</td>
<td>CA3</td>
<td>CA4</td>
</tr>
</tbody>
</table>
CPC2 NRC2 CPC1 CEC1 3.0000 2 180.00 !
NRC1 CPC1 CPC1 CEC1 3.0000 2 180.00 !
OC2 CPC1 CPC1 CEC1 2.0000 2 180.00 !
CEC1 CA1 CA2 HPC 4.2000 2 180.00 !
CEC1 CA1 CA2 CA3 3.1000 2 180.00 !

! ConneCtion CA-CB
CPC1 CPC1 CEC1 HAC1 6.84000 2 180.00 !
CPC1 CPC1 CEC1 CA1 6.84000 2 180.00 !
NRC2 CPC1 CEC1 HAC1 6.84000 2 180.00 !
NRC2 CPC1 CEC1 CA1 6.84000 2 180.00 !

! ConneCtion CB-CG2
CPC1 CEC1 CA1 CA2 1.40000 2 180.00 !
HAC1 CEC1 CA1 CA2 1.40000 2 180.00 !

! Linking the Chromophore and the glyCine(67) fragment
O C CT2 NRC1 0.0000 1 0.00 !
NH1 C CT2 NRC1 0.6000 1 0.00 !
CPC2 NRC1 CT2 HB 0.0670 3 0.00 !
CPC2 NRC1 CT2 C 0.0670 3 0.00 !
CPC1 NRC1 CT2 HB 0.0670 3 180.00 !
CPC1 NRC1 CT2 C 0.0670 3 180.00 !

! Linking the Chromophore and the met(65) fragment
C NH1 CTC1 CPC2 0.2000 1 180.00 !
NRC2 CPC2 CTC1 NH1 0.1000 3 180.00 !
NRC1 CPC2 CTC1 NH1 0.1000 3 0.00 !
H NH1 CTC1 CPC2 0.0000 1 0.00 !
NRC2 CPC2 CTC1 HB 0.1000 3 180.00 !
NRC1 CPC2 CTC1 HB  0.1000  3   0.00 !
H   NH1 CTC1 HB   0.0000  1   0.00
CT1 C   NH1 CTC1  2.5000  2   180.00
O   C   NH1 CTC1  2.5000  2   180.00
C   NH1 CTC1 HB  0.0000  1   0.00

IMPROPER
!
!V(improper) = Kpsi(psi - psi0)**2
!
!Kpsi: kCal/mole/rad**2
!psi0: degrees
!note that the seCond Column of numbers (0) is ignored
!
!atom types           Kpsi                   psi0
!
CPC2 NRC2 NRC1 CTC1  0.5000         0      0.0000
CPC2 NRC1 NRC2 CTC1  0.5000         0      0.0000
!
CPC1 NRC1 CPC1 OC2  0.5000         0      0.0000
CPC1 CPC1 NRC1 OC2  0.5000         0      0.0000
!
NRC1 CPC1 CPC2 CT2  0.4500         0      0.0000
NRC1 CPC2 CPC1 CT2  0.4500         0      0.0000
!
CPC1 NRC2 CPC1 CEC1  220.0000         0      0.0000
CPC1 CPC1 NRC2 CEC1  220.0000         0      0.0000
!
CEC1 CPC1 CA1 HAC1  30.0000         0      0.0000
CEC1 CA1 CPC1 HAC1  30.0000         0      0.0000
!

!V(Lennard-Jones) = Eps,i,j[(Rmin,i,j/ri,j)**12 - 2(Rmin,i,j/ri,j)**6]
!
!epsilon: kCal/mole, Eps,i,j = sqrt(eps,i * eps,j)
!Rmin/2: A, Rmin,i,j = Rmin/2,i + Rmin/2,j
!
!atom ignored    epsilon Rmin/2 ignored   eps,1-4 Rmin/2,1-4
!
!CAC  5.000000  -0.070000  1.992400 ! ALLOW   ARO
NONBONDED nbxmod  5 atom cdiel shift vatom vdistance vswitch -cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5
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  -0.068000     2.090000 ! ! for propene, yin/adm jr., 12/95
CPC1   0.000000  -0.050000     1.800000 ! ALLOW ARO
CPC2   0.000000  -0.050000     1.800000 ! ALLOW ARO
!CT3C   0.000000  -0.080000     2.060000  0.000000  -0.010000 1.900000 ! ALLOW ALI
HCH   -2.000000  -0.046000     0.224500 ! ALLOW PEP POL SUL ARO ALC
HAC    0.000000  -0.022000     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
OC2    0.000000  -0.120000     1.700000 ! ALLOW PEP POL,
OCH    0.000000  -0.152100     1.770000 ! ALLOW ALC ARO
CTC1   0.000000  -0.020000     2.275000  0.000000  -0.010000 1.900000
!HBOND CUTHB 0.5 ! If you want to do hbond analysis (only), then use
! READ PARAM APPEND CARD
! to append hbond parameters from the file:

Appendix E: Residue topology and parameter files for the chromophore of mCherry

Table 5.7 Residue topology file for the chromophore of mCherry (neutral)

<table>
<thead>
<tr>
<th>RESI</th>
<th>CH6</th>
<th>0.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP</td>
<td>! Imidazolinone ring</td>
<td></td>
</tr>
<tr>
<td>ATOM</td>
<td>C1</td>
<td>CPC2 0.76</td>
</tr>
<tr>
<td></td>
<td>N2</td>
<td>NRC2 -0.55</td>
</tr>
<tr>
<td></td>
<td>N3</td>
<td>NRC1 -0.64</td>
</tr>
</tbody>
</table>
ATOM C2 CPC1 0.80
ATOM O2 OC2 -0.61
ATOM CA2 CPC1 0.24
GROUP
ATOM CB2 CEC1 -0.10
ATOM HB2 HAC1 0.10
GROUP
ATOM CG2 CA1 0.00 ! Tyr ring : Charges from Charmm22
ATOM CD1 CA2 -0.115
ATOM HD1 HPC 0.115
ATOM CD2 CA2 -0.115
ATOM HD2 HPC 0.115
ATOM CE1 CA3 -0.115
ATOM HE1 HPC 0.115
ATOM CE2 CA3 -0.115
ATOM HE2 HPC 0.115
ATOM CZ CA4 0.110
ATOM OH OCH -0.54
ATOM HH HCH 0.43

! GlyCine part from Charmm22
GROUP
ATOM CA3 CT2 -0.18 ! |
ATOM HA31 HB 0.09 ! |
ATOM HA32 HB 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 CTC1 0.16 ! atom type Changed
GROUP
ATOM CB1 CT2 -0.18
ATOM HB11 HA 0.09
ATOM HB12 HA 0.09
GROUP
ATOM CG1 CT2 -0.14
ATOM HG11 HA 0.09
ATOM HG12 HA 0.09
ATOM SD S -0.09
ATOM CE CT3 -0.22
ATOM HE11 HA 0.09
ATOM HE12 HA 0.09
ATOM HE13 HA 0.09

BOND CA1 C1 N1 -C C3 +N
BOND N2 CA2 CB2 HB2 CB2 CG2 CD1 HD1 CD1 CE1 CE1 HE1 CZ OH OH HH
BOND CZ CB2 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
Table 5.8 Parameter file for the chromophore of mCherry (neutral)

*Charmm parameter file of mCherry Chromophore (met-tyr-gly) (NEUTRAL)*
* !parameter file

! GFP Chromophore parameters, protonated form !
BONDS 
! !V(bond) = Kb(b - b0)**2 
! !Kb: kCal/mole/A**2 
!b0: A 
! !atom type Kb          b0
CPC2 CTC1  320.000     1.4900 !ion for RFP C1-CA1 ConneCtion
NRC1 CT2   352.000     1.4500 !ion
NRC1 CPC2  400.000     1.3900 !
NRC1 CPC1  400.000     1.4100 !
CPC1 OC2   807.000     1.2200 !ion
NRC2 CPC2  400.000     1.3000 !
CPC1 CPC1  410.000     1.4900 !ion
NRC2 CPC1  400.000     1.4000 !
CPC1 CEC1  560.000     1.3600 !ion
HAC1 CEC1  360.500     1.1000 !
CEC1 CA1   370.000     1.4500 !ion
CA1 CA2   305.000     1.3750 !ion
HPC CA2   340.000     1.0800 !
CA2 CA3   305.000     1.3750 !ion
HPC CA3   340.000     1.0800 !
CA3 CA4   305.000     1.3750 !ion
OCH CA4   334.300     1.4110 !ion
OCH HCH   545.000     0.9600

CTC1 NH1   463.000     1.3650 !RFP CA1-N1 ConneCtion
CT2 CTC1  222.500     1.5380 !RFP CB1-CA1 ConneCtion

ANGLES 
! !V(angle) = Ktheta(Theta - Theta0)**2 
! !V(Urey-Bradley) = Kub(S - S0)**2
<table>
<thead>
<tr>
<th>atom types</th>
<th>Ktheta</th>
<th>Theta0</th>
<th>Kub</th>
<th>S0</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRC2 CPC2 NRC1</td>
<td>130.000</td>
<td>114.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC2 NRC2 CPC1</td>
<td>130.000</td>
<td>106.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC2 NRC1 CPC1</td>
<td>130.000</td>
<td>107.9000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>NRC2 CPC1 CPC1</td>
<td>130.000</td>
<td>108.3000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>NRC2 CPC1 CEC1</td>
<td>45.800</td>
<td>129.5000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>NRC1 CPC1 OC2</td>
<td>42.000</td>
<td>126.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>NRC1 CPC1 CPC1</td>
<td>130.000</td>
<td>103.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>OC2 CPC1 CPC1</td>
<td>38.000</td>
<td>132.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC1 CPC1 CEC1</td>
<td>45.800</td>
<td>122.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC1 CEC1 CA1</td>
<td>130.000</td>
<td>130.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CPC1 CEC1 HAC1</td>
<td>42.000</td>
<td>114.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>HAC1 CEC1 CA1</td>
<td>42.000</td>
<td>116.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CA1 CA2 CA3</td>
<td>40.000</td>
<td>120.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CA2 CA1 CA2</td>
<td>40.000</td>
<td>120.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CA2 CA3 CA4</td>
<td>40.000</td>
<td>120.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>CA3 CA4 CA3</td>
<td>40.000</td>
<td>120.0000</td>
<td>!ion</td>
<td></td>
</tr>
<tr>
<td>HPC CA3 CA4</td>
<td>30.000</td>
<td>120.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>HPC CA3 CA2</td>
<td>30.000</td>
<td>120.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>HPC CA2 CA3</td>
<td>30.000</td>
<td>120.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>HPC CA2 CA1</td>
<td>30.000</td>
<td>120.0000</td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>OCH CA4 CA3</td>
<td>45.200</td>
<td>120.0000</td>
<td>ALLOW</td>
<td>ARO ALC</td>
</tr>
<tr>
<td>HCH OCH CA4</td>
<td>65.000</td>
<td>108.0000</td>
<td>! add this for neutral</td>
<td></td>
</tr>
</tbody>
</table>

!Link to the met(65) fragment

| NH1 CTC1 CPC2 | 50.000 | 107.0000 | ALLOW | PEP POL ARO ALI |
| CT2 CTC1 CPC2 | 52.000 | 108.0000 | ALLOW | ALI PEP POL ARO |
| NRC2 CPC2 CTC1 | 40.000 | 125.0000 | !     |
| NRC1 CPC2 CTC1 | 35.000 | 121.4000 | !ion  |
| CT2 CT1 CPC2 | 52.000 | 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 | |
| HA CT2 CTC1 | 33.430 | 110.1000 | 22.53 | 2.17900 |
| CT2 CT2 CTC1 | 58.350 | 113.50 | 11.16 | 2.56100 |

!Link to the gly(67) fragment

| CPC2 NRC1 CT2 | 36.000 | 129.0000 | !ion  |
```
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CPC1 NRC1 CT2</td>
<td>32.000</td>
<td>123.4000</td>
<td>!ion</td>
</tr>
<tr>
<td>NRC1 CT2 C</td>
<td>50.000</td>
<td>107.0000</td>
<td></td>
</tr>
<tr>
<td>NRC1 CT2 HB</td>
<td>48.000</td>
<td>108.0000</td>
<td></td>
</tr>
</tbody>
</table>

DIHEDRALS

\[
V(\text{dihedral}) = K\Chi(1 + \cos(n(\Chi) - \delta))
\]

K\Chi: kCal/mole
n: multiplicity
\delta: degrees

atom types | K\Chi | n | \delta |
---|---|---|---|
CPC2 NRC2 CPC1 CPC1 | 14.0000 | 2 | 180.00 | !
CPC2 NRC1 CPC1 CPC1 | 14.0000 | 2 | 180.00 | !
NRC2 CPC2 NRC1 CPC1 | 14.0000 | 2 | 180.00 | !
NRC2 CPC1 CPC1 NRC1 | 4.0000 | 2 | 180.00 | !
NRC1 CPC2 NRC2 CPC1 | 4.0000 | 2 | 180.00 | !
CA1 CA2 CA3 CA4 | 3.1000 | 2 | 180.00 | !
CA2 CA1 CA2 CA3 | 3.1000 | 2 | 180.00 | !
CA2 CA3 CA4 CA3 | 3.1000 | 2 | 180.00 | !
CA2 CA3 CA4 OCH | 3.1000 | 2 | 180.00 | !
CA1 CA2 CA3 HPC | 4.2000 | 2 | 180.00 | !
CA2 CA1 CA2 HPC | 4.2000 | 2 | 180.00 | !
CA3 CA4 CA3 HPC | 4.2000 | 2 | 180.00 | !
HPC CA2 CA3 CA4 | 4.2000 | 2 | 180.00 | !
HPC CA2 CA3 HPC | 2.4000 | 2 | 180.00 | !
HCH OCH CA4 CA3 | 0.9900 | 2 | 180.00 | !
HPC CA3 CA4 OCH | 4.2000 | 2 | 180.00 | !

!ConneCtion CA-CB
CPC1 CPC1 CEC1 HAC1 | 6.8400 | 2 | 180.00 | !
CPC1 CPC1 CEC1 CA1 | 6.8400 | 2 | 180.00 | !
NRC2 CPC1 CEC1 HAC1 | 6.8400 | 2 | 180.00 | !
NRC2 CPC1 CEC1 CA1 | 6.8400 | 2 | 180.00 | !for green 180

!ConneCtion CB-CG2
CPC1 CEC1 CA1 CA2 | 1.4000 | 2 | 180.00 | !
HAC1 CEC1 CA1 CA2 | 1.4000 | 2 | 180.00 | !

CPC2 NRC1 CPC1 OC2 | 14.0000 | 2 | 180.00 | !
NRC2 CPC2 NRC1 CT2 | 14.0000 | 2 | 180.00 | !
NRC2 CPC1 CPC1 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 CPC1 NRC1 CT2  14.0000  2   180.00 !
CTC1 CPC2 NRC1 CT2  14.0000  2   180.00 !

! Linking the Chromophore and the glyCine(67) fragment
O    C    CT2  NRC1     0.0000  1     0.00 !
NH1  C    CT2  NRC1     0.6000  1     0.00 !
CPC2 NRC1 CT2  HB       0.0670  3   180.00 !
CPC2 NRC1 CT2  C        0.0670  3   180.00 !
CPC1 NRC1 CT2  HB       0.0670  3   180.00 !
CPC1 NRC1 CT2  C        0.0670  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
Charm22
NRC2 CPC2 CTC1 CT2  0.1000  3  180.00 !
NRC2 CPC2 CTC1 NH1  0.1000  3  180.00 !
NRC1 CPC2 CTC1 CT2  0.1000  3   0.00 !
NRC1 CPC2 CTC1 NH1  0.1000  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  HA       0.2000  3   0.00 !
CPC2 CTC1 CT2  CT2      0.2000  3   0.00 !
NH1  CTC1 CT2  HA       0.2000  3   0.00 !
NH1  CTC1 CT2  CT2      0.2000  3   0.00 !

IMPROPER
!
!V(improper) = Kpsi(ψi - ψi0)**2
!
!Kpsi: kCal/mole/rad**2
!ψi0: degrees
!note that the seCond Column of numbers (0) is ignored
!
<table>
<thead>
<tr>
<th>atom types</th>
<th>Kpsi</th>
<th>psi0</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPC2 NRC2 NRC1 CTC1</td>
<td>0.5000</td>
<td>0.0000</td>
</tr>
<tr>
<td>CPC2 NRC1 NRC2 CTC1</td>
<td>0.5000</td>
<td>0.0000</td>
</tr>
<tr>
<td>CPC1 NRC1 CPC1 OC2</td>
<td>0.5000</td>
<td>0.0000</td>
</tr>
<tr>
<td>CPC1 CPC1 NRC1 OC2</td>
<td>0.5000</td>
<td>0.0000</td>
</tr>
<tr>
<td>NRC1 CPC1 CPC2 CT2</td>
<td>0.4500</td>
<td>0.0000</td>
</tr>
<tr>
<td>NRC1 CPC2 CPC1 CT2</td>
<td>0.4500</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
! CPC1 NRC2 CPC1 CEC1  220.0000         0      0.0000
CPC1 CPC1 NRC2 CEC1  220.0000         0      0.0000
!
CEC1 CPC1 CA1 HAC1   30.0000         0      0.0000
CEC1 CA1 CPC1 HAC1   30.0000         0      0.0000

!V(Lennard-Jones) = Eps,i,j[(Rmin,i,j/ri,j)**12 - 2(Rmin,i,j/ri,j)**6]
!
!epsilon: kCal/mole, Eps,i,j = sqrt(eps,i * eps,j)
!Rmin/2: A, Rmin,i,j = Rmin/2,i + Rmin/2,j
!
!atom  ignored  epsilon  Rmin/2  ignored  eps,i-4  Rmin/2,i-4
!
!CAC  5.000000  -0.070000  1.992400  ! ALLOW   ARO
NONBONDED nbxmod 5 atom cdiel shift vatom vdistance vswitch - cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5

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  -0.068000  2.090000  ! for propene, yin/adm jr., 12/95
CPC1 0.000000  -0.050000  1.800000  ! ALLOW ARO
CPC2 0.000000  -0.050000  1.800000  ! ALLOW ARO

!CT3C 0.000000  -0.080000  1.900000  ! ALLOW ALI

HCH -2.000000  -0.046000  0.224500  ! ALLOW PEP POL SUL ARO

ALC

HAC  0.000000  -0.022000  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

OC2  0.000000  -0.120000  1.700000  ! ALLOW PEP POL,
OCH  0.000000  -0.152100  1.770000 ! ALLOW ALC ARO

CTC1  0.000000  -0.020000  2.275000  0.000000  -0.010000
      1.900000

!HBOND CUTHB 0.5  ! If you want to do hbond analysis (only),
then use
      ! READ PARAM APPEND CARD
      ! to append hbond parameters from the file:
VITA

CHOLA K. REGMI

2008 – 2014  Graduate Teaching Assistant (Ph. D. Student)
Department of Physics, Florida International University
Miami, FL 33199, USA.

2000 - 2004  M. Sc. in physics, Tribhuvan University, Kathmandu, Nepal

1997 – 2000  B. Sc. in physics, Tribhuvan University, Kathmandu, Nepal

Publications:


Awards:

New Investigator Travel Award 2014, APS Division of Chemical Physics, APS March Meeting 2014

Travel Award, ‘Workshop on Multiscale Theory and Simulation’, University of Chicago, Chicago, IL, June 17-19, 2013

Second Prize Award, Annual Graduate Scholarly Forum, Florida International University, Miami, FL 33199, USA, March 26, 2013

First Prize Award, Graduate Student Research Competition, Department of Physics, Florida International University, Miami, FL 33199, USA, April 19, 2012
<table>
<thead>
<tr>
<th>Year</th>
<th>Presentations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>C. Regmi, P. Chapagain, and B. Gerstman, “Temperature dependent solvation dynamics of the chromophore environment in the far-red fluorescent protein mPlum” APS March Meeting, March 03-07, 2014, Denver, CO, USA</td>
</tr>
<tr>
<td>2013</td>
<td>C. Regmi, “Diffusion of molecular oxygen in the red fluorescent protein mCherry” Annual Graduate Scholarly Forum, Florida International University, Miami, FL 33199, USA, March 26, 2013</td>
</tr>
<tr>
<td>2013</td>
<td>C. Regmi, Y. Bhandari, P. Chapagain, and B. Gerstman, “Diffusion of molecular oxygen in the red fluorescent protein mCherry” APS March Meeting, March 18-22, 2013, Baltimore, MD, USA</td>
</tr>
<tr>
<td>2012</td>
<td>C. Regmi, “Barrel fluctuations and oxygen diffusion pathways in the monomeric fluorescent proteins”, Graduate Student Research Competition, Department of Physics, Florida International University, Miami, FL 33199, USA, April 19, 2012</td>
</tr>
<tr>
<td>2011</td>
<td>C. Regmi, P. Chapagain, and B. Gerstman, “Protein barrel fluctuations and the barrel permeability: A comparison between Green and Red Fluorescent Proteins”, APS March Meeting, March 21-25, 2011, Dallas, TX, USA</td>
</tr>
</tbody>
</table>