**Distribution Agreement**

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Yingrong Chen November 14, 2023

Quantum Decoherence of Green Fluorescence Proteins

by

Yingrong Chen

Fang Liu, Ph.D.

Adviser

Chemistry

Fang Liu, Ph.D.

Adviser

Francesco Evangelista, Ph.D.

Committee Member

Jed Brody, Ph.D.

Committee Member

2023

Quantum Decoherence of Green Fluorescence Proteins

by

Yingrong Chen

Fang Liu, Ph.D.

Adviser

An abstract of

a thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of science with Honors

Chemistry

2023

Abstract

Quantum Decoherence of Green Fluorescence Proteins

By Yingrong Chen

Quantum batteries, rooted in quantum mechanics, are revolutionizing energy storage with their superextensive charging capability, which allows them to charge rapidly as their capacity increases. The previous prototype, utilizing an organic microcavity with polaritons as "qubits," faces limitations due to exciton-exciton annihilation. To address this, we propose replacing organic dyes in the microcavity with fluorescent proteins, leveraging their protein cylinder as a "molecular bumper" to mitigate annihilation. For practical battery function, the fluorescent protein-based microcavity must delay quantum decoherence, where the quantum system collapses into classical states due to environmental interactions.

Inspired by biological systems with enduring quantum coherence, our research explores how protein and solvent environments influence energy gap fluctuations and quantum decoherence. The hypothesis posits that mutations reducing interactions and collisions can delay quantum decoherence. Using hybrid quantum mechanics-molecular mechanics (QMMM) simulations with both explicit and implicit models, we studied three GFP mutants—citrine, eGFP, and GFP.

The findings reveal that citrine, with mutations (Q65M and T199Y) introducing π-π interactions and steric clashes, exhibits a shorter estimated decoherence time. In contrast, eGFP shows prolonged coherence, attributed to lower water density surrounding its chromophore. Explicit models also show shorter decoherence times due to dynamic solvent-chromophore interactions. Additionally, we establish a correlation between structural dynamics metrics like RMSD and RMSF and quantum decoherence, emphasizing their value for future rational design. The model is validated with experimental absorption spectra using time-dependent density functional theory (TDDFT). In conclusion, this study enhances our understanding of quantum decoherence in fluorescent proteins, providing a knowledge base for the rational design of quantum batteries.

# Quantum Decoherence of Green Fluorescence Proteins

by

Yingrong Chen

Fang Liu, Ph.D.

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Science with Honors

Chemistry

2023

Acknowledgment

I am sincerely thankful to:

- My thesis advisor, Dr. Fang Liu, who guided me at every turn in this project and supported my exploration of different avenues.
- My committee members, Dr. Jed Brody and Dr. Francesco Evangelista, for their enlightening quantum computing class and comprehensive quantum chemistry lab. Both courses laid the foundational knowledge for this project.
- Dr. Deepika Das, my research mentor at Oxford College, who introduced me to the fascinating world of computational chemistry and continues to be a guiding mentor in every aspects of my academic journey.
- Dr. Benjamin Brown, my research mentor at the Meiler Lab at Vanderbilt University, who broadened my horizons through enlightening discussions and patiently taught me essential research skills in Amber, cpptraj, and pyemma.
- Patrick Li, Fangning Ren and other members of the Liu Group, whose assistance was invaluable in implementing this project.

A heartfelt appreciation to my dedicated teachers at Emory, including Professor Brenda Harmon, Dr. Nitya P. Jacob, Dr. Alfred Farris, and Dr. David Lynn. Special thanks to my parents for supporting my research exploration and to my friends who have grown up with me throughout these four years.

## Table of Contents

# Introduction

## Quantum Battery

Quantum batteries are an emerging area of research that could revolutionize the energy storage and retrieval industries. While traditional batteries have been pivotal in technological advancements across personal electronics and automotive industries, they still rely on the principles of electrochemistry established in the 18th century [1]. Thus, the expanding energy industries require new approaches that harness the principles of quantum mechanics.

Quantum mechanics governs the behavior of particles at the atomic and subatomic levels, giving rise to nonclassical phenomena like coherence and entanglement. Decades of progress in quantum devices development have demonstrated the concept of "quantum advantage." "Quantum advantage" highlights how quantum devices can outperform their traditional counterparts in specific tasks, such as quantum sensing, computing, and communication. For instance, quantum computers can simulate many-particle quantum systems exponentially faster than classical computers [2]. Now, the focus of emerging research has shifted toward quantum energy processing, leading to the intriguing idea of quantum batteries [3].

In contrast to traditional batteries, which suffer from slower charging rates as their capacity increases, quantum batteries possess a remarkable property known as superextensive charging [4]. That is, quantum batteries charge at an accelerated rate as their capacity grows. This property holds promise for applications that require both substantial and rapid energy storage, such as electric vehicles. It could also reduce energy wastage during storage and retrieval, contributing to sustainability.

The superextensive charging property arises from how quantum batteries are engineered based on principles of quantum mechanics **(Fig. 1)**. Energy is stored in "qubits", which can be implemented with ions, neutral atoms, and other physical forms with multiple energy levels. An ensemble of many qubits is embedded in a single cavity coupling with an external energy source. Charging a quantum battery excites all qubits to higher energy states by a protocol of "interaction" between the qubit ensemble and the external field. Conversely, discharging relaxes these qubits to lower accessible states. The energy is

stored in work that can be reversibly extracted from the system. Superextensive charging is made possible through entanglement between qubits that reduces the number of traversed states [5] or cooperative behavior that increases the coupling between battery and energy source [6]. The more qubits, and hence the more entanglement, the faster the charging process.

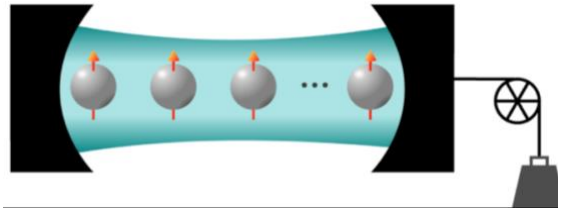


*Fig. 1 Quantum battery implemented with an ensemble of qubits coupled with external energy source [7].*

## Quantum Battery Built from Microcavity

Very recently, a prototype of quantum battery was realized in an organic microcavity coupled with ultrafast optical spectroscopy (**Fig. 2 a)**. [8]. A microcavity is an engineered structure to confine and control photons (**Fig. 2 b)**. It consists of two facing dielectric mirrors, the distance between which determines the wavelength of confined photons in the microcavity. Between the mirrors lies a layer of semiconductor dispersed in a polymer matrix. The confined photons will interact with the semiconductor molecules, which excites particles to higher energy states and creates electron-hole pairs known as excitons. Strong coupling between the incoming photons and the created excitons forms special particles called polaritons, which exhibit characteristics of both light and matter. Polaritons demonstrate quantum properties such as Bose-Einstein condensation and superfluidity. Moreover, they are capable of entanglement and coherent oscillations, so they are the "qubits" in this prototype [9]. To measure the evolution of stored energy, researchers excited the microcavity using a pump pulse and measured the number of excited molecules with a second probe pulse [8]. As the number of molecules in the microcavity increases, the charging power density also increases, and it takes less time to charge [8].

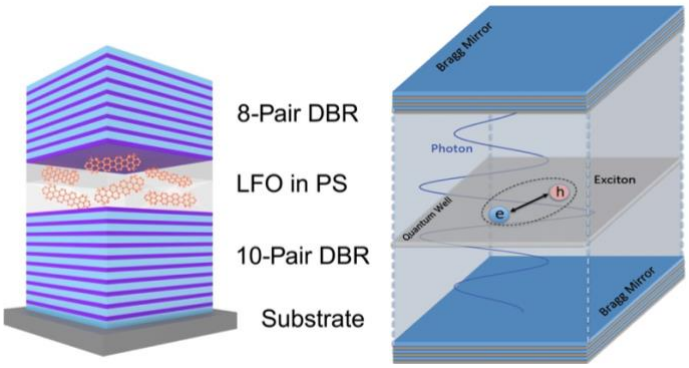



*Fig. 2 Quantum battery prototype implemented with microcavity.* (a) Quantum battery implemented with microcavity of Lumogen-F orange (LFO) in a polystyrene (PS) matrix between Bragg reflectors (DBRs). This prototype used polaritons as "qubits" and demonstrated superextensive charging [8]. (b) Typical exciton-polariton microcavity with two Bragg mirrors and quantum well. The coupling between photon and exciton gives arises to polaritons [10].

## Exciton-Exciton Annihilation Necessitates the Use of Fluorescent Protein

However, this prototype of quantum battery is limited by exciton-exciton annihilation. As the concentration of organic dye Lumogen-F orange (LFO) in the semiconductor increases, the quantum yield drops to zero [8]. This phenomenon, also known as exciton-exciton annihilation or intermolecular quenching, arises from the nonradiative recombination of excitons and the dissipation of energy as heat into the surrounding environment (**Fig. 3)**. Exciton-exciton annihilation is detrimental to microcavity because it limits device lifetime, makes it challenging to scale up the system, and reduces the efficiency for energy storage or other optical applications.

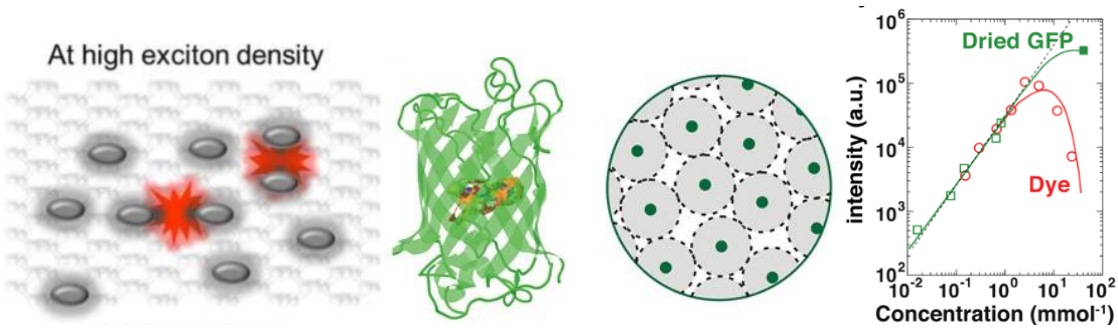



*Fig. 3 Exciton-exciton annihilation and green fluorescent protein.* (a) exciton-exciton annihilation [11](b) enhanced green fluorescent protein (eGFP) structure (c) eGFP's protein cylinder as "molecular bumpers" (d) higher fluorescence intensity of eGFP compared to organic dye at high concentrations[12].

To address the issue of exciton-exciton annihilation in microcavities, researchers have replaced LFO in the semiconductor with biologically sourced enhanced green fluorescent protein (eGFP) [12]. The active component in eGFP is a chromophore formed by autocatalytic cyclization of three amino acids. The chromophore is surrounded by a nanocylinder composed of 11 β sheets (**Fig. 3 b)**. This protein cylinder acts as a "molecular bumper," maintaining a separation between chromophores even within densely packed arrangement (**Fig. 3 c)**. eGFP has been engineered in the microcavity as the active semiconductor component positioned between two mirrors (**Fig. 4)** [12]. The resulting microcavity demonstrates polariton properties, such as low-threshold coherent light generation, superfluidity, and small Rabi splitting. Importantly, eGFP does not undergo exciton annihilation, so it exhibits higher fluorescence intensity than organic dye at high concentrations (**Fig. 3 d)** [12]. This suggests that a quantum battery prototype can be innovatively implemented with an eGFP-filled microcavity. Furthermore, by engineering fluorescent protein mutants other than eGFP inside the microcavity, we can study how the protein cylinder control the microcavity's energy transfer and quantum decoherence – an important property to consider in engineering quantum battery discussed in next section.

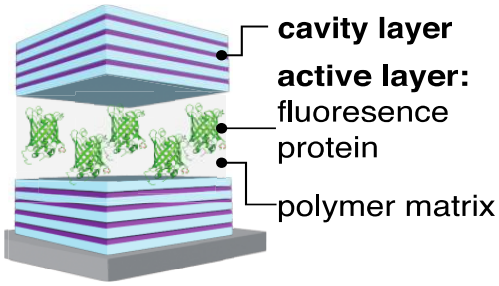



*Fig. 4 An innovative quantum battery prototype: fluorescent protein-based microcavity.*

## Energy Retention Issue – Quantum Decoherence

How can a quantum battery charge quickly while stabilizing the energy and discharging much more slowly? The energy retention problem is associated with a phenomenon called quantum decoherence. Quantum decoherence occurs when a quantum system "collapses" back into the classical regime due to its interactions with the surrounding environment (**Fig. 5**). Ideally, we prefer quantum systems to remain completely isolated. However, realistic quantum systems are exposed to external fields, rapidly entangling with numerous environmental degrees of freedom. Quantum decoherence explains why it's challenging to observe quantum phenomena like superposition in our everyday experiences – they quickly collapse even if they are successfully made.

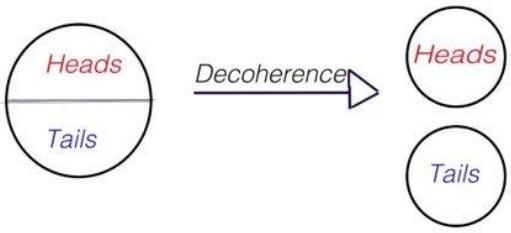



*Fig. 5 Quantum decoherence on a quantum superposition state of 'heads' and 'tails'.* And its collapse into two distinct 'heads' and 'tails' states due to quantum decoherence

To mitigate quantum decoherence in the fluorescent protein-based quantum battery, we can begin by studying biological systems known for their long-lasting quantum coherence. Biological functions like photosynthesis and fluorescence depend on quantum transitions from ground to excited states as well as

phenomena like entanglement and tunneling. Thus, researchers have long been interested in the extent and impact of quantum decoherence in biological systems. Long-lived quantum coherence has been observed in systems like light-harvesting complexes (**Fig. 6**) [13]. Excitation energy can become delocalized among closely spaced chromophores and their associated donor and acceptor states, driving multiple energy transfer pathways simultaneously [14]. This enhances the efficiency of energy transfer between the chromophore antenna and the photosynthetic reaction center [15].

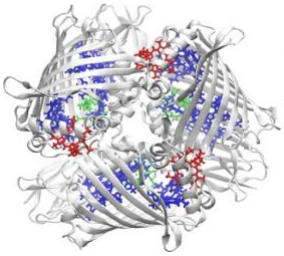


*Fig. 6 Light-harvesting complexes Fenna−Matthews−Olson (FMO) with multiple chromophores [16].*

This remarkable long-lived quantum coherence is attributed to small energy gap fluctuation [17]. When considering chromophores as two-level systems (TLS), the energy gap fluctuation between the ground and first excited states is determined by the correlated evolution of energy levels and the smoothness of each potential energy surface (**Fig. 7 a**). When the energy surface of the ground state vertically aligns with that of the excited state along certain reaction coordinate, it leads to correlated evolution of energy levels, thereby enhancing quantum coherence. On the other hand, quantum coherence is impaired by small bumps in the potential energy surfaces, which often occur due to the coupling of the reaction coordinate with other vibrational modes. This inverse relationship between energy gap fluctuation and quantum coherence are also found in the Fenna−Mat- thews−Olson complexes, the PE545 aggregate, the LH2 complexes, DNA, photolyase, and cryptochromes (**Fig. 7 b**) [1819]. Thus, we may improve quantum coherence in a biological system by minimizing the energy gap fluctuation.

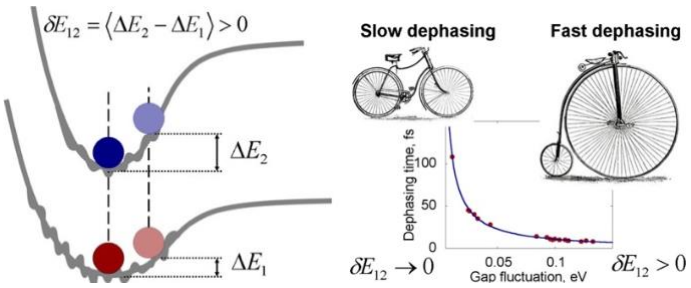



*Fig. 7 Energy gap fluctuation and quantum decoherence*. (a) energy gap fluctuation controlled by the correlated evolution of energy levels and the smoothness of each potential energy surface (b) inverse relationship between energy gap fluctuation and quantum coherence [17]

## Rational Design of Fluorescent Proteins to Minimize Decoherence

Biological systems, such as fluorescent proteins, can be examined as quantum subsystems interacting with the surrounding environment of proteins and solvents. Initially, the chromophore exists in a superposition of the ground and excited states. However, chemical interactions and collisions with surrounding proteins and solvents lead to quantum decoherence (**Fig. 8 a**). In our pursuit of designing fluorescent proteins for quantum batteries, it is crucial to understand how the protein cylinder and active water influence the chromophore decoherence. However, gaining such insights in experiments can be challenging due to the time and resources required to set up the microcavity. Thus, this research performs computer simulations to reveal the quantum decoherence of various fluorescent protein mutants. The protein cylinder of these mutants forms different hydrogen bonds and electrostatic interactions with the chromophore. The hypothesis is that the smoothness of potential energy surface can be improved by reducing random collisions between the environment and the quantum subsystem, while the environment can polarize the subsystem in a way that its energy levels become more correlated. The resulting design rules will enable rational design of fluorescent proteins to make highly efficient quantum batteries.

Green fluorescent proteins (GFP) have revolutionized biomedical imaging by allowing researchers to label and track biological components in living organisms. Notably, scientists have engineered a vibrant spectrum of fluorescent protein variants with diverse maturation, brightness,

stability, and emission bands, an achievement that was honored with the Nobel Prize in Chemistry in 2008. These properties are primarily controlled by the chemical structures of the chromophore, but also finely tuned by the noncovalent interactions between the chromophore and the protein cylinder as well as water molecules. This research studies two prominent GFP mutants, citrine and eGFP, along with the wild-type GFP, all sharing the same chromophore (CRO) (**Fig. 8 b**). Several residues on the surrounding β-barrel are key to the CRO excitation, so they are predicted to affect the quantum decoherence. For instance, the T203Y in citrine mutant removes the T203–CRO hydrogen bond and creates the π–π stacking between the six-membered rings of Y203 and CRO. This increases the CRO polarizability and results in a significant red-shift [20].

On the other hand, water molecules play a crucial role in chromophore excitation and emission [21]. Some water molecules can form hydrogen bond with nitrogen in the CRO imidazolinone ring, while others can mediate the interaction between surrounding amino acids and CRO [20]. To explore the effects of water on CRO decoherence, this research used both implicit and explicit solvent models.

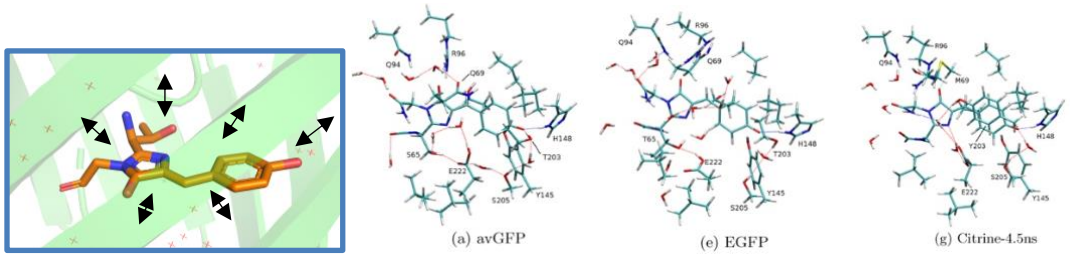



*Fig. 8 Mutants of GFP and their surrounding environments.* **(**a) effects of environment on the GFP chromophore's quantum decoherence (b) mutants of GFP studied in this research [20]

## Theory

### Quantum Decoherence

Formally, decoherence can be viewed as a filter acting on the state space, selecting stable classical states while excluding unstable entangled ones [22]. It happens either through continuous monitoring by the environment or through a rapid measurement, or both.

"Before" the subsystem A interacts with the environment B, it is a coherent pure state described by a wave function $\Psi(\mathrm{A})$:

$$\Psi(\mathrm{A}) = \sum_{\mathrm{I}\in\mathrm{A}} \mathrm{c}_I |\psi_I\rangle$$

With coefficient $\mathrm{c}_I$ and basis set $\{\psi_I\}$. And its density operator $\hat{\rho}$ can be expressed as:

$$\hat{\rho} = |\Psi\rangle\langle\Psi| = \sum_{IJ} \mathrm{c}_I \mathrm{c}_J^* |\psi_I\rangle\langle\psi_J|$$

The diagonal elements of the density matrix are called populations, or probabilities, while the off-diagonal elements are called coherences. On the other hand, a mixed state – a statistical mixture of pure states $\Psi_L$ – cannot be described by a wave function, but it can be described by a density matrix:

$$\hat{\rho} = \sum_{L} \mathrm{p}_L |\Psi_\mathrm{L}\rangle\langle\Psi_L|$$

where $p_L$ is the classical probability of finding the system in state $L$. Instead of the coherent superposition in pure state, this is a classical mixture of quantum states.

"After" the subsystem A interacts with the environment B with multiple state J expressed in the basis $\{|\mathrm{s}_\mathcal{K}\rangle\}$:

$$\phi_J^{(B)} = \sum_{\mathcal{K}\in\mathrm{B}} \alpha_{\mathrm{J}\mathcal{K}} |\mathrm{s}_\mathcal{K}\rangle$$

The combined system–environment state becomes:

$$\Psi(\mathrm{AB}) = \sum_{\mathrm{I}\in\mathrm{A}} \mathrm{c}_I |\psi_I\rangle|\phi_I\rangle = \sum_{\mathrm{I}\in\mathrm{A},\mathcal{K}\in\mathrm{B}} c_I \alpha_{\mathrm{I}\mathcal{K}} |\psi_I\rangle|\mathrm{s}_\mathcal{K}\rangle$$

Suppose there are two states each in the subsystem A and the environment B, the combined system–environment state is:

$$\Psi = c_1\alpha_{11}|\psi_1\rangle|s_1\rangle + c_1\alpha_{12}|\psi_1\rangle|s_2\rangle + c_2\alpha_{21}|\psi_2\rangle|s_1\rangle + c_2\alpha_{22}|\psi_2\rangle|s_2\rangle$$

The coefficients $c_I\alpha_{I\mathcal{J}}$ cannot be expressed in a product form, indicating entanglement between system A and environment B. The density operator for the subsystem A is given by:

$$\hat{\rho}\,(A) = \sum_{IJ\in A} c_I c_J^{*}\,\langle\phi_I|\phi_J\rangle|\psi_I\rangle\langle\psi_J|$$

$\hat{\rho}\,(A)$ contains terms describing the overlaps between environment states. When the environmental degrees of freedom are large and the environmental states are orthogonal, the off-diagonal elements vanish because $\delta_{IJ} = \langle\phi_I|\phi_J\rangle$. Although the combined system–environment state is still a pure state, the subsystem A becomes a mixed state.

In fluorescent protein, the chromophores can be described as two-level systems (TLS) with only the ground and first excited state, represented with coefficient $c_1$ and $c_2$. The density matrix of the chromophore subsystem "before" and "after" environmental interaction is:

$$\rho_{before} = \begin{pmatrix} |c_1|^2 & c_1{}^{*}c_2 \\ c_1c_2{}^{*} & |c_2|^2 \end{pmatrix} \rightarrow \rho_{after} = \begin{pmatrix} |c_1|^2 & 0 \\ 0 & |c_2|^2 \end{pmatrix}$$

The off-diagonal, coherence terms vanish due to decoherence (**Fig. 9**).

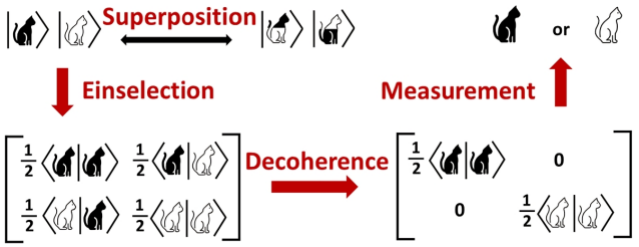



*Fig. 9 Math theories behind quantum decoherence represented with Schrödinger's cat.* The vanishing of off-diagonal, coherence terms in the density matrix due to decoherence. [23]

Decoherence can also be viewed as the loss of information from a system into the environment. That is, "the environment measures the system". As the molecule become entangled

with environmental degrees of freedom, the environment keeps track of the quantum information in the system, even though it's not a conscious observer. The continuous measurement-like interaction leads to the decay of quantum behavior.

## Decoherence Time and Energy Gap Fluctuation

The relationship between quantum decoherence and energy gap fluctuations in biological systems can be described by a generalized version of the Kubo stochastic line shape theory [24]. For the HOMO-LUMO energy gaps $E_{ij}\,(t)$ of green fluorescent protein, the autocorrelation function is calculated by:

$$C_{ij}(t) = \langle \delta E_{ij}(t)\delta E_{ij}(0)\rangle, \text{ where } \delta E_{ij}(t) = E_{ij}(t) - \langle E_{ij}\rangle.$$

The gap fluctuation $\langle \Delta E^2\rangle^{1/2}$ is used to estimate the decoherence time $\tau_D$ by:

$$\tau_D = \frac{B}{\sqrt{\langle \Delta E^2\rangle}}, \text{ with } B = \sqrt{2}\hbar \text{ or } B = \sqrt{\frac{12}{5}}\hbar$$

The following is a simplified derivation of the $\tau_D$ equation. Generally, the decoherence function D(t) is a convolution f * g of exponential decay f(t) and Gaussian decay g(t) [25]. The unnormalized correlation function C(t) in Kubo stochastic line shape theory consists of two parts with normalized prefactors $\alpha_1$ and $\alpha_2$, where $\alpha_1 + \alpha_2 = 1$, along with coherence times $\tau_{c,1}$ and $\tau_{c,2}$:

$$C(t) = \langle \Delta E^2\rangle(\alpha_1 \exp(-\frac{t}{\tau_{c,1}}) + \alpha_2 \exp(-\frac{t}{\tau_{c,2}}))$$

If the fluctuations in the system include slow and fast fluctuations at the same time, the line shape function can be approximated by:

$$g(t) = \frac{\langle \Delta E^2\rangle}{\hbar^2}(\alpha_1\tau_{c,1}t + \alpha_2 t^2/2))$$

The decoherence function is given by:

$$D(t) = \exp(-g(t))$$

The dephasing time (the onset of quantum decoherence) $\tau_D$ is defined as:

$$\tau_D = \frac{2}{\sqrt{\pi}}\int_0^\infty D(t)dt = \sqrt{\frac{2\hbar^2}{\langle \Delta E^2\rangle\alpha_2}}\, e^{-A\langle \Delta E^2\rangle}\operatorname{erfc}(\sqrt{A\langle \Delta E^2\rangle})$$

erfc(x) denotes the complementary error function and the constant A is given by:

$$\mathrm{A} = \frac{\tau_1^2}{2\hbar^2}\frac{\alpha_1^2}{a_2}$$

In the limit as $a_1$ approaches to 0, the dephasing time can be simplified to:

$$\tau_{\mathrm{D,G}} = \frac{B}{\sqrt{\langle \Delta \mathrm{E}^2 \rangle}}$$

The constant B is fitted in previous research to be $\mathrm{B} = \sqrt{2}\hbar$ or $\mathrm{B} = \sqrt{\frac{12}{5}}\hbar$ [17].

## Molecular Dynamics (MD)

The energy gap fluctuation is calculated by computational simulation of fluorescent protein mutants. Molecular dynamics (MD) simulates a system of interacting molecules by numerically solving Newton's equations of motion. For each atom, the starting position $R(0)$ is obtained from experimental data, and the starting velocity $v(0) = \frac{d\,R(0)}{dt}$ is assigned based on a specified temperature. The new positions and velocities after a time interval $\Delta t$ is predicted with acceleration $a(t)$. $a(t)$ is computed from Newton's second law and the force $F(t)$:

$$F(t) = ma(t)$$

In turn, the overall force acting on an atom is computed from a potential energy function $V(R)$ based on the position $R(t)$:

$$F(t) = -\frac{\partial V(R)}{\partial R}$$

The potential energy function can be derived from either quantum mechanics or molecular force field. A molecular force field comprises a set of parameters used to calculate interatomic forces, taking into account factors such as chemical bonds, bond angles, bond dihedrals, as well as non-bonded interactions like van der Waals forces and electrostatic charges. The computational cost for evaluating the molecular potential scales as $O(N^2)$, accounting for the interactions between each particle and all others. However, it's essential to recognize that a

molecular force field doesn't capture electron behavior and, therefore, cannot be utilized to depict chemical reactions, electron transfer processes, nor fluorescence phenomena.

Molecular movements are propagated in n steps, and the total simulation time is $\Delta t \times n$. Within each step, the system is updated using integrator like the Velocity Verlet method:

$$\begin{cases} R(t + \Delta t) = R(t) + v(t)\Delta t + \frac{1}{2} a(t)\Delta t^2 \\ v(t + \Delta t) = v(t) + \frac{1}{2} \{a(t) + a(t + \Delta t)\}\Delta t \end{cases}$$

By default, the number of atoms (N), the volume of the system (V), and the total energy (E) remain constant, which is known as the microcanonical ensemble (NVE). NVE ensemble represents an adiabatic process without any heat exchange. However, biological systems frequently interact with their environment through endothermic and exothermic processes, causing variations in their total energy. In such cases, simulations can be configured to maintain a constant number of atoms (N), volume (V), and temperature (T) using the canonical ensemble (NVT). Temperature is estimated from the kinetic energy of the system $\frac{nk_BT}{2}$ , and is controlled using a thermostat. Alternatively, simulations can be set to maintain constant pressure by employing the isothermal–isobaric ensemble (NPT) in place of constant volume.

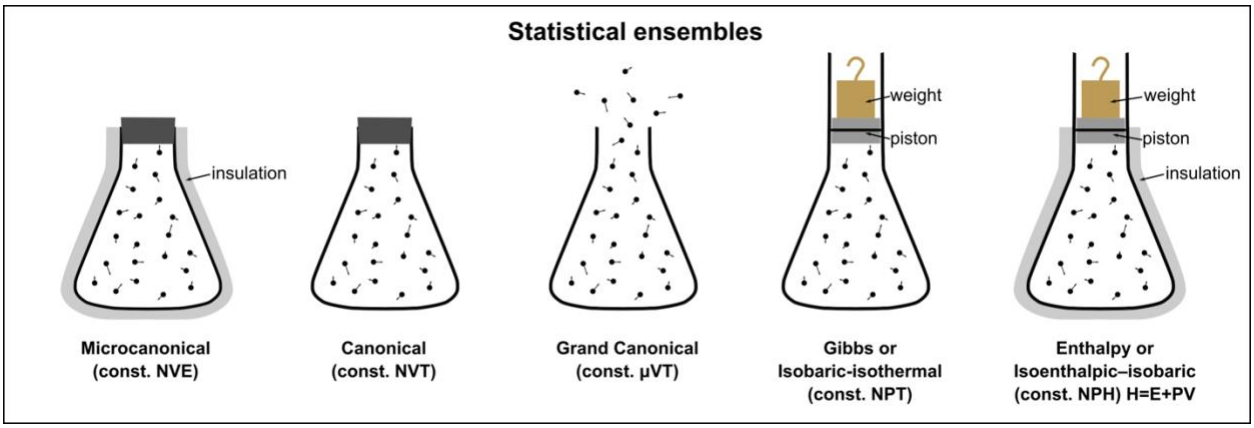



*Fig. 10 Statistical ensembles in Molecular Dynamics[26].*

Solvents can be depicted either explicitly or implicitly in simulations. Explicit solvent model substantially increases the computational cost, as it involves tenfold more particles to calculate in each step. In contrast, generalized born implicit solvent models treat the solute as a collection of spheres with a

designated dielectric constant. The implicit solvent model also account for solvation energies and forces acting on the solute atoms [27].

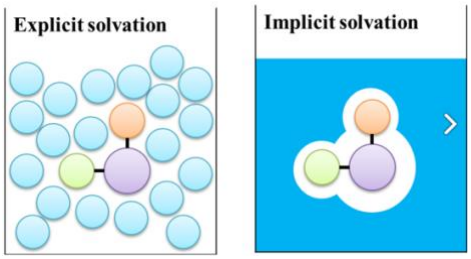



*Fig. 11 Explicit versus Implicit solvent models in Molecular Dynamics [26].*

Instead of having a single system comprising one protein and its surrounding solvents, the system exists within a bulk of other similar systems and interacts with the protein and solvents in other systems. Thus, periodic boundary conditions are used to connect one side of the simulation seamlessly to the opposite side. This condition mimics the behavior of multiple instances of the same system interacting with each other (**Fig. 12**). In the simulation, electrostatics are evaluated using the Particle Mesh Ewald method. This method combines the short-ranged potential calculated in real space with the long-ranged potential calculated in Fourier space, resulting in faster computation with a time complexity $O(\mathrm{N}\log N)$, as opposed to the direct calculation method with a time complexity of $O(N^2)$.

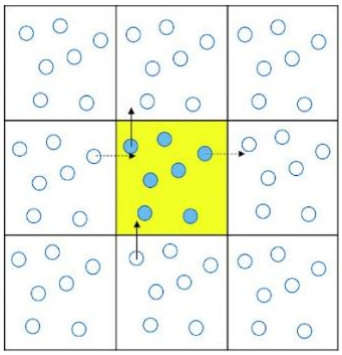


*Fig. 12 Periodic boundary condition in Molecular Dynamics [28].*

## Quantum Mechanics–Molecular Mechanics (QM/MM)

The fluorescent protein cylinder and solvent molecules collectively possess a multitude of degrees of freedom that can only be described using classical mechanics. In contrast, the chromophore subsystem experiences quantum transitions that necessitate a quantum mechanical description. Therefore, I employed a hybrid quantum mechanics-molecular mechanics (QM/MM) approach.

In this approach, the system is partitioned into two distinct regions: a quantum mechanical (QM) region and a molecular mechanics (MM) region. The total energy of the system is expressed as:

$$E = \langle \Psi | H_{QM} + H_{QM/MM} | \Psi \rangle + E_{MM}$$

The total energy in MM region $E_{MM}$ is independent of electronic degrees of freedom and is calculated classically using specified molecular force field. $H_{QM/MM}$ is determined through an electrostatic embedding scheme, which calculates the interaction between the point charges in the MM region and the nuclei and electrons in the QM region. When there is no covalent bond between the QM and MM regions, the $H_{QM/MM}$ comprises an electrostatic component and a Lennard-Jones term. If the QM and MM regions are connected by chemical bonds, the bond is truncated, and a link atom is introduced along the bond vector to cap each region.

$H_{QM}$ is evaluated using specified electronic structure calculation. In general, electronic structure calculation derives approximate solutions to the Schrödinger equation to obtain the total energy and the wavefunction (or electron density distribution) of the system. The Born-Oppenheimer approximation segregates the electron wavefunction from the nuclei wavefunction. For a fixed nuclei arrangement, the electronic ground state energy can be determined from the electron density ρ(r) based on density functional theory (DFT). However, the specific form of the density functional is generally unknown, so density functional approximations is used to estimate the ground state energy. The density $\rho(r)$ can be represented by a sum of squared molecular orbitals $\{\varphi_i\}$:

$$\rho(r) = \sum_{i}^{n} {\varphi_i}^2(r)$$

The molecular orbital $\varphi_i$ can be expressed as combinations of the basis $\{\chi_j\}$ with molecular orbital coefficients $c_{ij}$:

$$\varphi_i = \sum_j c_{ij}\chi_j$$

The basis functions $\{\chi_j\}$ typically resemble each of the occupied atomic orbitals, collectively referred to as the basis set. The coefficients $c_{ij}$ determine the weight of each basis function in a particular molecular orbital $\varphi_i$. $c_{ij}$ are obtained through a self-consistent field (SCF) method, where energy is iteratively calculated, and coefficients are updated until convergence is achieved.

Upon convergence of the energy, the forces acting on the atoms can be obtained by taking the derivative of the energy with respect to the positions of the QM and MM atoms. QM computations usually scale as $O(N^3)$ or worse, so QM/MM balances the strengths of QM (accuracy) and MM (speed).

## Unsupervised Learning in Chemistry Simulation

Unsupervised learning can uncover important conformations and structural transitions within large molecular dynamics (MD) simulation data. Given the high-dimensional 3D time series data, we use dimensionality reduction techniques to simplify and visualize the data while preserving important relationships. Features can be extracted from a MD data set, such as dihedral angles or pairwise contact distances of a protein. Time-structure independent components analysis (tICA) is a dimensionality reduction technique that extracts high-autocorrelation linear combinations of the input features, which represent the slowest-relaxing degrees of freedom in the system while maximizing decorrelation time. Once the dimensionality is reduced with tICA, clustering algorithms can be applied to group structures into distinct conformations. *k*-centers clustering algorithm selects a predefined number of cluster centers (k) that are representative of the data and assigns each data point to the nearest cluster center. From

the clustering results, we can study stable or frequently sampled states, their properties and dynamics, and transitions between states.

Based on the tICA and clustering result, the analysis can be extended by building a Markov State Model. The eigenvectors associated with the slowest processes in the model contain information about the configurational changes occurring on different timescales. Free energy landscape can be calculated by re-weighting the trajectory frames with stationary distribution. Each microstate can be assigned to macrostates using the PCCA++ algorithm. The representative molecular structures for each macrostate can be visualized for further inspection. We can also compute stationary as well as dynamic experimental observables to compare the model to experimental data.

### Time-Dependent Density Functional Theory (TDDFT)

The absorption spectrum is computed from the trajectory in order to benchmark the simulation with experimental measurements. The vertical absorption energy is calculated by subtracting the ground state electronic energy from that of the excited state at a fixed nuclear configuration, because the nuclear configuration stays the same during the fast electron excitation process. While the ground state energy is calculated using the QM method described earlier, the excitation energy can be obtained through the excited-electronic structure methods, including configuration interaction (CI), perturbation theory, or equation of motion coupled-cluster (EOM-CC). This process can be repeated to predict the complete evolution of the system. However, wave-function-based excited-state methods are computationally demanding due to electron correlation term. A more practical approach is time-dependent density functional theory (TDDFT), an extension of density-functional theory (DFT) that uniquely maps the (time-dependent) wave function to the (time-dependent) electronic density [29]. TDDFT calculates excitation energies in response to a time-varying applied electric field.

From the TDDFT energies and oscillatory strengths, the absorption spectrum is generated through Gaussian broadening. Each excitation energy, E, associated with an oscillation strength (amplitude) A, contributes to the absorption intensity, I, on the spectrum at $E_i$ in a Gaussian-like function:

$$I = \sum_{E_i=0}^{6} A * e^{-(\frac{(E - E_i)}{\sigma})^2}.$$

σ controls the width of the distribution, with larger values resulting in broader distributions. The contributions from each peak are cumulatively summed to obtain the overall spectral distribution.

## Computational Details

The protein structures were parametrized, relaxed with classical molecular dynamics, and simulated with hybrid QM/MM simulation (Fig. 13).

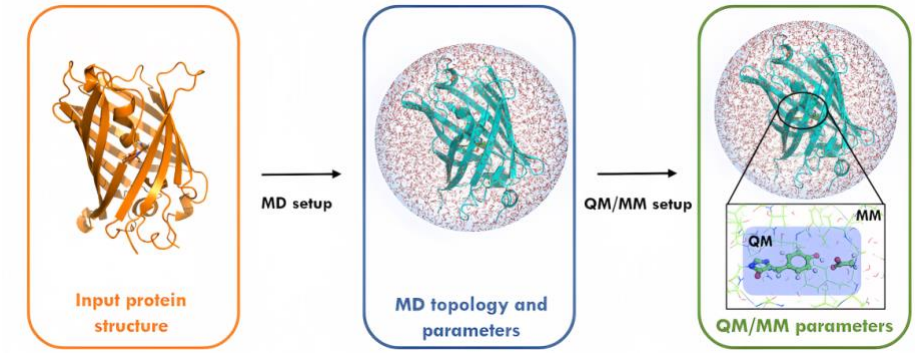



*Fig. 13 Hybrid quantum mechanics-molecular mechanics (QM/MM) workflow [30].*

## Protein Preparation

Initial protein structures were obtained from Protein Data Bank (PDB) – 1huy for citrine, 1emg for GFP, and 4eul for eGFP – all of which share the same chromophore structure (CRO) (**Table 1**). Extra ligands and ions were removed in Pymol. Water molecules were removed for the implicit solvent model but preserved for the explicit solvent model. Since the molecular force field does not contain parameters for CRO, its partial charges and atom types were computed in *antechamber* with the *bcc* charge scheme. The CRO template was obtained from the *Chemical Component Dictionary (CCD).* Prepgen was used to make an amino-acid-like molecule that be connected to other parts of the protein at its termini. The

covalent parameters (bonds, angles, and dihedrals) were computed with *parmchk2* using *parm10.dat file*, supplemented with GAFF (generalized Amber force field) database [31].

The protein was parameterized using the Amber ff14SB force field. For the implicit model, the generalized Born solvent model was used with the default radius set (PBRadii) as "mbondi3." For the explicit solvent model, the protein was solvated in a rectangular box with a 12 Å solvent on all sides using the TIP3P model. Counter ions (Cl- and Na+) were added to neutralize the protein charge.

*Table 1. Amino acid mutations in GFP mutants.*

| | ***63*** | ***62*** | ***64*** | ***65*** | ***68*** | ***76*** | ***199*** |
|---|---|---|---|---|---|---|---|
| *Citrine (1huy)* | CRO | F | L | M | A | R | Y |
| *GFP (1emg)* | CRO | F | V | Q | S | R | T |
| *eGFP (4eul)* | CRO | L | V | Q | S | Q | T |

## Molecular Dynamics (MD)

For the implicit solvent model, the system was first minimized with 100 steps of the steepest gradient descent. Then, the system was heated in the NVE ensemble from 10 K to 300 K over the course of 200 ps with a 2 fs time step. At the end, the system was relaxed at 300 K for 50 ns using a 2 fs time step. No periodic boundaries were imposed, and the Particle Mesh Ewald (PME) method was turned off. A 1000 Å cutoff distance was set for nonbonded interactions.

For the explicit solvent model, the solvent was initially minimized with the protein structure restrained, followed by a minimization with the solvent molecules fixed, and finally with the entire system unrestrained. Each minimization run consisted of 10000 steps of steepest gradient descent followed by 10000 steps of conjugate gradient descent. After minimization, the system was heated in NVT ensemble from 0K to 100K over 1000 ps with a 1 fs timestep and then in NPT ensemble from 100K

to 310K over 1000 ps with 1 fs timestep. The system was equilibrated in NPT ensemble at 310K for an additional 5 ns with 2 fs timestep. Temperature was controlled using Langevin dynamics with a collision frequency of 5 ps−1 during heating and 2 ps−1 in production simulations. Pressure is controlled using Monte Carlo barostat. Periodic boundary conditions were applied to the system throughout heating and equilibration. Electrostatics were calculated using the Particle Mesh Ewald (PME) method, with a distance cutoff of 9.0 Å.

### Quantum Mechanical−Molecular Mechanical (QM/MM)

From the molecular runs, four frames with 1 ns in between were used to setup the QM/MM simulation. For two solvent model and three GFP mutants, four independent trajectories were ran for 30 ps with 0.2 fs time step. The MM region was run with similar setup as the above MD runs. The QM region is set with zero net charge and a spin multiplicity of 1, indicating a closed-shell configuration. The basis set was 6-31G*, encompassing both valence and core electrons. The electronic structure calculation uses the global hybrid density functional, B3LYP. Dispersion interactions were added through Grimme's D3 correction. The self-consistent field (SCF) calculation uses the Direct Inversion in the Iterative Subspace (DIIS) algorithm, with a maximum iteration limit of 200.

### Simulation Analysis

Python script was developed to watch the total energy and the HOMO-LUMO energy gap calculated in each QM/MM step. The dephasing time was calculated with $B = \sqrt{\frac{12}{5}}\hbar$.

Mass-weighted root-mean-square-deviation (RMSD) was used to measure the structure flexibility of the whole chromophore or the protein nanocylinder during simulation. Root Mean Square Fluctuation (RMSF) was used to measure the structure flexibility of each atom or each amino acid. The native contacts between chromophore and protein are tracked by recording any

protein atoms nearby the chromophore within 5 A cut-off in each frame. I calculated the the total fraction of frames the contact is present, the average distance of the contact when present, and the standard deviation of the contact distance when present. Relative contact strengths are normalized and a PDB is generated with relative contact strength in B factor column and visualized with Chimera. The radial distribution function (RDF, aka pair correlation function) is calculated from the histogram of the number of water particles found as a function of distance to the chromophore.

The 3D coordinates data is transformed to distances between CRO and other proteins, resulting in 223 features. To simplify and visualize the data, Time-Lagged Independent Component Analysis (TICA) is used. TICA aims to map the high-dimensional input space into a lower-dimensional space that preserves 95% of the kinetic variance, with the use of a lag time of 5. The TICA coordinates are clustered into discrete states using the k-means algorithm with 200 clusters. The distribution of these states is visualized in the low-dimensional TICA subspace.

### Time-Dependent Density Functional Theory (TDDFT)

Time-dependent density functional theory (TDDFT) is used to calculate the absorption spectra on 40 selected frames from the QM/MM simulation with the QM xyz coordinate and MM point charge. The QM region is set with zero net charge and a spin multiplicity of 1. Both the wPBEh method and the B3LYP method were used with lacvps_ecp basis set. SCF convergence methods was implemented with DIIS with a convergence threshold of 3.0e-5 and tolerance for the X-matrix of 1.0e-6. Configuration interaction singles (CIS) calculations was performed for 30 excited states with a spin multiplicity of 1. The absorption spectrum is calculated by Gaussian broadening with $\sigma$ of 0.5.

## Results and Discussion

### Simulation Overview

To assess the accuracy of MM and QM models, I analyzed thermodynamic parameters obtained from the simulations. The total energy and temperature were recorded by Amber. The MM production simulations showed steady energy levels and minimal temperature fluctuations, which indicates that the MM models have achieved equilibrium state (**Fig. 14 a**). The eGFP system exhibited higher energy levels compared to GFP and citrine (**Fig. 14 b**). In the QMMM production simulations, the initial few picoseconds displayed a decrease in total energy and significant temperature fluctuations (**Fig. 14 b**). This indicates that the system was unstable when transitioning from MM to QMMM due to changes in force fields. However, the system was stablized after 5 ps, as evident from the consistent energy levels (**Fig. 14 b**). eGFP still exhibited higher energy levels compared to the other two systems (**Fig. 14 b**).

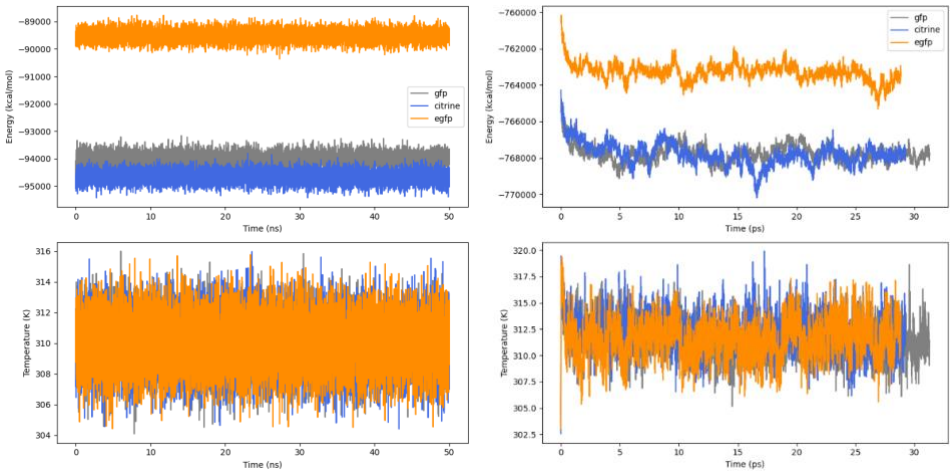



*Fig. 14 Total energy and temperature over time*. (a) results from MM simulation (b) results from hybrid QMMM simulation

Fig. 15 displays the total energies of the QM region, comprising solely the chromophore, along with the HOMO-LUMO energy gap obtained from Terachem. In two separate implicit models initialized from different MM frames (Fig. 15a, b), we observed consistent patterns in both total energy and HOMO-LUMO gap, confirming the reproducibility of the findings (**Fig 15**

**a, b**). The QM energy was very high but rapidly decreased within the first few picoseconds, which again indicated that the system was unstable first (**Fig 15 a, b**). However, the energy stabilized at a lower level after 5 ps, signifying stabilization. In the implicit models, eGFP displayed higher total QM energy compared to GFP and citrine, consistent with the MM results (**Fig 15 a, b**).

Explicit models were run for a longer time to ensure that the system is equilibrated with the solvents. Two consecutive 4 ps trajectories (20-24ps and 24-28ps) were then analyzed (**Fig 15 c, d**). Explicit models are higher in energy (~-1072) than the implicit models (~-1160). This is because explicit models account for the energy cost of the solvent-solute interactions via point charges representing water molecules. Different from the trends observed in implicit models and MM productions, citrine is slightly higher in energy than the other two systems.

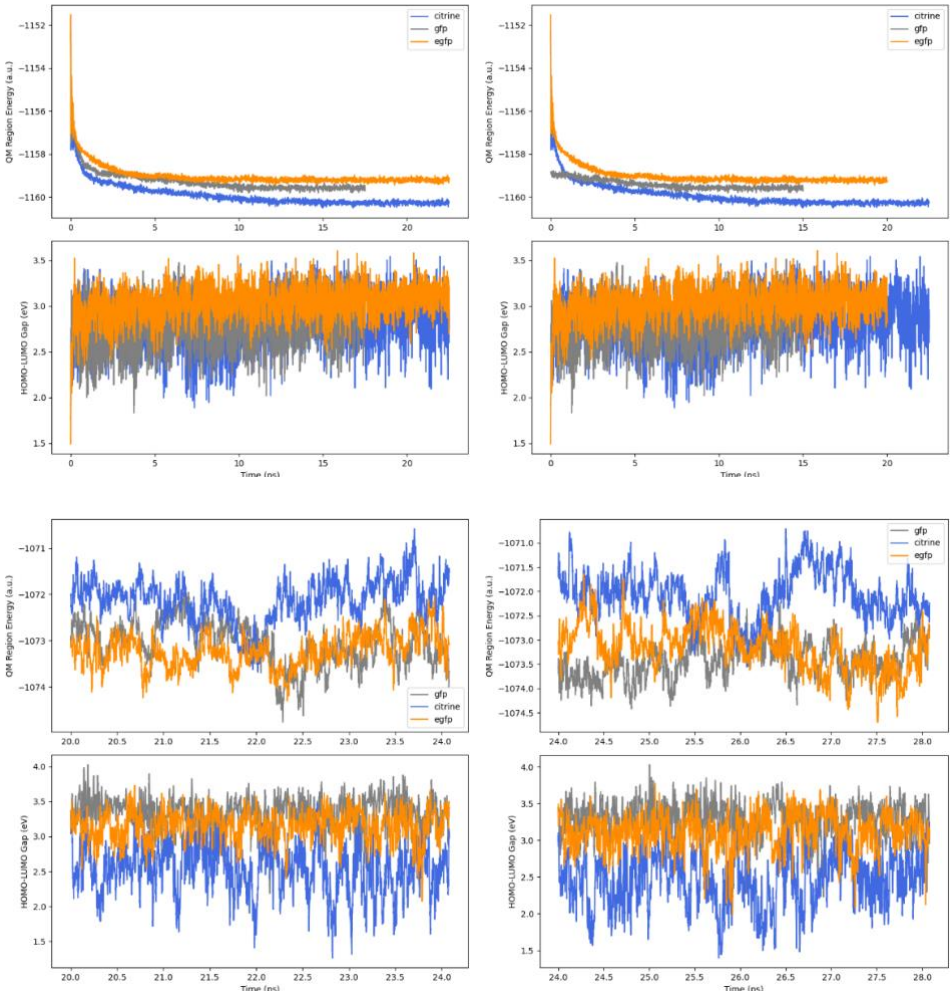


*Fig. 15 QM region energy and HOMO-LUMO gap over time.* (a-b) results from two separate implicit models (c) 20-24ps simulation in the explicit model (d) 24-28ps simulation in the same explicit model.

Quantum coherence can also be visualized with autocorrelation function, a measure of the degree of similarity between the HOMO-LUMO gap series and a lagged version of itself over successive time intervals. Different trajectories within each model display similar trends, underscoring the reproducibility of the results (**Fig 16**). In general, explicit models were less auto-correlated than the implicit models (**Fig 16**). This means that the HOMO-LUMO gap energies in explicit models tend to fluctuate more over time due to dynamic interactions with the solvent.

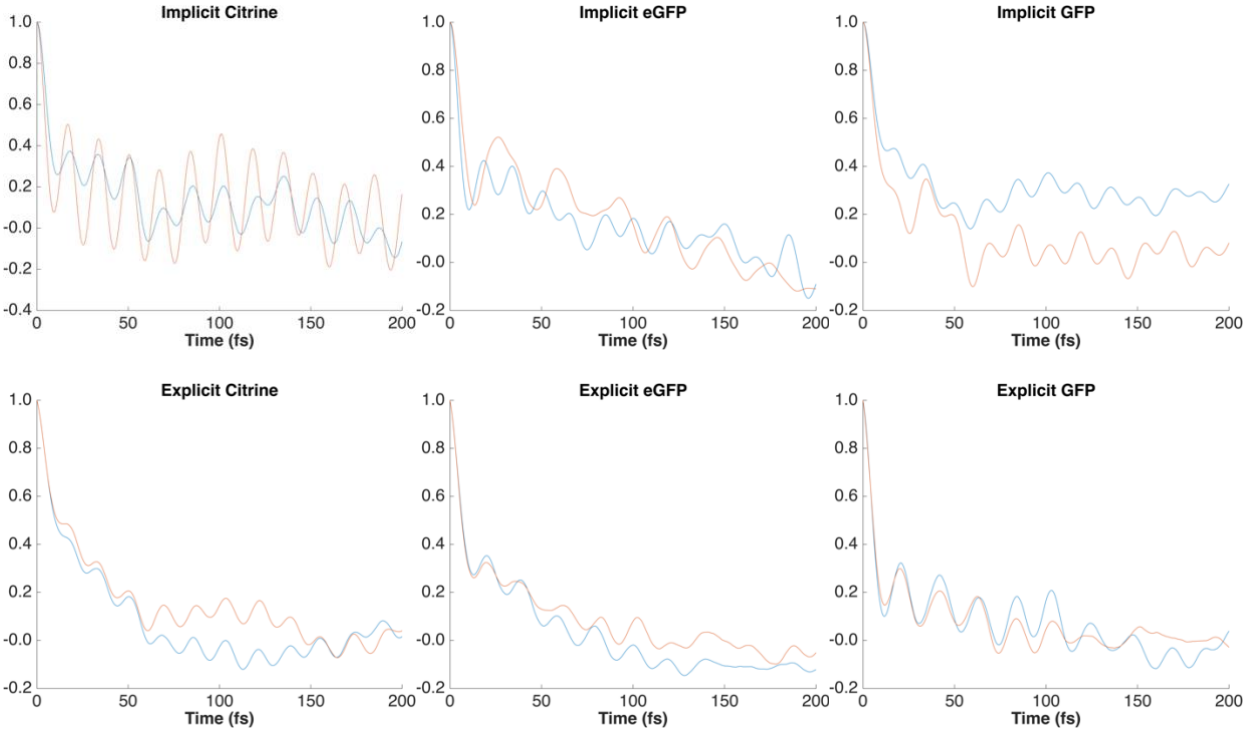



*Fig. 16 Autocorrelation function of HOMO-LUMO gap.* The x-axis displays the number of lags and the y-axis displays the autocorrelation at that number of lags.

## Larger Energy Gap Fluctuation in Explicit Solvent Model and Citrine

All means and standard deviations were calculated using 4 ps long trajectories to ensure a fair comparison. Notably, the HOMO-LUMO energies of GFP were the lowest in implicit models but became the highest in the explicit model (**Table 2**). This shift indicates that GFP interacts more profoundly with the solvent, stabilizing the HOMO state of the chromophore. Conversely, the HOMO-LUMO energies of citrine decrease in explicit models, suggesting lesser interaction with the solvent (**Table 2**).

Furthermore, the explicit condition displays higher standard deviations compared to the implicit model overall (**Fig 16, Table 2**). These higher standard deviations reflect additional sources of variability arising from dynamic and unpredictable solvent-solute interactions. Citrine consistently exhibits the highest standard deviation of HOMO-LUMO energies compared to n to eGFP and GFP across all three conditions (**Fig 15 c, d, Table 2**). However, citrine also has the smallest HOMO-LUMO gap energies. Such a combination of smaller gap energies and higher standard deviation left a question: Does a larger HOMO-LUMO gap result in less variability due to the energy surface distance and better alignment between the excited state and the ground state? In contrast, eGFP exhibits the lowest standard deviation, except in the last condition (**Fig 16, Table 2**). Thus, some of the mutations in citrine lead to higher energy gap fluctuation, while that in eGFP leads to lower energy gap fluctuation. The larger the HOMO-LUMO fluctuation, the smaller the estimated decoherence time. Thus, explicit models had shorter decoherence time compared to implicit models (**Table 2**). And citrine system had shorter decoherence time relative to eGFP and GFP.

*Table 2. Mean values and standard deviations of HOMO-LUMO gaps, as well as estimated decoherence times.*

| | *Implicit(16-20ps)* | | | *Explicit (20-24ps)* | | | *Explicit (24-28ps)* | | |
|---|---|---|---|---|---|---|---|---|---|
| | *Citrine* | *Egfp* | *gfp* | *Citrine* | *Egfp* | *gfp* | *Citrine* | *Egfp* | *gfp* |
| *Mean (eV)* | 2.896 | 3.069 | 2.741 | 2.536 | 3.159 | 3.334 | 2.464 | 3.117 | 3.291 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Standard Deviation (eV)* | 0.224 | 0.163 | 0.208 | 0.350 | 0.206 | 0.224 | 0.356 | 0.236 | 0.231 |
| *Decoherence Time (fs)* | 2.531 | 2.967 | 2.627 | 2.025 | 2.641 | 2.533 | 2.007 | 2.466 | 2.492 |

### Correlation between Structural Dynamics and Decoherence

To investigate the structural mechanisms underlying the observed trend—faster decoherence in citrine and slower decoherence in eGFP—I conducted structural analyses of conformational dynamics, interactions between the chromophore and the protein, and water density. Additionally, I performed dimensionality reduction and conformation clustering to identify crucial conformations and transitions within the simulation.

Root mean square deviation (RMSD) calculates the deviation of a structure to a reference, with RMSD=0.0 indicating a perfect overlap. Here, I measured the RMSD with respect to the starting frame. The RMSD values for the chromophore and the surrounding protein were separated and normalized by the number of atoms in each region. The RMSD were large for all systems initially; upon reaching a stable state, no significant conformational changes were observed (**Fig 17 a, b**). The chromophore has lower RMSD compared to the surrounding protein (**Fig 17 a, b**). The chromophore changed less over the simulation because it was constrained by steric factors and interactions with surrounding proteins. The RMSD can't be directly compared across the system with varying initial conditions.

The trajectories were partitioned into 2.5ps windows. For each window, the average chromophore RMSD and protein RMSD were calculated with respect to the starting frame of the window and plotted against the energy gap fluctuation (**Fig 17 c, d**). Both chromophore RMSD and protein RMSD correlate with energy gap fluctuation, highlighting the influence of structural dynamics on quantum decoherence (**Fig 17 c, d**). Within each 2.5ps window, eGFP exhibits smaller chromophore and protein RMSD, which indicate a more stable structure, likely due to specific interactions and steric constraints (**Fig 17 c, d**). This reduced structural dynamicss in eGFP may lead to a more stable

electronic structure and lower quantum decoherence effects, as evident in its lower HOMO-LUMO gap fluctuation compared to GFP and citrine (**Fig 17 c, d**).

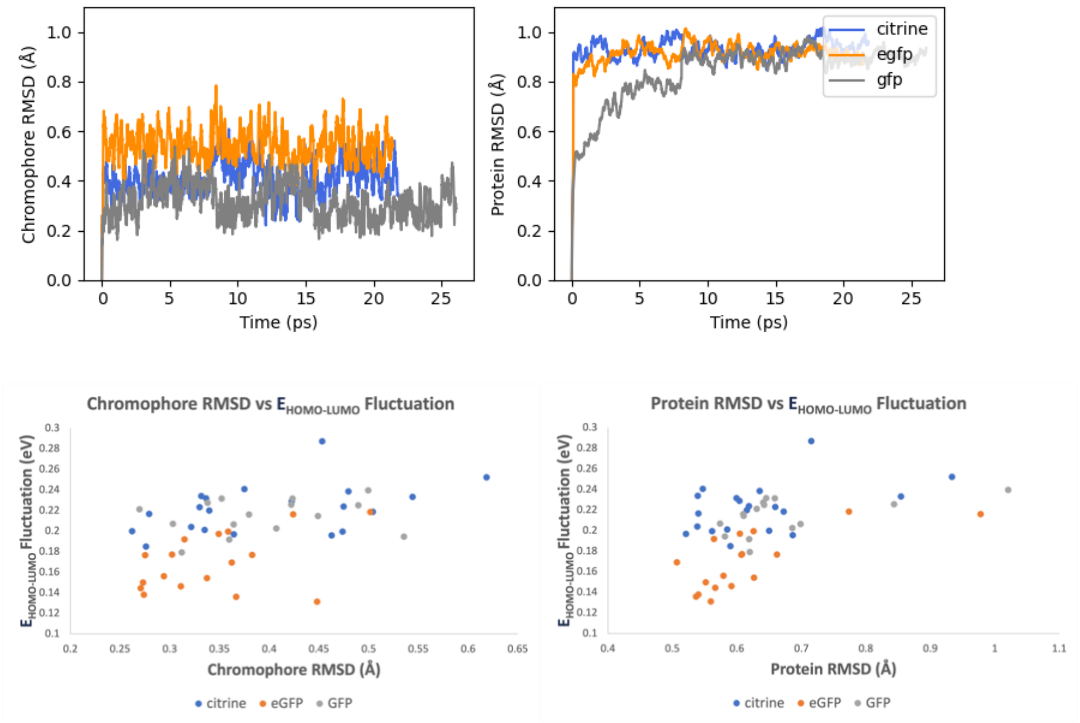



*Fig. 17 Root mean square deviation (RMSD) analysis.* a) Chromophore RMSD b) Protein RMSD c) Chromophore RMSD vs Energy Fluctuation d) Protein RMSD vs Energy Fluctuation

## Higher Structural Dynamics of Individual Residues in Explicit Solvent Models

Root Mean Square Fluctuation (RMSF) calculates the structural dynamics of individual residues throughout the simulation. All three systems display similar RMSF patterns (**Fig 18**). Notably, the explicit water models have higher RMSF values compared to the implicit models, indicating that, on average, residues are more dynamic due to their interactions with water (**Fig 18**).

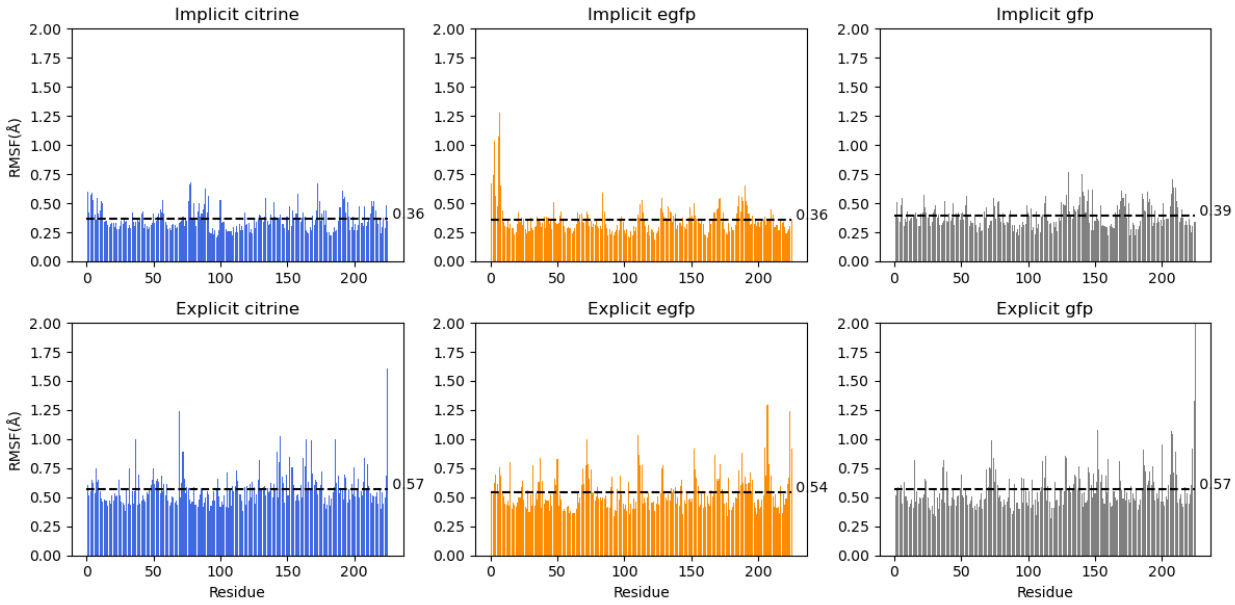


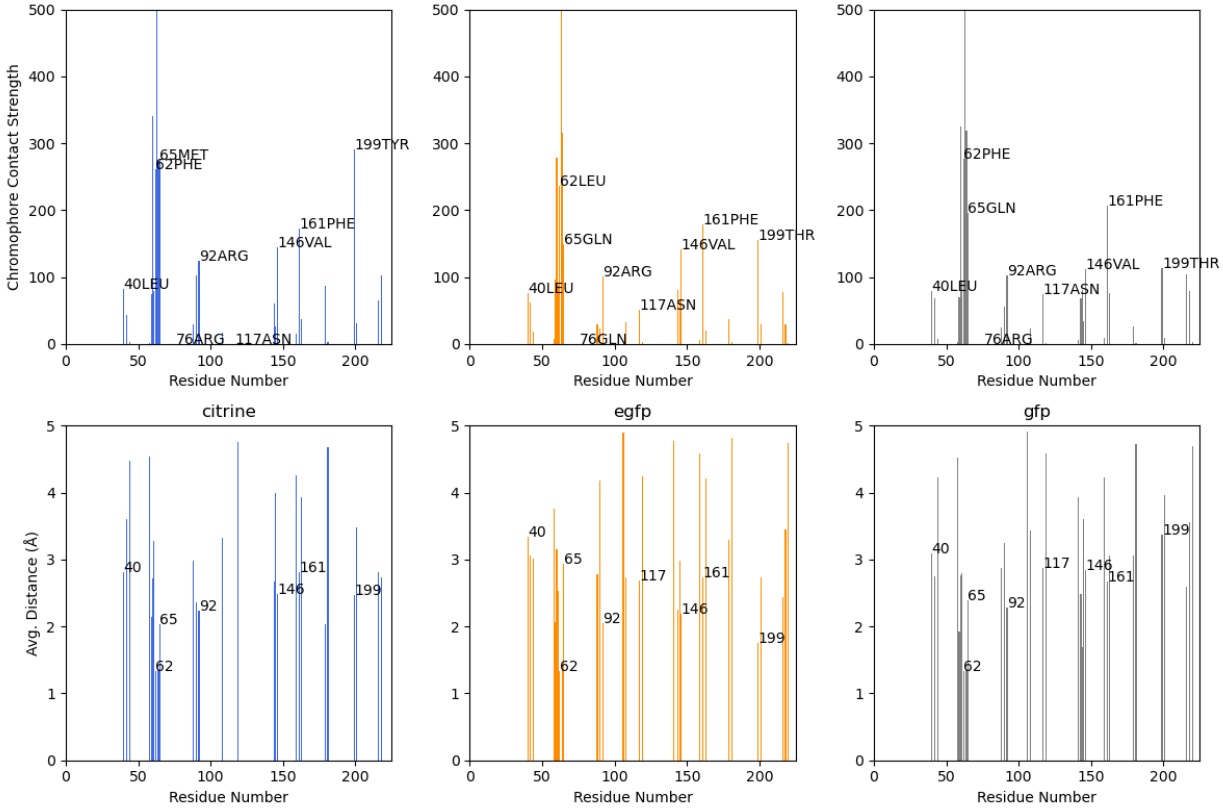

*Fig 18. Root Mean Square Fluctuation (RMSF) analysis.*

## Q65M and T199Y in Citrine Increase Protein-Chromophore Contact

The interactions between the chromophore and the protein were evaluated using the contact strength metric, as well as the closest distance between their atoms (**Fig 19, Table 3**). Citrine exhibits high contact strength at Q65M and T199Y, and the Q65M has a smaller distance to the chromophore (**Fig 19 a, b**). The mutated methionine is larger than the original glutamine, which brings closer to the chromophore, potentially resulting in steric clashes or hydrophobic interactions. (**Fig 19 c**). The mutated tyrosine engages in a $\pi$–$\pi$ interaction with the aromatic ring of the chromophore, and its hydroxyl group can establish polar interactions with the mutated methionine (**Fig 19 c**). The peak corresponding to N117 is prominent in eGFP and GFP but is absent in citrine (**Fig a, b**). Notably, in citrine, N117 is oriented away from the chromophore, potentially driven by unfavorable interactions with water as an allosteric effect of the mutations (**Fig 19 c**). These heightened interactions and potential steric clashes provide an explanation for the larger energy gap fluctuations and shorter decoherence time observed in citrine.

On the other hand, the contact strength pattern of eGFP closely resembles that of GFP (**Fig 19 a**). Within eGFP, the F62L mutation is directly linked to the chromophore, while the Q72R mutation is positioned far from the binding site (**Fig 19 c**). As a result, neither of eGFP's two mutations makes additional contacts with the chromophore. In addition, the standard deviation of eGFP's contact distance is smaller compared to the other two systems (**Table 3**).

Overall, residues with high contacts with the chromophore display low RMSF (**Fig 18, Fig 19**). This is because they are oriented inward, interacting with the chromophore, rather than outward, where they would interact with the solvent.



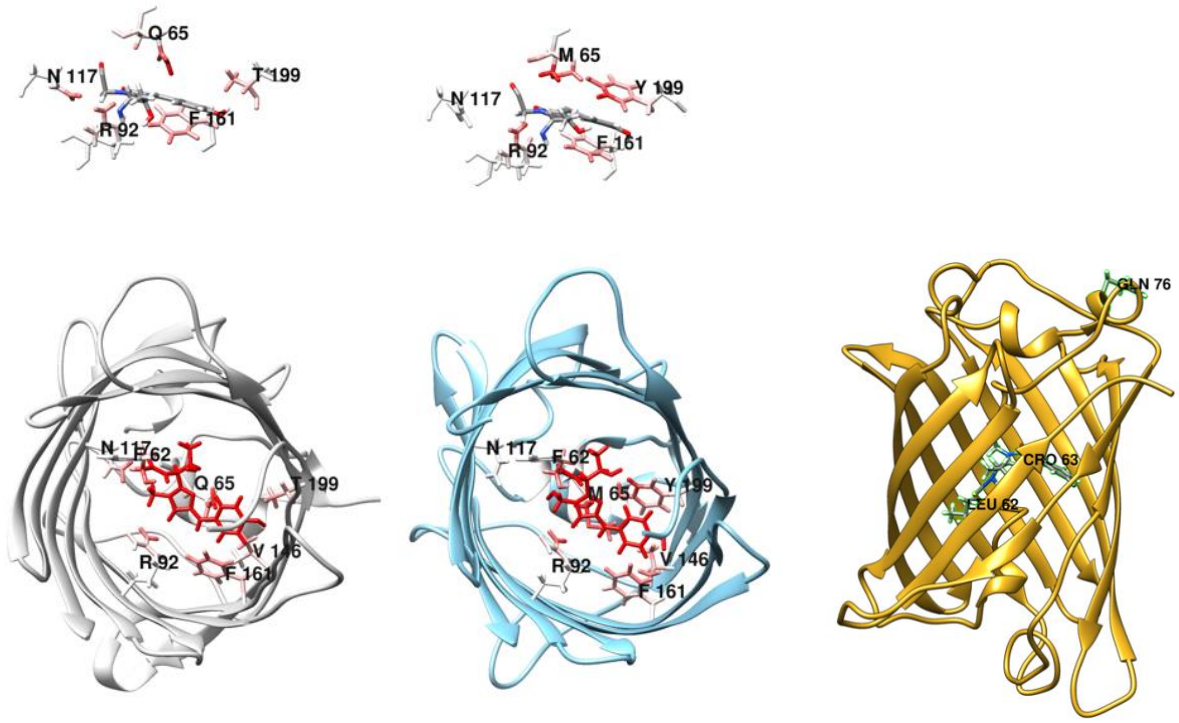



*Fig 19. Native contacts between the protein and the chromophore*. a) contact strength, b) average distance, c) structure illustrations

*Table 3. Mean values and standard deviations of contact distance between chromophore and the protein at selected residues.*

| | *Mean distance* | | | *Standard deviation* | | |
|---|---|---|---|---|---|---|
| *Res* | *citrine* | *egfp* | *gfp* | *citrine* | *egfp* | *gfp* |
| *40* | 2.81 | 3.34 | 3.09 | 0.42 | 0.30 | 0.46 |
| *62* | 1.34 | 1.33 | 1.33 | 0.02 | 0.03 | 0.02 |
| *65* | 2.04 | 2.94 | 2.39 | 0.21 | 0.31 | 0.25 |
| *92* | 2.24 | 2.05 | 2.28 | 0.27 | 0.23 | 0.18 |
| *117* | | 2.69 | 2.87 | | 0.34 | 0.22 |
| *146* | 2.49 | 2.19 | 2.84 | 0.41 | 0.15 | 0.42 |
| *161* | 2.81 | 2.73 | 2.68 | 0.22 | 0.18 | 0.30 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *199* | 2.47 | 1.75 | 3.37 | 0.29 | 0.15 | 0.39 |
| *Avg* | 2.31 | 2.38 | 2.61 | 0.26 | 0.21 | 0.28 |

## Greater Water Density Around GFP

The radial distribution function was used to assess the water density around the chromophore. It is evident that there is a greater concentration of water molecules near GFP than eGFP and citrine (**Fig 20**). This observation explain the higher decoherence observed in GFP compared to eGFP, due to the dynamics water interaction.

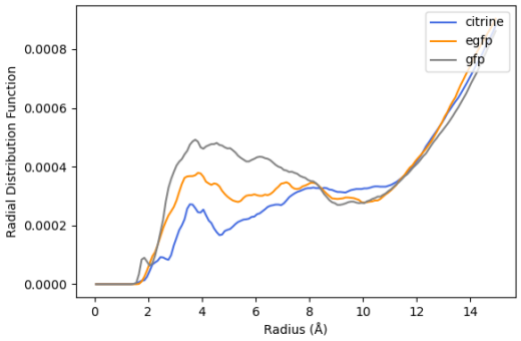



*Fig. 20 Solvent radial distribution function.*

## Dimensionality Reduction and Clustering Analysis of Simulations

To extract important structures and dynamics from the simulation, I used tICA to reduce the dimensionality of the data and K-means clustering to group various frames into macrostates. The first few tICA components represent the slowest motions in the simulation. The resulting macrostates and their transitions can be further examined in relation to their HOMO-LUMO energy fluctuations (**Fig 21**).

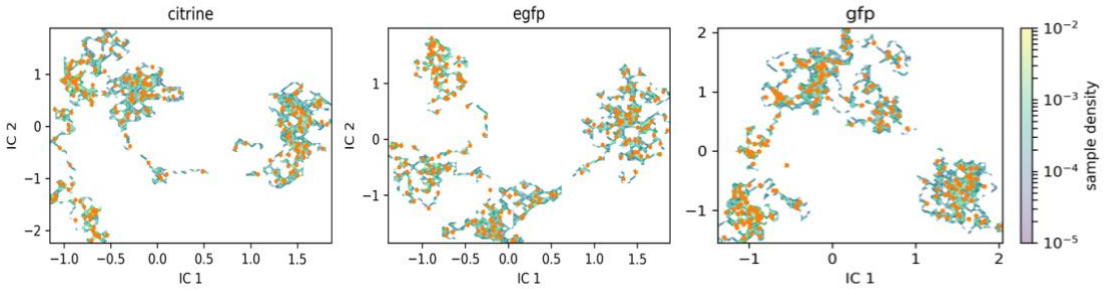


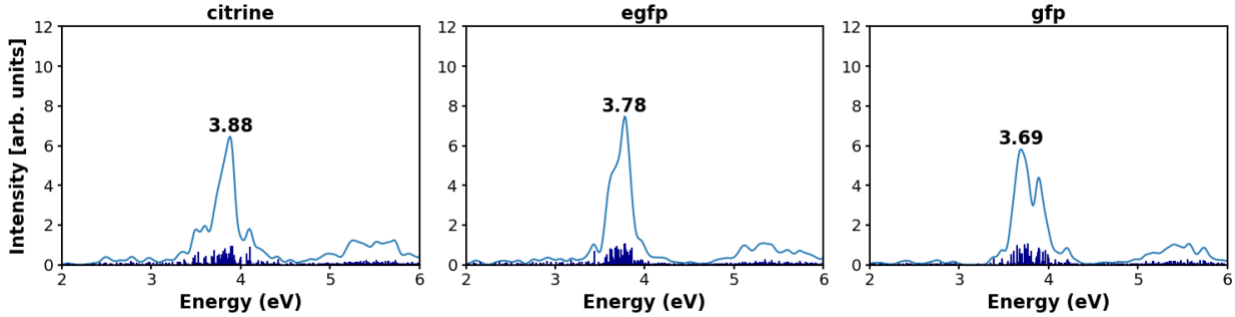
*Fig 21. tICA and clustering analysis of the simulations.*

## Mismatch between TDDFT Results and Experimental Absorption Spectrum

To validate and refine the computational model, we compared the absorption spectrum calculated by TDDFT to existing experimental observations. Previous experiments and computational simulations have consistently shown that the absorption peak of GFP is higher in energy than that of eGFP and citrine (**Table 4**). In the implicit model, we found that citrine peak was higher in energy than eGFP and GFP, which is opposite to experimental observations. In explicit model, we found that eGFP absorbs higher energy than citrine than GFP. The order between citrine and eGFP is correctly predicted since explicit water molecules conjugate with the chromophore. However, the model for GFP still does not match the experimental results. The TDDFT results remain consistent when employing different QM methods, such as B3LYP and ***ω***PBEH, and considering various QM regions, such as the chromophore and atoms within a 5 Å radius of the chromophore. This suggests that while the TDDFT results are reproducible, the simulation fails to accurately reproduce experimental observations.

Table 4. Experimental absorption spectrum. 32

| | *citrine* | *egfp* | *gfp* |
|---|---|---|---|
| *absorption wavelength (nm)* | 516 | 488 | 396,475 |
| *photon energy (eV)* | 2.40 | 2.54 | 3.13, 2.61 |



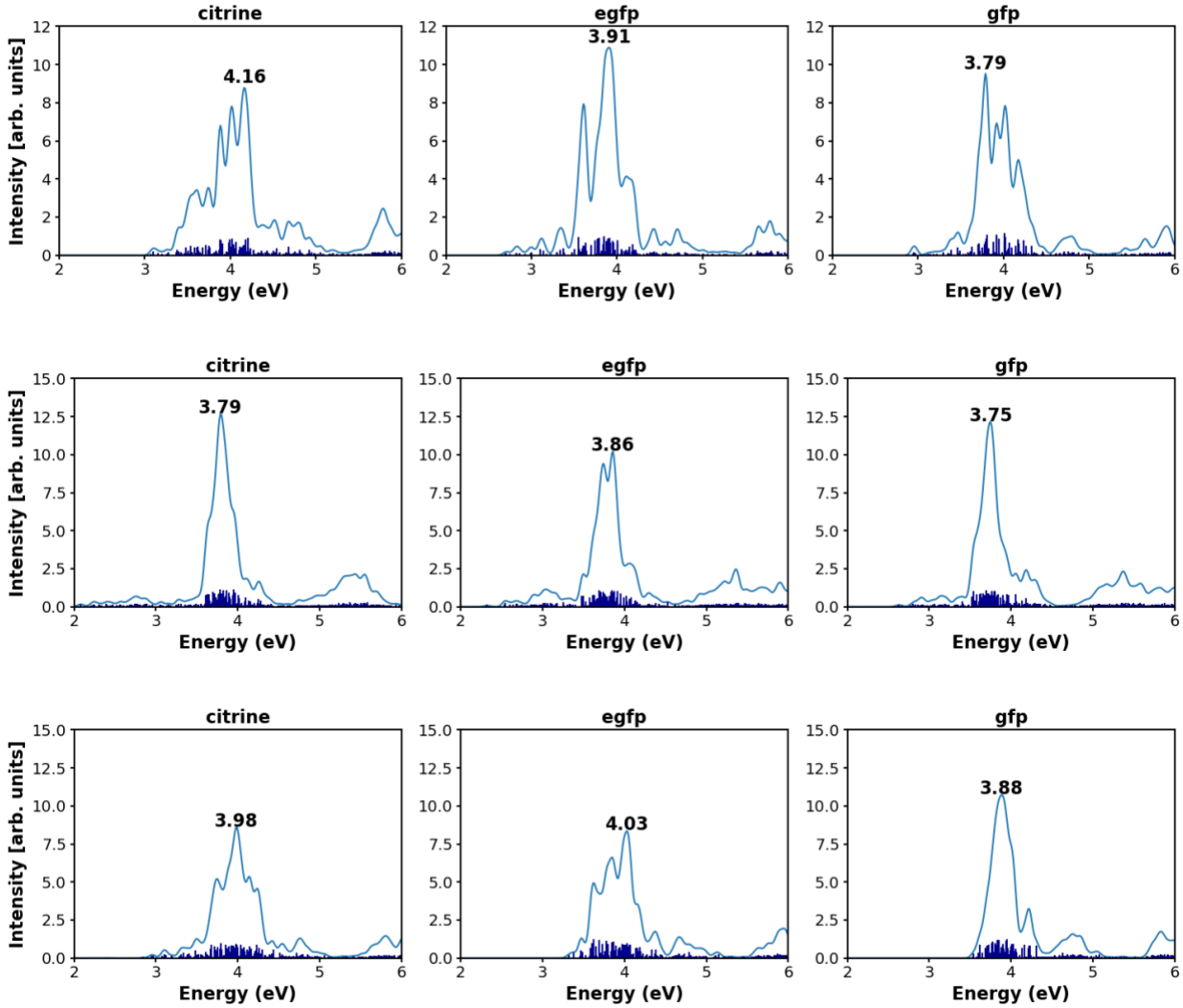



*Fig. 22 TDDFT modeled absorption spectrum.* (a) B3LYP implicit solvent (b) $\omega$PBEH implicit solvent (c) B3LYP explicit solvent (d) $\omega$PBEH explicit solvent

## Conclusion

This study contributes to the understanding of quantum decoherence in fluorescent proteins and establishes a knowledge base for the rational design of quantum batteries. Quantum batteries, rooted in the principles of quantum mechanics, stand as an emerging and transformative technology in the field of energy storage and retrieval. Unlike conventional batteries, quantum batteries exhibit superextensive charging, which allows them to charge rapidly as their capacity increases. Acknowledging limitations in a previous prototype employing an organic microcavity with polaritons as "qubits," attributed to exciton-

exciton annihilation, we proposed replacing organic dyes in the microcavity with fluorescent proteins. Fluorescent proteins can utilize their protein cylinder as a "molecular bumper" to effectively mitigate annihilation, enabling them to operate efficiently even at high concentrations. In our research, we delve into the impact of protein and solvent environments on energy gap fluctuations. Our hypothesis suggests that targeted mutations, reducing interactions and collisions, hold the potential to delay quantum decoherence.

We conducted a comprehensive study on three distinct GFP variants—citrine, eGFP, and GFP—using hybrid quantum mechanics-molecular mechanics (QMMM) simulations with both explicit and implicit models. Surprisingly, citrine, despite having the smallest energy gap, exhibits the largest gap fluctuation and the shortest estimated decoherence time. Mutations, such as Q65M and T199Y, introduce π-π interactions and steric clashes with the chromophore, modifying the energy surface of the ground or excited state and resulting in less correlated energy level changes. Furthermore, the explicit models demonstrate shorter decoherence times. Notably, the wild-type GFP, characterized by higher water density around the chromophore, also displays a shorter decoherence time compared to eGFP. Solvent interactions couple chromophore excitation with other vibrational modes, introducing small perturbations in the potential energy surfaces. Therefore, it is advisable to avoid bulkier amino acids and π-π interactions with the chromophore while minimizing solvent exposure in the development of fluorescent proteins for quantum batteries.

Additionally, we show that structural dynamics metrics like RMSD and RMSF are correlated with energy gap fluctuations, so they can be valuable metrics for future rational design to minimize the quantum decoherence. In the next stages of our research, we aim to refine the model to fit the experimental absorption spectrum, accounting for factors such as temperature, pH, and the presence of cofactors or other molecules. We also plan to enhance force field parameters, utilize more advanced quantum mechanical methods, and extend simulations to capture additional conformational states. Based on the tICA and clustering result, we can build a Markov State Model to compute stationary distribution, free energy landscape and other experimental observables. Looking ahead, we plan to simulate the

behavior of fluorescent proteins within the microcavity, considering the intricate interactions between light and matter. This research not only provides a fundamental understanding of quantum decoherence in fluorescent proteins but also establishes a knowledge base for the rational design of quantum batteries.

References

1. *Understanding Batteries*. (Royal Society of Chemistry, 2007). doi:10.1039/9781847552228.

2. Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 467–488 (1982).

3. Alicki, R. & Fannes, M. Entanglement boost for extractable work from ensembles of quantum batteries. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **87**, 042123 (2013).

4. Campaioli, F. *et al.* Enhancing the charging power of quantum batteries. *Phys. Rev. Lett.* **118**, 150601 (2017).

5. Quach, J. Q. & Munro, W. J. Using dark states to charge and stabilize open quantum batteries. *Phys. Rev. Appl.* **14**, 024092 (2020).

6. Ferraro, D., Campisi, M., Andolina, G. M., Pellegrini, V. & Polini, M. High-Power Collective Charging of a Solid-State Quantum Battery. *Phys. Rev. Lett.* **120**, 117702 (2018).

7. Binder, F. C., Vinjanampathy, S., Modi, K. & Goold, J. Quantacell: powerful charging of quantum batteries. *New J. Phys.* **17**, 075015 (2015).

8. Quach, J. Q. *et al.* Superabsorption in an organic microcavity: Toward a quantum battery. *Sci. Adv.* **8**, eabk3160 (2022).

9. Cuevas, Á. *et al.* First observation of the quantized exciton-polariton field and effect of interactions on a single polariton. *Sci. Adv.* **4**, eaao6814 (2018).

10. Gargoubi, H., Guillet, T., Jaziri, S., Balti, J. & Guizal, B. Polariton condensation threshold investigation through the numerical resolution of the generalized Gross-Pitaevskii equation. *Phys. Rev. E* **94**, 043310 (2016).

11. Okinawa Institute of Science and Technology. Shifting dimensions: Exciting excitons in phosphorene. https://phys.org/news/2020-03-shifting-dimensions-excitons-phosphorene.html.

12. Dietrich, C. P. *et al.* An exciton-polariton laser based on biologically produced fluorescent protein. *Sci. Adv.* **2**, e1600666 (2016).

13. Fassioli, F., Dinshaw, R., Arpin, P. C. & Scholes, G. D. Photosynthetic light harvesting: excitons and coherence. *J. R. Soc. Interface* **11**, 20130901 (2014).

14. Olaya-Castro, A., Lee, C. F., Olsen, F. F. & Johnson, N. F. Efficiency of energy transfer in a light-harvesting system under quantum coherence. *Phys. Rev. B* **78**, (2008).

15. Panitchayangkoon, G. *et al.* Direct evidence of quantum transport in photosynthetic light-harvesting complexes. *Proc Natl Acad Sci USA* **108**, 20908–20912 (2011).

16. Maity, S. *et al.* DFTB/MM Molecular Dynamics Simulations of the FMO Light-Harvesting Complex. *J. Phys. Chem. Lett.* **11**, 8660–8667 (2020).

17. Akimov, A. V. & Prezhdo, O. V. Persistent electronic coherence despite rapid loss of electron–nuclear correlation. *J. Phys. Chem. Lett.* **4**, 3857–3864 (2013).

18. Mallus, M. I. *et al.* Relation between Dephasing Time and Energy Gap Fluctuations in Biomolecular Systems. *J. Phys. Chem. Lett.* **7**, 1102–1108 (2016).

19. Mallus, M. I., Schallwig, M. & Kleinekathöfer, U. Relation between Vibrational Dephasing Time and Energy Gap Fluctuations. *J. Phys. Chem. B* **121**, 6471–6478 (2017).

20. Grabarek, D. & Andruniów, T. The Role of Hydrogen Bonds and Electrostatic Interactions in Enhancing Two-Photon Absorption in Green and Yellow Fluorescent Proteins. *ChemPhysChem* **23**, e202200003 (2022).

21. Amat, P. & Nifosì, R. Spectral "Fine" Tuning in Fluorescent Proteins: The Case of the GFP-Like Chromophore in the Anionic Protonation State. *J. Chem. Theory Comput.* **9**, 497–508 (2013).

22. Zurek, W. H. Pointer basis of quantum apparatus: Into what mixture does the wave packet collapse? *Phys. Rev. D* **24**, 1516–1525 (1981).

23. Shu, Y. & Truhlar, D. G. Decoherence and its role in electronically nonadiabatic dynamics. *J. Chem. Theory Comput.* **19**, 380–395 (2023).

24. Schmidt, J. R., Sundlass, N. & Skinner, J. L. Line shapes and photon echoes within a generalized Kubo model. *Chem. Phys. Lett.* **378**, 559–566 (2003).

25. Yan, B. & Zurek, W. H. Decoherence factor as a convolution: an interplay between a Gaussian and an exponential coherence loss. *New J. Phys.* **24**, 113029 (2022).

26. Ensembles — Statistical Physics Notes. https://statisticalphysics.leima.is/equilibrium/ensembles.html.

27. Onufriev, A. V. & Case, D. A. Generalized born implicit solvent models for biomolecules. *Annu. Rev. Biophys.* **48**, 275–296 (2019).

28. Katiyar, R. S. & Jha, P. K. Molecular simulations in drug delivery: Opportunities and challenges. *WIREs Comput Mol Sci* **8**, (2018).

29. Runge, E. & Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett* **52**, 997–1000 (1984).

30. QM/MM with GROMACS + CP2K. https://docs.bioexcel.eu/2021-04-22-qmmm-gromacs-cp2k/.

31. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).

32. Salk Institute for Biological Studies, La Jolla, California, USA. Table of Fluorescent Proteins. https://www.salk.edu/pdf/fluorochrome-table.html.

## Supplement Code

### Implicit solvent (generalized Born) minimization

```
&cntrl
imin=1, ntb=0, maxcyc=100, ntpr=10, cut=1000., igb=8,
/
```

### Implicit solvent heating

```
&cntrl
imin=0, irest=0, ntx=1,
ntpr=1000, ntwx=1000, nstlim=100000,
dt=0.002, ntt=3, tempi=10,
temp0=300, gamma_ln=1.0, ig=-1,
ntp=0, ntc=2, ntf=2, cut=1000,
ntb=0, igb=8, ioutfm=1, nmropt=1,
/
&wt
TYPE='TEMP0', ISTEP1=1, ISTEP2=100000,
VALUE1=10.0, VALUE2=300.0,
/
```

Implicit solvent molecular dynamics

```
 &cntrl
   imin=0, irest=1, ntx=5,
   ntpr=1000, ntwx=1000, ntwr = 1000
   nstlim=500000, dt=0.002,
   ntt=3, ig=-1, !Langevin thermostat
   tempi=300, temp0=300,
   gamma_ln=1.0, ntp=0,
   ntc=2, ntf=2, cut=1000,
   ntb=0, igb=8, ioutfm=1,
 /
```

Explicit solvent setup

```
source leaprc.protein.ff14SB
source leaprc.water.opc
loadAmberPrep cro.prepin
loadAmberParams frcmod2.cro
loadAmberParams frcmod1.cro
x = loadPDB ${protein}.pdb
solvateBox x TIP3PBOX 12
charge x
addions2 x Cl- 0
addions2 x Na+ 0
saveAmberParm x ${protein}.parm7 ${protein}.crd
```

Explicit solvent minimization

```
&cntrl
ntx = 1, irest = 0, ntrx = 1, ntxo = 1,
ntpr = 50, ntwx = 0, ntwv = 0, ntwe = 0,
ntf = 1, ntb = 1,
cut = 9.0, nsnb = 10,
ibelly = 0, ntr = 1,
imin = 1,
maxcyc = 20000,
ncyc = 10000,
ntmin = 1, dx0 = 0.1, drms = 0.0001,
ntc = 1, tol = 0.00001,
&end
Hold solvent fixed
1.0
RES 225 9999999
END
```

Explicit solvent heating from 0 to 100 K

```
&cntrl
imin=0, ! Molecular dynamics
ntx=1, ! Coordinates formatted with no initial velocities
irest=0, ! No restart velocities assigned
tol=0.000001, ! SHAKE tolerance
nstlim=2000000, ! Number of MD steps
ntt=3, ! Langevin dynamics
```

```
gamma_ln=5.0, ! Collision frequency Langevin dynamics, external heat bath, transfer KE to system in
velocity, collide
ig=-1, ! Random seed Langevin dynamics, stochastic
ntpr=25000,
ntwr=25000,
ntwx=25000, ! Write to trajectory file every ntwx steps
dt=0.001, ! Timestep (fs)
ntb=1,    ! Constant volume periodic boundary conditions
ntp=0,
cut=9.0,
tempi=0.0,          ! Starting temperature is 0 K from minimization
TEMP0=100.0, ! Reference temperature (held here)
&end
```

Explicit solvent heating from 100 to 300 K

```
&cntrl
imin=0, ! Molecular dynamics
ntx=5, ! Positions and velocities read NetCDF (or formatted)
irest=1, ! Restart calculation
ntc=2, ! SHAKE algorithm bonds with hydrogen
ntf=2, ! No force evaluation bonds with hydrogen
tol=0.0000001, ! SHAKE tolerance
nstlim=4000000, ! Number of MD steps
ntt=3, ! Langevin dynamics
gamma_ln=5.0, ! Collision frequency Langevin dynamics
ig=-1, ! Random seed Langevin dynamics
```

```
ntpr=2500,
ntwr=2500,
ntwx=2500, ! Write to trajectory file every ntwx steps
dt=0.001, ! Timestep (ps)
ntb=2,    ! Constant pressure periodic boundary conditions
ntp=1, ! Monte Carlo Barostat
cut=9.0,
&end
&wt
type='TEMP0', ! Varies the target temperature TEMP0
istep1=1000001, ! Initial step
istep2=3000000, ! Final step
value1=100.0, ! Initial temp0 (K)
value2=310.0, ! final temp0 (K)
&end
&wt
type='TEMP0', ! Varies the target temperature TEMP0
istep1=3000001, ! Initial step
istep2=4000000, ! Final step
value1=310.0, ! Initial temp0 (K)
value2=310.0, ! final temp0 (K)
&end
```

Explicit solvent molecular dynamics

```
 &cntrl
  imin=0,         ! Molecular dynamics
```

```
 ntx=5,          ! Positions and velocities read NetCDF (or formatted)
 irest=1,        ! Restart calculation
 ntc=2,          ! SHAKE on for bonds with hydrogen
 ntf=2,          ! No force evaluation for bonds with hydrogen
 tol=0.0000001,  ! SHAKE tolerance
 nstlim=5000000, ! Number of MD steps - 10 ns at 4 fs timestep
 ntt=3,          ! Langevin dynamics
 gamma_ln=2.0,   ! Collision frequency for Langevin dyn.
 temp0=310.0,    ! Simulation temperature (K)
 ntpr=2500,      ! Print to mdout every ntpr steps
 ntwr=2500,      ! Write a restart file every ntwr steps - 10ps
 ntwx=2500,      ! Write to trajectory file every ntwx steps
 ntwprt=3542,
 dt=0.002,       ! Timestep (ps) is 4 fs
 ig=-1,          ! Random seed for Langevin dynamics
 ntb=2,          ! Constant pressure periodic boundary conditions
 ntp=1,          ! Isotropic pressure coupling
 cut=9.0,        ! Nonbonded cutoff (Angstroms)
 ioutfm=1,       ! Write binary NetCDF trajectory
 ntxo=2,         ! Write binary restart file
 barostat=2,     ! Use Monte Carlo barostat (Amber 14+)
 iwrap=1,
&end
```

Explicit solvent QM/MM simulation

```
gfp and its mutant CRO QM/MM EXTERN(tc_job.tpl) production 12500 steps (dt=0.2fs) = 2.5 ps
```

```
&cntrl
imin = 0, ! no minimization
ntx = 5, ! 1-read in coordinates, but not velocity, 5-both
irest = 1, ! Restart calculation
ig = -1, ! random seed
cut = 9.0, ! Nonbonded cutoff (Angstroms)
dt = 0.0002, ! 0.2fs time step
nstlim = 20000, ! num steps
temp0 = 310.0, ! Simulation temperature (K)
ntc = 2,
ntf = 2, ! Shake is used for solvent
tol = 0.0000001, ! SHAKE tolerance
ntb = 2, ! Constant pressure periodic boundary conditions
ntt = 3, ! 0-const E, 1-rescale 3-Langevin dynamics 7-bussi
tautp = 0.01, ! time constant, in ps, for heat bath coupling for the system
gamma_ln = 5.0, ! Langevin dynamics collision frequency
ntpr = 10, ! print details to log every step
ntwx = 10, ! write coordinates to mdcrd every 500 steps (every 250fs)
ntwr = 10, ! write restart file at last step
ifqnt = 1, ! turn on QM/MM
ntb = 2, ! Constant pressure periodic boundary conditions
ntp = 1, ! Isotropic pressure coupling
cut = 9.0, ! Nonbonded cutoff (Angstroms)
barostat=2, ! Use Monte Carlo barostat (Amber 14+)
iwrap = 1,
/
```

```
&qmmm
qmmask = ':CRO', ! only cro residue, 300
qmcharge = 0,
qmcut = 8.0, !point charge in mm - long range electrostatic, non continous ! qm region
qmshake = 0, !Shake QM H atoms if shake is turned on (NTC>1) (default)
qm_theory = 'EXTERN',
qm_ewald = 0,
qmgb = 0,
verbosity = 2,
writepdb = 1,
/
&tc
executable = 'terachem',
use_template = 1,
/
tc_job.tpl
basis lacvps_ecp
method b3lyp
dispersion yes
scf diis+a
threall 1e-13
convthre 1e-6
xtol 1e-4
maxit 200
dftgrid 1
charge 0
```

```
spinmult 1
scrdir ./scr
keep_scr yes
end
```

TDDFT

```
run             energy # compute the electronic energy
gpus            1 # number of Graphics Processing Units (GPUs)
method          wpbeh # DFT functional
basis           lacvps_ecp # basis set
charge          0 # total charge of zero, neutral species
spinmult        1 # spin multiplicity, closed-shell configuration
sphericalbasis yes
guess           generate
maxit           100
scf             diis+a # DIIS method to accelerate SCF convergence
precision       dynamic
convthre        3.0e-5
xtol            1.0e-6
dftgrid         1
dynamicgrid     no
dispersion      no
purify          yes
pointcharges    ptchrg.xyz


cis             yes # configuration interaction singles (CIS) method is being used for excited state calculations
```

```
cisnumstates   30
cismax         150
cismult        1
cistarget      1
cismaxiter     100
cisprintthresh 0.1
cistransdensity        no
cisdiffdensity         no
```

cpptraj total energy, temperature, RMSD, RMSF, native contacts, RDF

```
readdata ${out}.out
writedata ${out}_etot.dat ${out}.out[Etot] time ${time}
writedata ${out}_temp.dat ${out}.out[TEMP] time ${time}
parm ${PROTEIN}.parm7
trajin ${TRIAL}.nc
autoimage
rms :CRO first out ${TRIAL}_rmsd_CRO.dat
rms :CRO<:5.5 out ${TRIAL}_rmsd_surround.dat
atomicfluct !:WAT out ${PROTEIN}_rmsf.dat byres
nativecontacts :CRO !:CRO byresidue writecontacts ${PROTEIN}_nativebyresidue.dat \
distance 5.0 contactpdb ${PROTEIN}_nativecontacts.pdb \
map mapout ${PROTEIN}_nativebyresidue.map
radial out ${TRIAL}_water.dat 0.1 15.0 :CRO :WAT
run
```

tICA and Kmeans clustering

```
def featurize_data(topfile):
    feat = coor.featurizer(topfile)
    residue_pairs = [[62, i] for i in range(1, 224)]
    feat.add_residue_mindist(residue_pairs)
    print(feat)
    print('number of features ', feat.dimension())
    return feat
def build_tica(data):
    tica = pyemma.coordinates.tica(data, lag=5)
    tica_output = tica.get_output()
    tica_concatenated = np.concatenate(tica_output)
    cluster = pyemma.coordinates.cluster_kmeans(
        tica_output, k=200, max_iter=50, stride=10, fixed_seed=1)
    dtrajs_concatenated = np.concatenate(cluster.dtrajs)
    fig, ax = plt.subplots(figsize=(6,4))
    pyemma.plots.plot_density(
        *tica_concatenated[:, :2].T, ax=ax, cbar=True, alpha=0.3,logscale=True)
    ax.scatter(*cluster.clustercenters[:, :2].T, s=5, c='C1')
    ax.set_xlabel('IC 1')
    ax.set_ylabel('IC 2')
    fig.tight_layout()
top = '1huy.parm7'
feat = featurize_data(top)
data = coor.load(['1huy_prod_2.nc', '1huy_prod_2.0007.nc', '1huy_prod_2.0008.nc'], features=feat)
build_tica(data)
plt.savefig('1huy-tica.png')
```