

# Unraveling the roGFP2 Cu(I) Sensing Mechanism

Alex M. Maldonado

Department of Biological Sciences

University of Pittsburgh

Pennsylvania, United States

[alex.maldonado@pitt.edu](mailto:alex.maldonado@pitt.edu)

Here, we propose atomistic mechanisms for a redox-sensitive green fluorescent protein variant roGFP2 including a novel Cu(I) sensing application.

## 1. GFP fluorescence mechanism

The fluorescence mechanism in green fluorescent protein (GFP) is a complex interplay of photophysical and photochemical processes that occur at the molecular level. This section delves into the current understanding of fluorescence in GFP, exploring the fundamental principles of excitation and de-excitation, as well as the various factors that influence these processes. We will examine how the protein environment modulates the fluorescent properties, the critical role of chromophore protonation states, and the intricate dynamics of excited-state phenomena such as non-adiabatic crossings and proton transfer. From this point forward, we will refer to eGFP as “GFP” and eGFP chromophore<sup>1</sup> as “chromophore” that is the result of the F64L and S65T mutations from the *Aequorea victoria* wild type GFP<sup>2</sup>.

### A. Chromophore

Fluorescence in GFP begins with the absorption of a photon by the chromophore, typically in the blue region of the visible spectrum (around 488 nm). This absorption process is intimately linked to the unique molecular structure of the chromophore. The chromophore is formed autocatalytically from three amino acid residues—Ser65, Tyr66, and Gly67—through a series of reactions involving cyclization, dehydration, and oxidation. The resulting structure consists of a hydroxybenzylidene imidazolinone moiety, which forms an extended  $\pi$ -conjugated system.

#### a. Excitation:

This  $\pi$ -conjugated system is crucial for the chromophore's light-absorbing properties. The delocalized electrons in the conjugated bonds can be excited by photons of specific energies, corresponding to the energy gap between the ground state ( $S_0$ ) and the first excited state ( $S_1$ ) of the chromophore. The exact absorption wavelength is fine-tuned by several factors that are discussed later: planarity, protonation state, and protein environment.

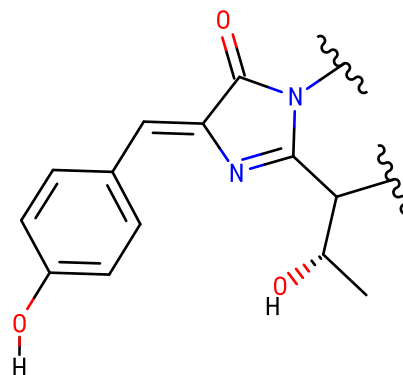


Fig. 1: Neutral (i.e., A state) GFP chromophore.

When a blue photon is absorbed, it promotes an electron from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO) of the chromophore. This electronic transition is predominantly  $\pi \rightarrow \pi^*$  in nature, reflecting the excitation within the  $\pi$ -conjugated system.

Following absorption, the excited chromophore undergoes rapid vibrational relaxation within the  $S_1$  state, typically on a femtosecond to picosecond timescale. This relaxation involves small structural adjustments in the chromophore and its immediate protein environment, preparing the system for the subsequent fluorescence emission.

#### b. Protonation states:

The protonation state of the GFP chromophore is a critical determinant of its photophysical properties, playing a crucial role in the protein's spectral characteristics and fluorescence behavior. The chromophore can exist in two primary forms: a neutral (protonated) state and an anionic (deprotonated) state, each exhibiting distinct spectroscopic signatures.

In its neutral form (shown in Fig. 1), the chromophore's phenolic oxygen is protonated, resulting in an absorption maximum typically around 395-400 nm. This state is often referred to as the A state. The neutral chromophore generally exhibits weaker fluorescence compared to its anionic counterpart, with emission maxima around 460 nm. The reduced fluorescence efficiency of the neutral form is attributed

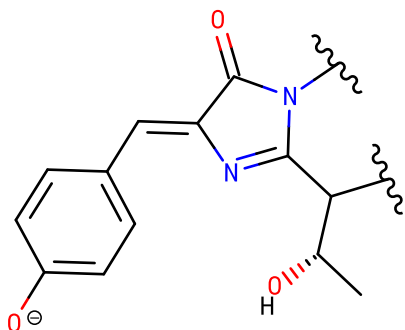


Fig. 2: Anionic (i.e., B state) GFP chromophore.

to excited-state dynamics that favor non-radiative decay pathways.

The anionic form of the chromophore, where the phenolic oxygen is deprotonated, is primarily responsible for the characteristic green fluorescence of GFP.

This state, often called the B state, has an absorption maximum at approximately 475-490 nm and emits strongly at around 510 nm. The anionic chromophore demonstrates a higher fluorescence quantum yield, making it the predominant contributor to GFP's bright fluorescence.

The relative population of these two states is influenced by several factors, including the local pH, specific interactions within the protein environment, and mutations in the protein sequence. In wild-type GFP, the chromophore exists in an equilibrium between these two states, with the population distribution heavily dependent on pH. Under physiological conditions, the anionic form is typically favored.

### c. Planarity:

The planarity of the eGFP chromophore plays a pivotal role in determining its fluorescent properties, particularly through enhanced conjugation and increased quantum yield. These factors contribute significantly to the chromophore's spectroscopic characteristics and efficiency.

Enhanced conjugation in a planar chromophore structure is primarily due to the maximized overlap of p-orbitals in the  $\pi$ -conjugated system. This optimal alignment of p-orbitals, perpendicular to the molecular plane, facilitates efficient delocalization of  $\pi$ -electrons across the entire conjugated system. Such extensive electron delocalization has profound effects on the chromophore's electronic structure, most notably in reducing the energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO).

Planarity significantly impacts the chromophore's quantum yield by restricting non-radiative decay pathways. In non-planar configurations, rotation around single bonds, can serve

as an efficient route for non-radiative decay. However, when the chromophore is held in a planar conformation by the protein environment, these rotational movements are severely limited. This restriction of molecular motion creates energy barriers in the excited state potential energy surface, effectively preventing the chromophore from accessing geometries that favor non-radiative decay.

Furthermore, the planar structure reduces coupling between electronic and vibrational states, which might otherwise lead to non-radiative relaxation through internal conversion. The minimized structural changes between ground and excited states in a planar chromophore also contribute to reduced internal conversion rates. Consequently, with fewer available non-radiative pathways, the excited chromophore is more likely to return to the ground state via fluorescence emission, directly increasing the fluorescence quantum yield.

### d. Neutral-state fluorescence:

Excitation and subsequent emission of the neutral state of the chromophore without any structural changes is one—infrequent—possibility.

Excitation of the neutral state occurs at approximately 395 nm, corresponding to the absorption of violet-blue light. This excitation promotes the chromophore to its first excited singlet state ( $S_1$ ) without immediate proton transfer. The subsequent emission from this excited neutral state results in weak blue fluorescence with a peak around 460 nm.

This protonated chromophore can be further stabilized by hydrogen bonding with a water molecule, Thr203, or Ser205. The presence or absence of these interactions contribute to

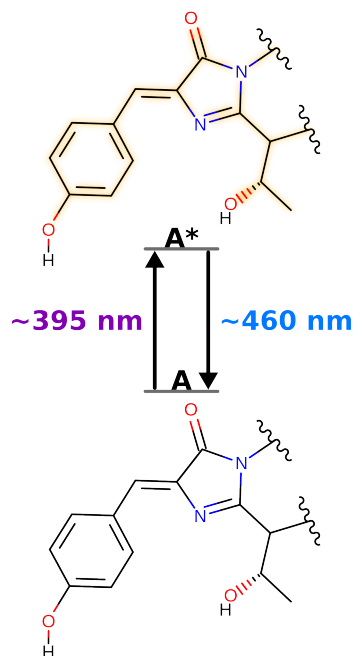


Fig. 3: Excitation and emission of neutral chromophore.

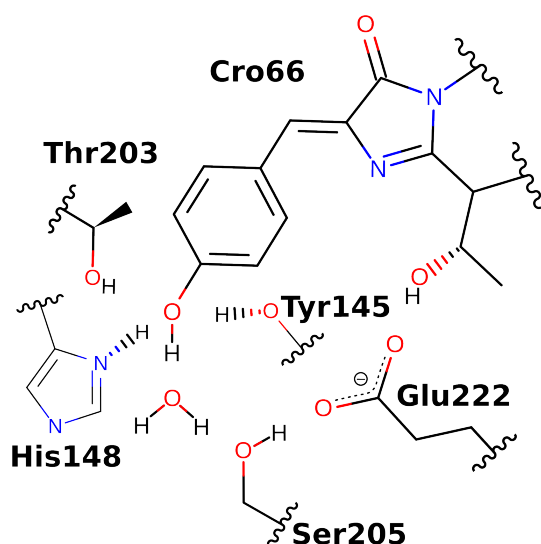


Fig. 4: Example stabilizing configuration for the neutral chromophore.

the distinct excitation and emission. Although, this emission is relatively weak compared to the anionic state discussed next.

#### e. Anionic-state fluorescence:

The anionic state of the chromophore represents a key configuration responsible for the protein's characteristic green fluorescence. In this state, the chromophore exists in its deprotonated form, with the phenolic oxygen carrying a negative charge.

Excitation of the anionic chromophore occurs at approximately 475 nm, corresponding to the absorption of blue light. This excitation promotes the chromophore to its first excited singlet state ( $S_1$ ). The subsequent relaxation and emission result in the bright green fluorescence typically associated with GFP, with a peak around 508 nm.

The anionic state is stabilized by specific interactions within the protein barrel. Notably, Thr203 plays a crucial role in stabilizing the anionic form through hydrogen bonding with the deprotonated phenolic oxygen. Additionally, a protonated Glu222 could form a hydrogen bond with the anionic chromophore, further contributing to its stabilization.

The protein environment around the chromophore is critical in maintaining this anionic configuration. The  $\beta$ -barrel structure of eGFP provides a hydrophobic pocket that shields the chromophore from bulk solvent, contributing to the high quantum yield of fluorescence in this state.

#### f. Excited-state proton transfer:

The excited-state proton transfer (ESPT) mechanism represents a fundamental process in GFP, contributing significantly to its unique spectroscopic properties. This process involves the initial excitation of the neutral (protonated) chromophore, followed by rapid, successive proton transfer

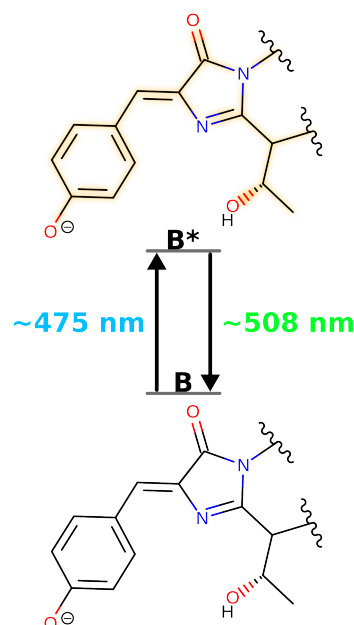


Fig. 5: **TODO: Add caption**

events in the excited state, ultimately resulting in emission from an anionic species. An overview of the process is shown below.

1. Upon absorption of a photon at approximately 395 nm, the neutral chromophore is promoted to its first excited singlet state ( $S_1$ ).
2. In this excited state, the chromophore exhibits markedly different acid-base properties compared to its ground state, becoming a much stronger acid. This enhanced acidity facilitates the transfer of a proton from the chromophore to a proximal acceptor within the protein matrix. The proton transfer pathway involves a sophisticated hydrogen-

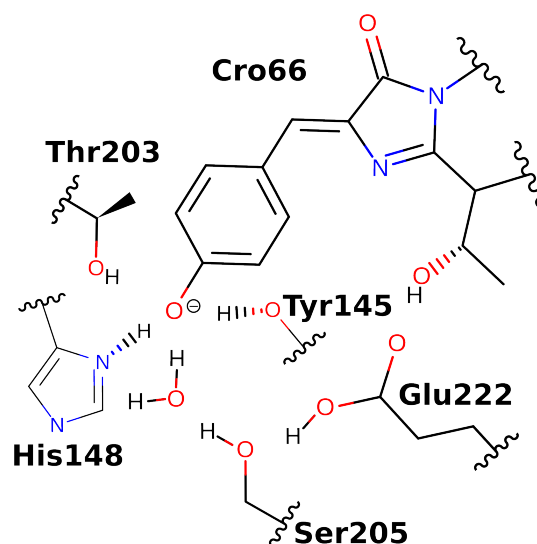


Fig. 6: Example stabilizing configuration for the anionic chromophore.

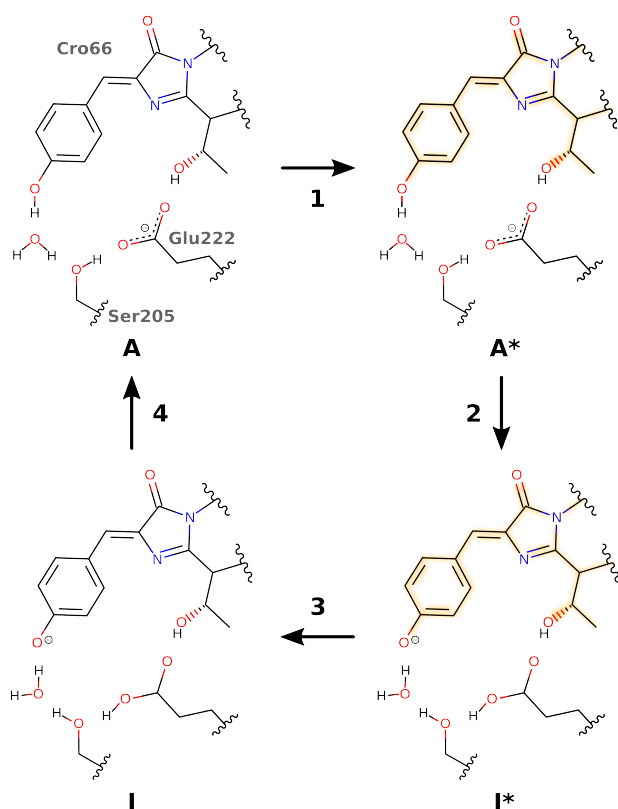


Fig. 7: **TODO: Add caption**

bonding network. A critical component of this network is a strategically positioned water molecule, which acts as the initial proton acceptor. This water molecule is part of a proton wire that includes Ser205 and terminates at Glu222. The transfer occurs on an ultrafast timescale, typically in the order of picoseconds.

- Following the ESPT, the system exists transiently in an intermediate state ( $I^*$ ), characterized by an anionic chromophore and a protonated Glu222.  $I^*$  subsequently relaxes and emits fluorescence at approximately 508 nm, closely resembling the emission profile of the intrinsically anionic chromophore.
- The final step is a reverse ground-state proton transfer from Glu222 through Ser205, a water molecule, and terminated at the chromophore.

## B. Roles of crucial residues

The exceptional fluorescent properties of GFP arise from a complex interplay of molecular interactions within its  $\beta$ -barrel structure.

### a. His148:

His148 plays a pivotal role in the structure and function of GFP, significantly contributing to chromophore stabilization and modulation of photophysical properties.

Structurally, His148 helps maintain  $\beta$ -strand stability by hydrogen bonding to Arg168's backbone (**CITE**). Others have also explored introducing a hole in the  $\beta$  barrel through a H148G mutation; thereby allowing metal ions to diffuse into the protein and interact directly with the chromophore<sup>3</sup>.

His148 directly interacts with the chromophore, forming a critical hydrogen bond. Specifically, the imidazole side chain of His148 donates a hydrogen bond to the chromophore's phenolate oxygen (**CITE**). This interaction is vital for stabilizing the anionic form of the chromophore.

The role of His148 in GFP's photophysics extends beyond static structural interactions. Upon chromophore excitation, His148 facilitates proton movement, contributing to the complex photochemistry that gives rise to GFP's unique fluorescent properties<sup>4</sup>.

His148, along with Tyr145, aids in locking the chromophore into a planar conformation. This planar structure is essential for efficient  $\pi$ -electron delocalization across the chromophore system, directly influencing its fluorescent properties. The rigid environment provided by these residues minimizes non-radiative decay pathways, contributing to GFP's high fluorescence quantum yield.

The importance of His148 has been further elucidated through mutational studies. Of particular interest is the H148G mutation, which introduces a hole in the  $\beta$ -barrel structure. Barondeau et al. demonstrated that this mutation allows metal ions to diffuse into the protein and interact directly with the chromophore<sup>3</sup>.

### b. Tyr145:

Tyr145 plays a crucial role in stabilizing the chromophore and influencing its photophysical properties. Located in close proximity to the chromophore within the  $\beta$ -barrel structure of GFP, Tyr145 contributes significantly to the protein's unique fluorescent characteristics through several key interactions.

One of the primary functions of Tyr145 is its direct involvement in the hydrogen bonding network surrounding the chromophore. The phenolic hydroxyl group of Tyr145 serves as a hydrogen bond donor to the imidazolinone nitrogen of the chromophore. This interaction is critical for maintaining the chromophore in its optimal conformation for fluorescence. The hydrogen bond between Tyr145 and the chromophore contributes to the rigidity of the chromophore's environment, which is essential for efficient fluorescence by minimizing non-radiative decay pathways.

Tyr145 also plays a significant role in modulating the electronic structure of the chromophore. Through its proximity and interactions, Tyr145 influences the  $\pi$ -electron system of the chromophore, affecting its spectral properties. The specific orientation and strength of the hydrogen bond formed by Tyr145

can lead to subtle shifts in the absorption and emission spectra of GFP.

### c. Thr203:

Thr203 is located close to the chromophore within the  $\beta$ -barrel structure; its hydroxyl group can form a hydrogen bond with the chromophore's phenolate oxygen in its anionic state. The position and orientation of the side chain influence the chromophore's electronic environment, affecting both absorption and emission spectra. Mutations at this position, such as T203V or T203Y, disrupt this hydrogen bond and destabilize the deprotonated B-state of the chromophore, shifting the equilibrium towards the neutral A-state.

T203 mutations also affect ESPT efficiency fluorescence quantum yields. The T203V mutation maintains high ESPT efficiency similar to wild-type, while T203Y significantly reduces ESPT. T203V and T203Y mutations generally lead to higher fluorescence quantum yields of the deprotonated chromophore form compared to wild-type T203, likely by suppressing non-radiative decay pathways like rotation around the chromophore's exocyclic bond<sup>5</sup>.

## 2. roGFP2 contains redox-sensing cysteines

Redox-sensitive green fluorescent proteins (roGFPs) are engineered variants of GFP designed to report cellular redox states through changes in their fluorescence properties<sup>6</sup>. Among these, roGFP2 has emerged as a particularly useful probe due to its ratiometric readout and midpoint potential, which are suitable for measuring redox conditions in reducing cellular compartments like the cytosol and mitochondria.

roGFP2 contains two cysteine residues (S147C and Q204C) on adjacent  $\beta$ -strands near the chromophore of a GFP variant already containing mutations C48S, S65T, and Q80R. (A structural depiction of the relevant residues is in Fig. 8.) The strategically placed cysteines can form a reversible disulfide bond in response to the surrounding redox environment changes. Formation of this disulfide alters the protonation state of the chromophore, resulting in reciprocal changes in the excitation peaks at 400 nm and 490 nm.

## 3. roGFP2 fluorescence response

Fig. 9 illustrates the redox-dependent fluorescence properties of roGFP2, showing the protein's excitation spectra under various redox conditions. These spectra were obtained by monitoring emission at 511 nm while scanning excitation wavelengths from 350 to 500 nm. The protein samples were equilibrated in buffers containing different ratios of oxidized and reduced dithiothreitol (DTT) to achieve a range of defined redox potentials.

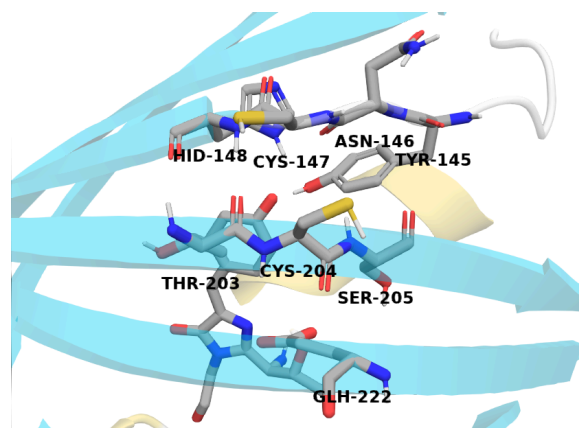


Fig. 8: Fluorescence-relevant residues in reduced roGFP2.

roGFP2 exhibits two distinct excitation peaks: (1) the A band at 400 nm and (2) the B band at 490 nm. These peaks correspond to 511 nm fluorescence of the chromophore's excited neutral (protonated) and anionic (deprotonated) forms. We will refer to the neutral and anionic forms of the chromophore as "A state" and "B state", respectively.

We observe a clear shift in the excitation spectrum as the environment becomes more oxidizing (moving from  $-0.240$  V to  $-0.310$  V). The intensity of the A band increases significantly, while the B band experiences a concomitant decrease. This spectral shift reflects the formation of the disulfide bond between Cys147 and Cys204 and an increasing preference for the protonated chromophore.

Notably, the spectra display a clear isosbestic point at approximately 425 nm, where the fluorescence intensity remains constant regardless of the redox state. This isosbestic point is a hallmark of a two-state system, confirming

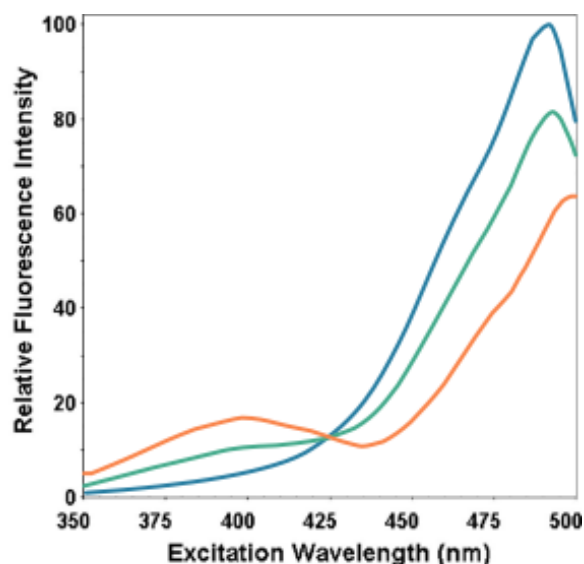


Fig. 9: Relative fluorescence at 511 nm after excitation scan from 350 to 500 nm at  $-0.310$  (blue),  $-0.275$  (green), and  $-0.240$  V (orange) redox potentials. Adapted with permission from Hanson et al.<sup>6</sup> distributed under the CC-BY-4.0 license.



that roGFP2 is transitioning cleanly between its oxidized and reduced forms without significant intermediate states. This spectral feature is a critical indicator of the probe's behavior: it suggests that the engineered disulfide bond is forming and breaking as intended, without competing side reactions or alternative conformations significantly affecting the fluorescence. The maintenance of this isosbestic point across various redox potentials implies that the structural changes induced by oxidation and reduction are consistent and reversible. Any deviation from this behavior, such as a shift in the isosbestic point or its disappearance, would suggest more complex interactions—potentially involving intermediate states, protein structural changes, or interactions with other molecules—that could complicate data interpretation.

#### A. Possible perturbations

Changes in A- or B-band absorption could indicate a variety of environmental changes in the chromophore.

- **Equilibrium ratio of A or B state populations.** Changes in neutral (A state) or anionic (B state) chromophore stability impacts the relative proportion of A- and B-band absorbance. For example, roGFP1 has a higher A band instead of B band<sup>6</sup>. A-state stability is directly correlated to A-band fluorescence; whereas the B band would be inversely correlated.
- **Excited-state proton transfer (ESPT) from A\*  $\rightarrow$  I\*.** Emissions at the typical 511 nm (green) fluorescence from the A state requires an ESPT from Cro66 to Glu222 through a coordinated water molecule and Ser205. Prohibiting ESPT would result in radiative emission at 460 nm which often is not monitored experimentally.
- **Ground-state proton transfer (GSPT) from I  $\rightarrow$  A.** Reprotonating the chromophore through a GSPT is crucial for maintaining the A band and B-band lifetime. Disrupting the Glu222  $\rightarrow$  Ser205, Ser205  $\rightarrow$  H<sub>2</sub>O, or H<sub>2</sub>O  $\rightarrow$  Cro66 pathway would decrease the A-state population&mdash;likely with a corresponding B state increase.
- **Non-radiative emissions.** Enhanced flexibility of the chromophore through protein conformational shifts would lead to additional non-adiabatic crossings; thereby lowering the relative fluorescence in that state.

## 4. Molecular simulations

#### A. Protein preparation

Initial protein structures for reduced (1JC0) and oxidized (1JC1) states of roGFP2 were retrieved from the Protein Data Bank (PDB). The structures were processed using in-house Python and bash scripts (available free of charge at [gitlab.com/oasci/studies/metalfare](https://gitlab.com/oasci/studies/metalfare)). The first chain, along with the crystallographic water molecules, were centered to

the origin while minimize the box values using NumPy (CITE), SciPy (CITE), and MDAnalysis (CITE) packages.

Given the relevance and availability of the GFP mechanism's force field parameters, the chromophore was modeled in its anionic state. All selenomethionine residues (MSE) were converted to methionine (MET), and Cys147 and Cys204 were transformed into the appropriate residues. Glu222 was protonated to model the I state of the GFP photocycle; this still probes the anionic chromophore's stability while offering insight into GSPT dynamics. The protonation states of all other residues were determined with PDB2PQR<sup>7</sup>, using the default parameters to ensure standardization. The pdb4amber tool was then used to validate the PDB file before proceeding.

#### B. Cu(I) placement

**TODO: Add GFN2-xTB calculations.**

#### C. Simulation preparation

System preparation was performed using the tleap module of AmberTools v23.6 (CITE). The protein structure was parameterized using the ff19SB force field (CITE). For the solvent environment, we employed the OPC3 water model, (CITE), which is known for its balanced representation of water properties in biomolecular simulations. The 12-6 nonbonded model and parameters for all ions were taken from Sengupta et al<sup>8</sup>. The system was neutralized by adding Na<sup>+</sup> and Cl<sup>-</sup> ions as needed. Additional ions were introduced to achieve a solvent ionic strength of 0.150 M to mimic physiological conditions. The protein was solvated in a rectangular box, with a minimum distance of 10 Å between the protein and the box edges to minimize periodic boundary condition artifacts. Parameters from Breyfogle et al.<sup>9</sup> were employed for the anionic chromophore.

#### D. Minimization

The prepared system underwent a four-stage energy minimization protocol using Amber23 (CITE) to relieve any unfavorable interactions and optimize the structure. All minimization stages used the steepest descent method for the first 1000 steps, followed by the conjugate gradient method for the remaining steps, with a maximum of 5000 steps per stage. A non-bonded cutoff of 10.0 Å was applied throughout. Periodic boundary conditions were employed, and coordinates were wrapped to the primary unit cell. The minimization progress was monitored by writing energies every step and coordinates every 200 steps.

**Stage 1:** Initial minimization was performed with restraints (force constant: 5.0 kcal/mol/Å<sup>2</sup>) on all non-hydrogen atoms of the entire system, allowing hydrogen atoms to relax and adjust their positions. **Stage 2:** The system was further minimized with restraints (force constant: 5.0 kcal/mol/Å<sup>2</sup>) on all non-

hydrogen atoms of the solute (excluding water molecules and ions), allowing solvent and ions to equilibrate around the solute. *Stage 3:* Minimization continued with reduced restraints (force constant: 2.0 kcal/mol/Å<sup>2</sup>) applied only to the protein backbone, allowing side chains and other flexible parts to relax. *Stage 4:* Final minimization was performed with further reduced restraints (force constant: 1.0 kcal/mol/Å<sup>2</sup>). The resulting minimized structure served as the starting point for subsequent relaxation and production simulations.

### E. Relaxation simulations

Following energy minimization, the system underwent a three-stage relaxation protocol using Amber23 to gradually equilibrate the structure and solvent. Three independent runs were initiated with random initial velocities to ensure adequate sampling. All subsequent simulations were continued using the respective run's restart files.

*Stage 1:* An initial 20 ps NVT (constant Number of particles, Volume, and Temperature) simulation was performed with a 2 fs time step. The system was heated from 100 K to 300 K using Langevin dynamics with a collision frequency of 5 ps<sup>-1</sup>. Restraints (force constant: 1.0 kcal/mol/Å<sup>2</sup>) were applied to the protein backbone. SHAKE algorithm was used to constrain bonds involving hydrogen atoms. The non-bonded cutoff was set to 10.0 Å. *Stage 2:* A 1 ns NPT simulation followed, maintaining the temperature at 300 K using Langevin dynamics (collision frequency: 5 ps<sup>-1</sup>). Pressure was regulated at 1.01325 bar using the Monte Carlo barostat with a relaxation time of 1 ps. Restraints on the same atoms were reduced (force constant: 0.5 kcal/mol/Å<sup>2</sup>). *Stage 3:* The final relaxation stage consisted of a 1 ns NPT simulation with all positional restraints removed.

### F. Production simulations

All production runs were performed under the same setup as the last relaxation stage. Each run was simulated for 500 ns with coordinates saved every ten ps. The resulting trajectories from all three replicates were used for subsequent analyses, providing a cumulative 1.5 μs of simulation data for the system under study.

### G. Analysis

From the MD trajectories, we extracted two primary sets of data:

1. Dihedral angles ( $\varphi$ ,  $\psi$ ) were calculated using MDAnalysis (CITE) for specific residues known to be crucial for roGFP2 function.
2. Key interactions between various residues through intermolecular distances.

Dihedral angles were transformed using

$$\frac{1 - \cos(\theta)}{2}$$

This transformation maps the circular dihedral data to a [0, 1] range, preserving the periodicity while differentiating between cis and trans conformations.

All input features ( $X$ ) were standardized using sklearn's StandardScaler to ensure each feature contributes equally to the model. This transformation ensures continuity across the periodic boundary, facilitating more accurate modeling.

#### a. Potential of mean force (PMF):

**TODO: Write this.**

#### b. Feature correlation:

To investigate the relationship between structural descriptors and various features, we employed Partial Least Squares (PLS) regression analysis. This multivariate statistical technique was chosen for its ability to handle high-dimensional, correlated data and reveal underlying patterns in complex datasets.

A PLS regression model was fitted to  $X$  and response variable  $y$  using sklearn's PLSRegression with two components for each simulation state. The model's performance was evaluated using the  $R^2$  score.

Data points were projected onto the space of the first two PLS components. A 2D histogram was created in this space, with bin colors representing the mean  $y$ .

Loading vectors for each feature were plotted as arrows in the PLS component space. The magnitude and direction of these arrows indicate the importance and relationship of each feature to the PLS components. The magnitude of each feature's loading vector was calculated as the Euclidean norm of its first two PLS components.

A dashed line representing the direction of maximum change in the response variable was added to the plot. This line, referred to as the derivative line, indicates the direction in the PLS component space along which  $y$  increases most rapidly.

### H. Feature importance

We developed a machine learning pipeline to elucidate the relationship between backbone dihedral angles and the target feature. Our approach employed two complementary models:

- **XGBoost Regressor:** A gradient boosting algorithm chosen for its high performance and ability to capture non-linear relationships (CITE).
- **Elastic Net:** A regularized linear regression model selected to identify linear correlations while mitigating multicollinearity (CITE).

The dataset was randomly partitioned into training (80%) and testing (20%) sets, ensuring model generalizability. We utilized GridSearchCV with 3-fold cross-validation to tune model

hyperparameters. For XGBoost, we optimized the number of estimators (250-700), learning rate (0.05-0.2), maximum tree depth (5-9), and regularization terms ( $\alpha$ : 0.0-0.2,  $\gamma$ : 0.8-1.0). For Elastic Net, we tuned the regularization strength ( $\alpha$ :  $10^{-5}$  - 5) and L1 ratio (0.2-1.0). Performance was assessed using mean squared error (MSE) and coefficient of determination ( $R^2$ ) on the held-out test set.

**Feature Importance Analysis** We extracted XGBoost feature importance scores directly from the model, but used absolute values of the ElasticNet coefficients as a proxy for feature importance. This dual-model approach allows for a robust comparison of feature rankings, mitigating model-specific biases.

### 1. Hydrogen bond cutoff

A hydrogen bond of  $X-H \cdots Y-Z$ , where X is the donor and Y is the acceptor atom, can be classified based on distances and angles. One characteristic recommended by IUPAC is that the  $H \cdots Y$  distance is less than the sum of H and Y van der Waals radii. Hydrogen (1.10 Å) and oxygen (1.52 Å)<sup>10</sup> would have a cutoff of 2.62 Å. Others<sup>11</sup> recommend a cutoff of 2.50 Å based on structural analysis<sup>12</sup> and quantum chemical calculations<sup>13</sup>. Since the difference between a 2.5 and 2.62 Å cutoff is likely a substantially weak hydrogen bond, we will use a  $H \cdots Y$  cutoff of 2.5 Å.

## 5. Fluorescence mechanism of Cu(I) distinct from oxidation

Joel Rosenbaum performed in vitro assays of roGFP2 under  $H_2O_2$  (i.e., oxidation) and Cu(I) conditions by monitoring 528 nm emissions **TODO: Check emission value** with scanning excitation wavelengths between 380 and 500 nm. Relative fluorescence with respect to **TODO: How was this normalized?** **Check methods** is shown in Fig. 10.

Apo represents our control with roGFP2 in its reduced state. Adding 1.0 mM  $H_2O_2$  results in the expected shift in the A- and B-band proportions. Again, this indicates that oxidizing Cys147 and Cys204 and forming a disulfide bridge increases the neutral chromophore equilibrium proportion—with a corresponding decrease in the anionic state. This oxidation-induced change in the chromophore protonation state is the basis for roGFP2's utility as a redox sensor in biological systems.

Cu(I), on the other hand, exhibits an entirely different fluorescence response. Strikingly, a mere 1  $\mu$ M of Cu(I) drastically reduces the B-band intensity while keeping the A band relatively stable. The distinct responses to oxidation and Cu(I) binding provide compelling evidence that Cu(I) binding to roGFP2 does not simply alter the chromophore's protonation state preference, as oxidation does. Instead, these results suggest that Cu(I) may disturb the anionic chromophore's

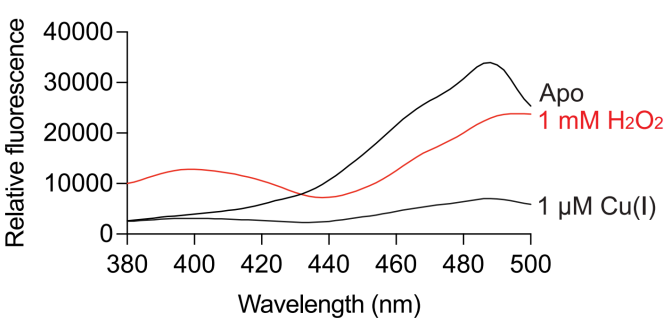


Fig. 10: Relative fluorescence of roGFP2 under reduced, oxidized, and Cu(I) conditions from 380 to 500 nm excitation scan. **TODO: Check emission value?** Apo (i.e., reduced) roGFP2 exhibits typical bimodal absorption of A-band (excited at 400 nm) and B-band (excited at 488) peaks. Upon roGFP2 oxidation from 1 mM  $H_2O_2$ , a shift in A- (increased) and B-band (decreased) absorption and subsequent 528 nm emission marks a corresponding change in neutral and anionic chromophore populations. Binding of Cu(I), however, exhibits a larger decrease in B-band without the A-band increase observed when oxidized.

electronic state or induce conformational changes that specifically affect its environment, underscoring the need for further exploration.

## 6. Cu(I) binding enhances roGFP2 backbone flexibility

First, we investigate the backbone dynamics over 1.5  $\mu$ s of MD simulations.

### A. Cys147 and Cys207 $C_\alpha$ distance

Experimental structures of both the reduced (PDB ID: 1JC0) and oxidized (PDB ID: 1JC1) states of for roGFP2 exhibited a mean  $C_\alpha$ - $C_\alpha$  distance of  $4.30 \pm 0.12$  and  $4.07 \pm 0.09$  Å, respectively<sup>6</sup>. Our MD simulations agreed well with experimental observations as shown in Table I with mean differences less than 0.04 Å.

TABLE I: MEAN AND STANDARD DEVIATIONS OF  $C_\alpha$ - $C_\alpha$  DISTANCES BETWEEN CYS147 AND CYS204.

State	Experimental (Å)	Simulations (Å)
Reduced	$4.30 \pm 0.12$	$4.34 \pm 0.47$
Oxidized	$4.07 \pm 0.09$	$4.11 \pm 0.29$
Cu(I)	N/A	$4.78 \pm 0.82$

The standard deviations,  $\sigma$ , also provide an indication of conformational flexibility. Unsurprisingly, the  $\sigma$  of oxidized roGFP2 exhibits is substantially smaller than the reduced state.

An experimental structure of roGFP2-Cu(I) is currently not available; however, our simulations suggest that it is on average 0.44 Å larger than the reduced state. Cu(I) stayed coordinated between Cys147 and Cys204 throughout all simulations while enhancing conformational flexibility. For example, Fig. 11 shows a broader distribution of  $C_\alpha$ - $C_\alpha$  distances with a slight bimodal peaks centered around 4.48 Å and 4.96 Å.



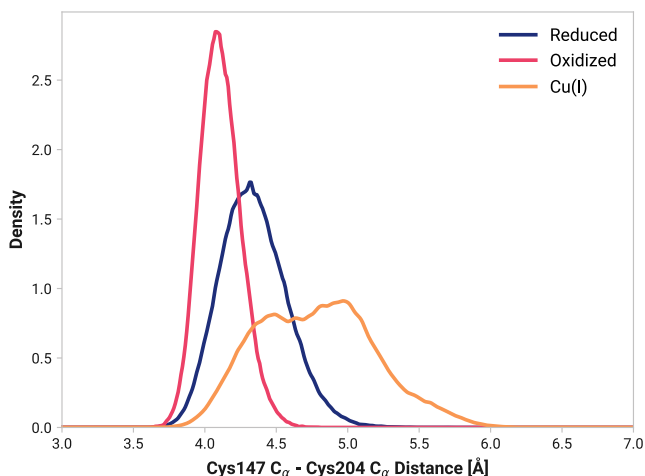


Fig. 11: **TODO: Add caption**

## B. $\beta$ -strand fraying

GFP and its variants are characterized by a distinctive  $\beta$ -barrel structure, which plays a crucial role in their fluorescence properties. This  $\beta$ -barrel is composed of 11  $\beta$ -strands, forming a robust scaffolding that encapsulates and protects the centrally located chromophore. The structural integrity of this  $\beta$ -barrel is essential for maintaining the protein's fluorescence characteristics and its sensitivity to environmental changes. In roGFP2, Cys147 is strategically positioned near the C-terminus of a  $\beta$ -pleated sheet, flanked by His148 and Thr203 (Figure Fig. 8).

Our molecular dynamics simulations reveal a striking structural change upon Cu(I) binding to roGFP2. Specifically, we observe a significant disruption in the  $\beta$ -sheet structure near Cys147. Fig. 12 illustrates that Cu(I) binding leads to the breaking of a key  $\beta$ -sheet hydrogen bond in this region. This observation is further substantiated by quantitative analysis

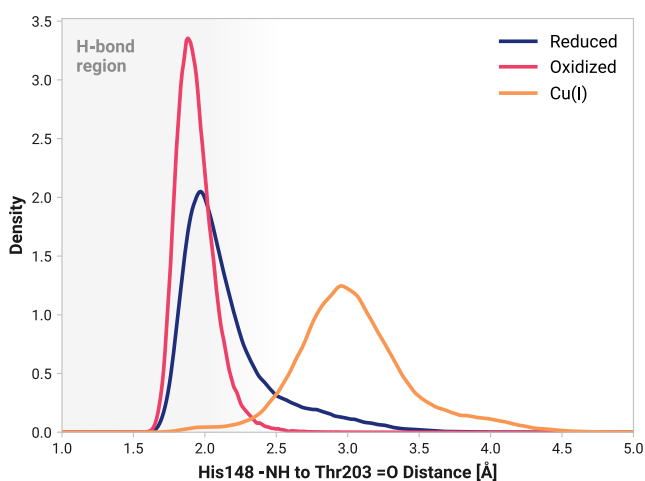


Fig. 12: Probability density of His148 and Thr203 backbone hydrogen bonding under various roGFP2 conditions.

TABLE II: HYDROGEN BONDING PROBABILITY BETWEEN RESIDUE BACKBONES

State	His148 - Thr203	Asn146 - Ser205
Reduced	0.865	0.047
Oxidized	0.997	0.144
Cu(I)	0.063	0.000

presented in Table II, which shows a dramatic decrease in the hydrogen bond probability between the backbone  $\text{-NH}$  and  $\text{C=O}$  groups within a 2.5 Å distance. The probability drops from 0.865 in the reduced state to a mere 0.063 in the Cu(I)-bound state, indicating a near-complete loss of this stabilizing interaction. (Unsurprisingly, oxidized roGFP2 stabilizes this hydrogen bond.)

This disruption of the  $\beta$ -sheet hydrogen bonding network can be characterized as “fraying” of the  $\beta$ -strand. Fraying is a phenomenon where the regular hydrogen bonding pattern at the termini of secondary structure elements, particularly  $\beta$ -strands, becomes disrupted, leading to increased local flexibility.

In the case of roGFP2, the Cu(I)-induced fraying at the C-terminus of this  $\beta$ -strand is likely to have several important implications. Given the proximity of this structural change to the chromophore, it may directly influence the chromophore's electronic environment, contributing to the observed changes in fluorescence properties upon Cu(I) binding.

The structural perturbation at this site could also propagate through the protein structure. Indeed, Fig. 13 shows the intermolecular distance peak between Asn146 and Ser205 increasing from 3.80 to 4.48 Å with bound Cu(I). Distances with Cu(I) extends out to at least 0.66 Å further than the reduced state. Under oxidizing conditions, the peak distance only decreases to 3.56 Å and hydrogen bonds 14.4 % during the course of the simulations.

No changes are observed in the Val150 and Leu201  $\beta$ -strand.

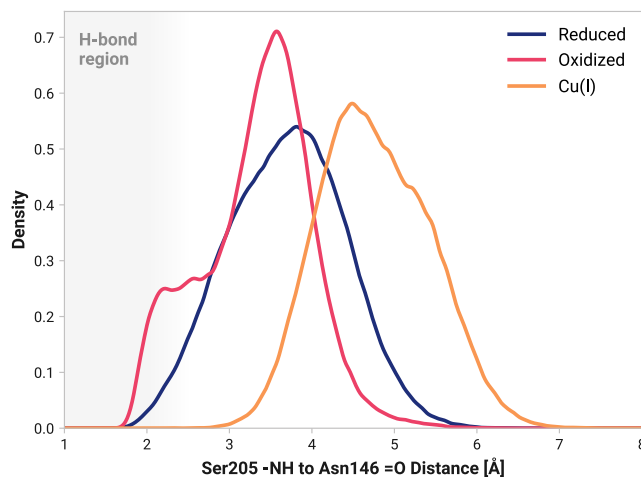


Fig. 13: **TODO: Add caption**

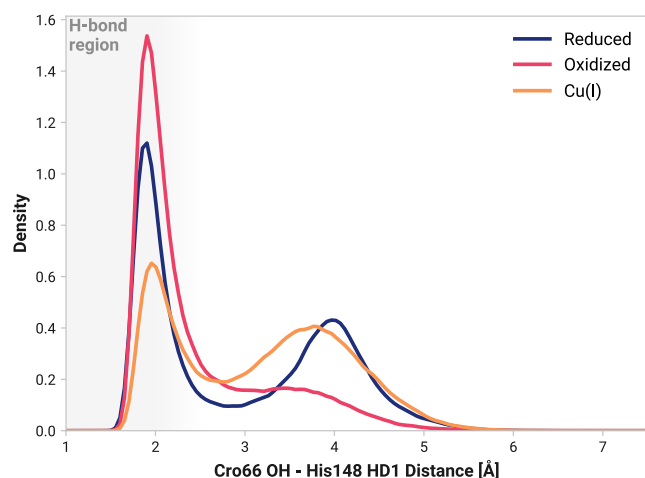


Fig. 14: **TODO: Add caption**

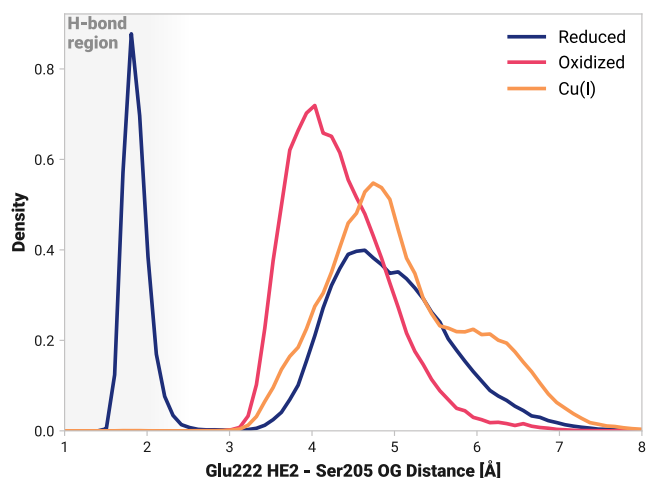


Fig. 16: **TODO: Add caption**

## 7. Cu(I) binding disrupts anionic chromophore hydrogen binding

As previously mentioned, several residues are in close proximity to the chromophore. Numerous studies have observed complex interactions and have led to several variants tailored for specific applications. roGFP2 is no different.

His148 often coordinates with the chromophore, both reduced and oxidized simulations spend 48.6 and 69.1% of the time within 2.5 Å. With bound Cu(I), this percentage drops to 33.9%. Outside of this time, the  $\delta$ -protonated His148 hydrogen bonds to the Asn146 backbone, while keeping the  $\epsilon$ -nitrogen coordinated to Arg168 for at least 68.7% of the time (data not shown).

His148 and Thr203  $\beta$ -strand fraying appears to correlates correlate with decreased stabilization of the anionic chromophore through His148. Training

## 8. Disruption of ground-state proton transfer

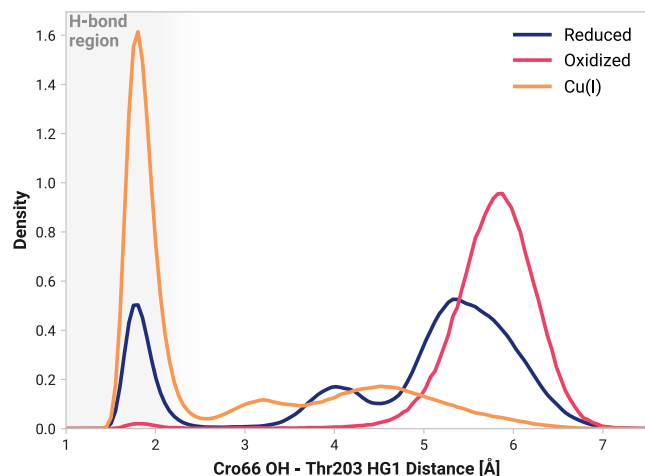


Fig. 15: **TODO: Add caption**

## References

- (1) Cormack, B. P. ; Valdivia, R. H. ; Falkow, S. FACS-Optimized Mutants of the Green Fluorescent Protein (GFP). *Gene* **1996**, 173 (1), 33–38
- (2) Prasher, D. C. ; Eckenrode, V. K. ; Ward, W. W. ; Prendergast, F. G. ; Cormier, M. J. Primary Structure of the Aequorea Victoria Green-Fluorescent Protein. *Gene* **1992**, 111 (2), 229–233
- (3) Barondeau, D. P. ; Kassmann, C. J. ; Tainer, J. A. ; Getzoff, E. D. Structural Chemistry of a Green Fluorescent Protein Zn Biosensor. *Journal of the American Chemical Society* **2002**, 124 (14), 3522–3524
- (4) Shinobu, A. ; Palm, G. J. ; Schierbeek, A. J. ; Agmon, N. Visualizing Proton Antenna in a High-Resolution Green Fluorescent Protein Structure. *Journal of the American Chemical Society* **2010**, 132 (32), 11093–11102
- (5) Jung, G. ; Wiehler, J. ; Zumbusch, A. The Photophysics of Green Fluorescent Protein: Influence of the Key Amino Acids at Positions 65, 203, And 222. *Biophysical journal* **2005**, 88 (3), 1932–1947
- (6) Hanson, G. T. ; Aggeler, R. ; Oglesbee, D. ; Cannon, M. ; Capaldi, R. A. ; Tsien, R. Y. ; Remington, S. J. Investigating Mitochondrial Redox Potential with Redox-Sensitive Green Fluorescent Protein Indicators. *Journal of Biological Chemistry* **2004**, 279 (13), 13044–13053. <https://doi.org/10.1074/jbc.m312846200>
- (7) Jurrus, E. ; Engel, D. ; Star, K. ; Monson, K. ; Brandi, J. ; Felberg, L. E. ; Brookes, D. H. ; Wilson, L. ; Chen, J. ; Liles, K. ; others. Improvements to the APBS Biomolecular

Solvation Software Suite. *Protein Science* **2018**, 27 (1), 112–128

- (8) Sengupta, A. ; Li, Z. ; Song, L. F. ; Li, P. ; Merz Jr, K. M. Parameterization of Monovalent Ions for the Opc3, OPC, Tip3p-FB, And Tip4p-FB Water Models. *Journal of chemical information and modeling* **2021**, 61 (2), 869–880
- (9) Breyfogle, K. L. ; Blood, D. L. ; Rosnik, A. M. ; Krueger, B. P. Molecular Dynamics Force Field Parameters for the EGFP Chromophore and Some of Its Analogues. *The Journal of Physical Chemistry B* **2023**, 127 (26), 5772–5788. <https://doi.org/10.1021/acs.jpcb.3c01486>
- (10) Mantina, M. ; Chamberlin, A. C. ; Valero, R. ; Cramer, C. J. ; Truhlar, D. G. Consistent Van Der Waals Radii for the Whole Main Group. *Journal of Physical Chemistry A* **2009**, 113 (19), 5806–5812
- (11) Hubbard, R. E. ; Haider, M. K. Hydrogen Bonds in Proteins: Role and Strength. *Encyclopedia of life sciences* **2010**, 1, 1–6
- (12) McDonald, I. K. ; Thornton, J. M. Satisfying Hydrogen Bonding Potential in Proteins. *Journal of molecular biology* **1994**, 238 (5), 777–793
- (13) Liu, Z. ; Wang, G. ; Li, Z. ; Wang, R. Geometrical Preferences of the Hydrogen Bonds on Protein- Ligand Binding Interface Derived from Statistical Surveys and Quantum Mechanics Calculations. *Journal of Chemical Theory and Computation* **2008**, 4 (11), 1959–1973