An Alternative Strategy for Spectral Tuning of Flavin-binding Fluorescent Proteins

Mohammad Pabel Kabir,[1] Daniel Ouedraogo,[1] Yoelvis Orozco-Gonzalez,[1,*] Giovanni Gadda,[1,2,3*] Samer Gozem[1,*]

[1]Department of Chemistry, Georgia State University, Atlanta, GA 30302, United States.

[2]Department of Biology, Georgia State University, Atlanta, GA 30302, United States.

[3]The Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA 30302, United States.

*Correspondence to: yoelvis.orozco@gmail.com, ggadda@gsu.edu, sgozem@gsu.edu

Running title: iLOV-Q430E, a red shifting mutant of iLOV protein

Keywords: Flavoprotein, flavin, flavin mononucleotide (FMN), fluorescence, quantum chemistry, molecular dynamics, photoreceptor, Light-Oxygen-Voltage (LOV) domains.

Abstract

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

Introduction

Fluorescent proteins (FPs) have been used as tags for biosensing and bioimaging applications for over two decades in molecular virology and medicine.(1-8) The most widely used FPs are derived from the green fluorescent protein (GFP), and their spectral properties have been extensively studied experimentally and computationally.(8-16) However, GFPs have some limitations; they require molecular oxygen and produce hydrogen peroxide during chromophore maturation.(17-19) They are also ineffective genetic tags in small viruses that cannot handle the genetic load (GFP's molecular weight is around 22 kD). For such cases, flavin-binding fluorescent proteins (FbFPs),(20-29) like those derived from light, oxygen, and voltage sensing (LOV) domains,(30) are an attractive alternative because of their smaller size (10 kD). LOV domains non-covalently bind flavin mononucleotide (FMN), which is readily available *in vivo* and does not require any chemical maturation reaction.(21) iLOV is a recently engineered FbFP derived from the DNA shuffling of phototropin LOV1 and LOV2 domains.(21,22) Unlike wild-type LOV domains, iLOV does not contain a cysteine residue near the FMN cofactor and is consequently unable to form the cysteinyl adduct in the excited state that initiates the LOV domain photocycle.

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

There are two main strategies currently used for spectral tuning of FbFPs. The first approach is a chemical modification of the fluorophore to modify its electronic structure and, therefore, its spectroscopic properties.(37-41) However, chemically modifying FMN involves synthesizing and loading the chromophore in the protein. This is difficult to achieve *in vivo,* where natural flavin derivatives are instead readily available. The second, more convenient approach would be modulating the natural FMN fluorophore's electronic energies by modifying the surrounding protein (i.e., by mutagenesis).

Several computational and experimental studies in recent years focused on the spectral tuning of iLOV through protein point mutations, with recent attempts primarily focused on attempting to redshift the absorption. Khrenova and coworkers first recognized that placing a positive charge near flavin's N5 and C4a atoms would redshift iLOV's absorption and emission wavelength (see **Fig. 1** for flavin atom labels).(42) They proposed a Q489K single-point mutation, reasoning that the positively charged amino group (Lys) near flavin's N5 would stabilize its excited state π-electron system more than in the ground state, resulting in a redshift. QM/MM calculations supported their hypothesis. However, Davari and coworkers computationally and experimentally showed that the Q489K lysine side chain flips away from the N5 and C4a atoms of the chromophore, resulting in a blueshift in the absorption and emission instead of a redshift.(43) In a follow-up QM/MM study, Khrenova et al. proposed additional mutations to stabilize the lysine side chain close to the N5 and C4a atoms of the chromophore.(44) Recently, Wehler and coworkers experimentally attempted to prepare these mutants but could not prepare a functional red-shifted FbFP, as they found that a double-point mutant (iLOV-L470T/Q489K) gives ~2 nm blueshift and a triple-point mutation (iLOV-V392K/F410V/A426S) lost the ability to bind the chromophore due to the V392K mutation.(45) Overall, while the strategy of placing a positive charge in the vicinity

of flavin's N5 and C4a atoms was theoretically shown to work, most of the amino acids on that side of the protein turned out to be conformationally unstable or essential for chromophore binding. However, there has been recent success preparing a red-shifted iLOV with a double point mutation, iLOV-V392T-Q489K.(23) Red-shifted FbFPs were also derived from the thermostable protein CagFbFP.(23,46)

Our group recently reported electrostatic spectral tuning maps (ESTMs)(47,48) and flavin-solvent hydrogen bonding interactions.(49) Those serve as a starting point to find suitable mutations for spectral tuning. Here, based on these ESTMs, we suggest an alternative mutagenesis approach to redshift the absorption of iLOV; instead of focusing on placing a positive charge near C4a or N5, we introduce a negatively charged amino acid in the vicinity of flavin's N1 atom through a Q430E single-point mutant. We first test the effect of this mutation using hybrid QM/MM calculations employing the average solvent electrostatic configuration (ASEC) free energy gradient (FEG) approach. We then express iLOV-Q430E in the lab and measure its absorption, excitation, and fluorescence spectra to verify if they are red-shifted relative to iLOV.

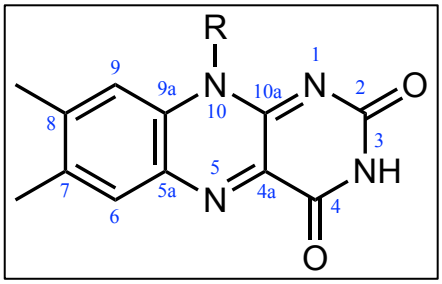

Figure 1. The isoalloxazine ring of FMN and atom number labels. R=CH_3 for lumiflavin and R= ribose-5′-phosphate for FMN.

Results and Discussion

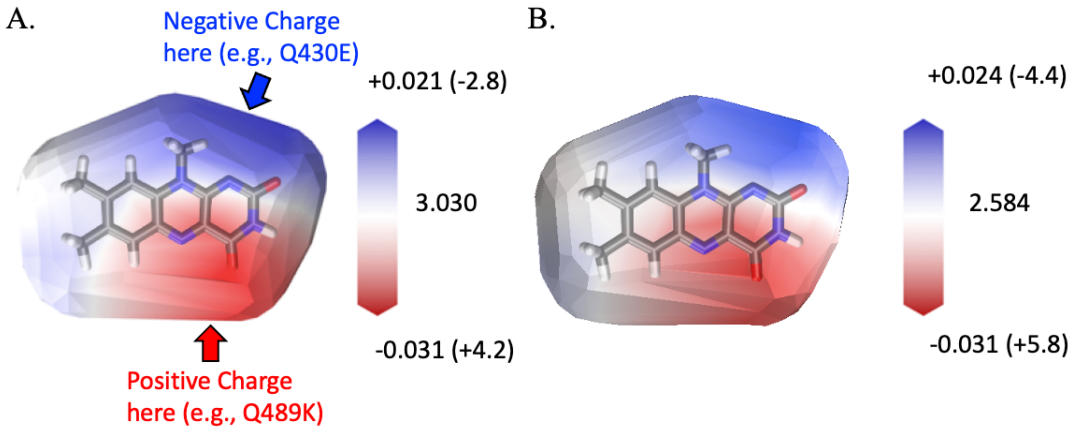

Figure 2. A. An ESTM map reproduced from ref. (48). The map indicates the change in the vertical excitation energy between the ground state (S_0) and the first singlet excited state (S_1) introduced by a +0.1-probe charge placed at the van der Waals surface of lumiflavin in its S_0 equilibrium geometry. The map suggests mutations that redshift the excitation energy (see arrows and labels). The legend indicates the magnitude of the excitation energy shifts relative to the gas-phase reference excitation energy in eV (and in nm in parentheses). **B.** The same ESTM map was computed at the S_1-optimized lumiflavin geometry, corresponding to the fluorescent minimum. The legend indicates the magnitude of the emission energy shifts relative to the gas-phase reference emission energy in eV (and in nm in parentheses).

ESTMs and charge analysis. ESTMs for flavin were reported recently.(47,48) These maps are intuitive visual tools that indicate how external positive or negative charges in the vicinity of a molecule influence its absorption or emission spectra. Briefly, ESTMs are constructed by moving a point charge on the van der Waals surface of the molecule and calculating the change in excitation and emission energies. The ESTM for the first singlet excited state (S_1) of flavin, which is experimentally at ~448 nm in iLOV, is shown in **Fig. 2A**. Here, we also recomputed the ESTM at the excited-state optimized flavin geometry to map how the fluorescence energy, which is

experimentally at ~500 nm in iLOV, is modified by nearby point charges (**Fig. 2B**). Both ESTMs were computed using time-dependent-density functional theory (TD-DFT) with the B3LYP functional(50) and cc-pVTZ basis set.(51)

The red and blue colored regions of the ESTMs in **Fig. 2** indicate that a spectral shift can be achieved if there is an electrostatic potential change in those regions. Both **Fig. 2A** and **2B** have similar features: a red region near C4a and N5 atoms, a blue region near the N1 atom, and a white region near the xylene portion of flavin. This indicates that (a) a positive charge near the C4a or N5 atoms would redshift the absorption/emission, (b) a positive charge near N1 would blueshift the absorption/emission, and (c) charges introduced near the xylene portion of the flavin would have a negligible effect on the spectral properties for this state. A negative probe charge would have the exact opposite effect.(47) These calculations are largely consistent with transition dipole moment measurements in flavins.(52,53) While the excitation energy ESTM (**Fig. 2A**) and emission ESTM (**Fig. 2B**) appear very similar, the magnitude of the shift reported in the plot legends reveals a subtle difference: the emission energy (in eV) is more sensitive to the presence of negative charges near the N1 flavin atom than the corresponding absorption energy.

The ESTMs in **Fig. 2** were used to generate strategies for spectral tuning lumiflavin. The maps indicate that a positive charge near the C4a or N5 flavin atoms would lead to a red-shifted absorption/emission; this has been the strategy proposed by Khrenova et al. and the ensuing computational and experimental work.(23,41-45) A second strategy, not yet explored in FbFPs, would be to introduce a negatively charged amino acid or negative dipole in the vicinity of the N1 atom of flavin. Here, we pursue this strategy with Q430E.

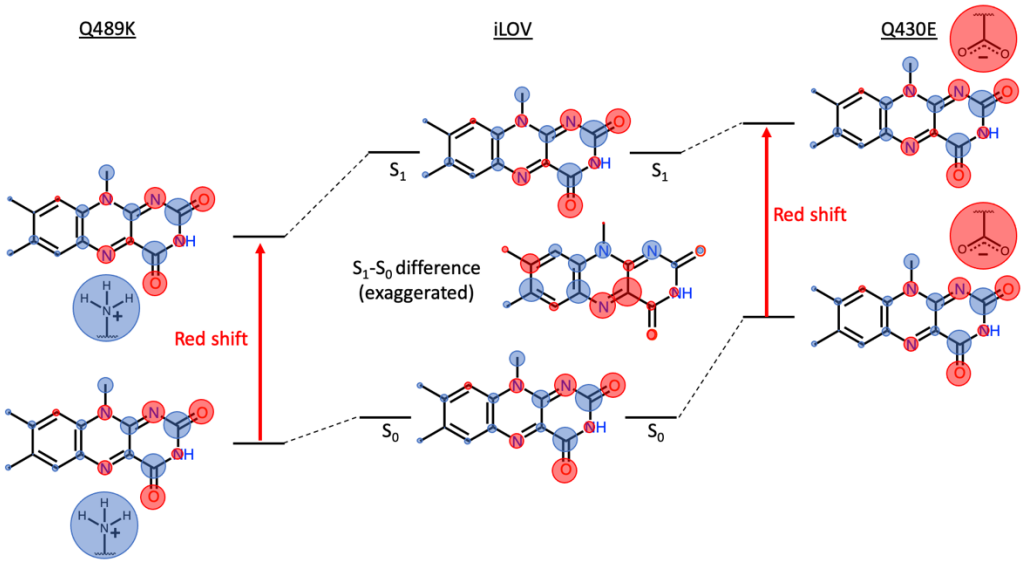

Figure 3. Center: LoProp charge population analysis (with hydrogen atom charges summed onto the heavy atoms they are connected to) for lumiflavin in its ground (S_0) and first singlet excited (S_1) state. Red circles indicate negative charge density, and blue circles indicate positive charge density. The area of the circles is directly proportional to the charge on the corresponding atom. The S_1 and S_0 states have slightly different charge distributions that are difficult to discern without close inspection. Therefore, the difference in the atomic charges (S_1–S_0) is also shown in the middle. In the S_1–S_0 difference plot, the areas of the circles are proportional to the magnitude of the charge difference between S_0 and S_1, with red circles indicating reduced charge (higher electron density) on the atom after excitation from S_0 to S_1 and blue circles indicating increased charge (lower electron density). **Left:** A scheme illustrating the effect of placing a positively charged lysine side chain close to the flavin N5/C4a, which would result in a redshift (see text for details). **Right:** A scheme illustrating the effect of placing a negatively charged glutamate side chain close to the flavin N1, which should also result in a redshift (see text for details).

To better understand the electronic structure changes underlying the electrostatic spectral tuning properties of flavin, we computed atomic charges from both the ground and excited-state wave functions of a lumiflavin gas-phase model (**Fig. 3**). The charges were obtained using LoProp population analysis from CASSCF(10,10)/ANO-L-VDZP wave functions. The LoProp approach provides physically meaningful localized properties and mitigates issues like basis set dependence sometimes encountered with other population analysis methods.(54)

Exciting flavin from the S_0 to the S_1 state changes the electron distribution along the π-conjugated isoalloxazine ring. This effect is subtle but can be best visualized by plotting the change in the charges at each atomic center from S_0 to S_1 (**Fig. 3** center). Specifically, there is an increase in electron density at the C4a and N5 flavin atoms and a decrease in electron density at the N1 flavin atom and several atoms in the xylene portion of the flavin. There is also a slight decrease in electron density at the C2=O carbonyl. The charge redistribution is perfectly consistent with the ESTMs in **Fig. 2**; the increased charge density at the C4a/N5 flavin atoms means there is potential for spectral tuning by placing a positive charge nearby. Such a positive charge, e.g., a protonated lysine side chain, would stabilize the excited S_1 state slightly more than the ground S_0 state, leading to a redshift in the excitation energy (**Fig. 3** left). Conversely, there are several atoms where the electron density decreases upon excitation to S_1. In most cases, those atoms are shielded from external charges by methyl groups or hydrogen atoms, which explains the less intense color of the ESTM map near the C7, C9, C9a, and N10 atoms despite the decrease in the electron density on those atoms. However, the N1 atom is exposed, allowing charged amino acids to approach and creating an opportunity for spectral tuning at that site. The decreased electron density on the N1 flavin atom means that a negative charge nearby, e.g., a glutamate side chain, would destabilize the ground state more than the excited state, decreasing the S_0–S_1 energy gap and resulting in a red-shifted absorption (**Fig. 3** right).

The approach shown on the right of **Fig. 3** (placing a negatively charged amino acid near flavin's N1 atom) seems less desirable than the approach on the left (placing a positively charged amino acid near flavin's C4a/N5 atoms) from a bioengineering standpoint, since it relies on an unfavorable interaction between flavin and a negatively charged residue. However, given the

limited success with spectral tuning at the C4a/N5 site, we attempted spectral tuning with a negatively charged residue with the Q430E single-point mutation.

QM/MM simulations of iLOV and iLOV-Q430E. The crystal structure of iLOV indicates that Q430 is 3.4 Å away from the N1 atom of FMN. Therefore, we chose to replace Q430 with isosteric glutamic acid. Since glutamic acid has a pK_a of 4.07 in solution and Q430 has polar residues nearby, we anticipated that the mutated glutamic acid Q430E would be deprotonated and introduce a negative charge near the flavin's N1 atom without causing a significant structural change in the protein. To test this hypothesis, we performed QM/MM geometry optimizations followed by excited-state energy calculations for both iLOV and iLOV-Q430E using the ASEC-FEG method, as outlined in the computational approach section.

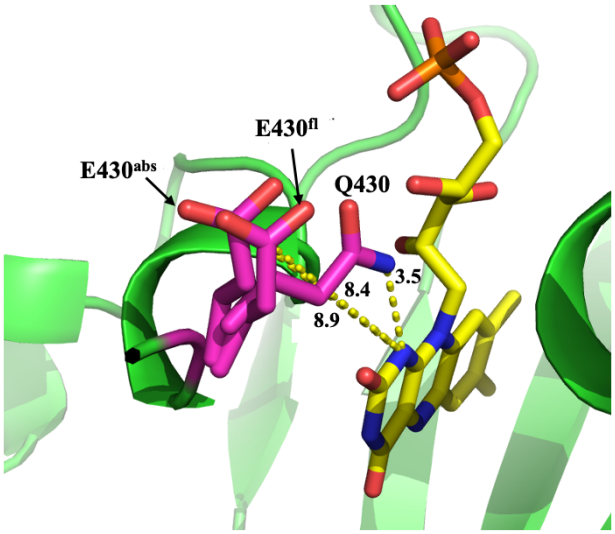

Figure 4. This figure shows three representatives QM/MM optimized snapshots to show the distance between the flavin N1 and the Q430 (iLOV) or E430 (iLOV-Q430E) side chain. The labeled distances are reported averages from one of the production runs of the MD simulations. Specifically, we label the distance between the flavin N1 and the Q430 side chain nitrogen in iLOV (3.5 Å), the E430 side chain carboxylate carbon in the ground-state optimized system (E430^{abs}, 8.9 Å), and the E430 side chain carboxylate carbon in the excited-state optimized system (E430^{fl}, 8.4 Å). The figure was prepared using PyMol.(55)

In snapshots obtained from molecular dynamics (MD) simulations of iLOV, the average computed distance of the Q430 nitrogen from the flavin N1 atom is 3.5 Å (Q430 in **Fig. 4**), in good agreement with the crystal structure (3.4 Å). In contrast, MD simulations of iLOV-Q430E revealed that the E430 glutamate side chain flips away from flavin and maintains an average distance of 8.9 Å from the flavin N1 atom ($E430^{abs}$ in **Fig. 4**). This conformational change is likely driven by a lack of hydrogen bonding with neighboring amino acids, which causes E430 to point outwards towards the surface of the protein.

Next, we proceeded to calculate the vertical excitation energy. The iLOV and iLOV-Q430E calculations reached self-consistency quickly after just one step of the ASEC-FEG cycle. After that, the following four steps, each involving MD to regenerate the ASEC environment, re-optimizing the flavin chromophore, and re-computing the vertical excitation energy, yielded similar results within 1 nm of each other. These calculations indicated that the vertical excitation energy of iLOV-Q430E is 4 nm red-shifted compared to iLOV (**Table S1**).

In a recent joint computational and experimental study,(56) we found that solution ions may affect the outcome of ASEC-FEG calculations, especially when there are charged residues inside the active site of a flavoprotein. Therefore, we repeated the calculations for both iLOV and iLOV-Q430E after adding 4 pairs of Na+ and Cl- solution ions, approximately 1 NaCl per 2775 water solvent molecules, equivalent to 20 mM of salt used in the experiments in this work. These calculations showed that the Q430E mutation has almost no effect on the absorption spectrum of iLOV, causing a shift of less than 0.5 nm (**Table S2**).

The MD and ASEC-FEG simulations show that the E430 does not remain in the same position as Q430 and causes just a slight 0-4 nm redshift in the vertical excitation energy relative to iLOV, which was initially discouraging. However, given the difficulty in red-shifting iLOV

even by a few nm, we decided to proceed with expressing iLOV-Q430E and iLOV experimentally to compare their absorption and emission properties.

Experimental absorption and emission spectra of iLOV and iLOV-Q430E. iLOV and iLOV-Q430E were expressed and purified successfully. The absorption and excitation/emission spectra were measured and are shown in **Fig. 5** and **Fig. 6**, respectively. The data are also tabulated in **Table 1**. We found that iLOV-Q430E gave a modest 1-2 nm redshift in the first excited state absorption wavelength, consistent with our computational prediction of a weak redshift. However, the fluorescence wavelength is shifted to red by 4-8 nm compared to the emission in iLOV. This is comparable to the shift recently achieved in CagFbFP.(23)

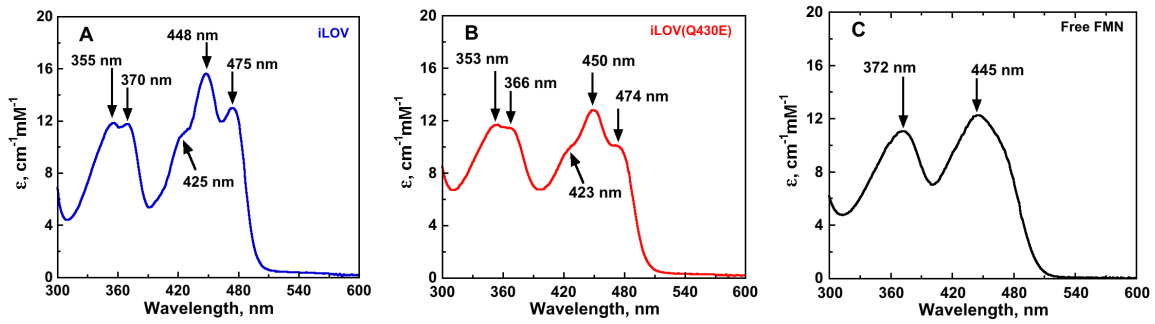

Figure 5. UV-visible absorption spectra of iLOV (panel A, blue), iLOV-Q430E (panel B, red), and free FMN (Panel C, black). The spectra were recorded in 20 mM pH 7.0 phosphate buffer solution and 15 °C.

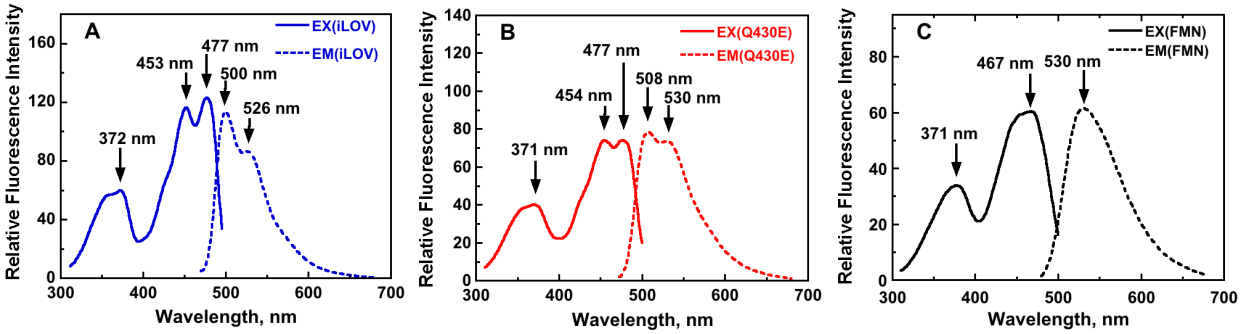

Figure 6. The excitation and emission spectra of iLOV (panel A, blue), iLOV-Q430E (panel B, red), and free FMN (panel C, black). The spectra were recorded in 20 mM pH 7.0 phosphate buffer and 15 °C.

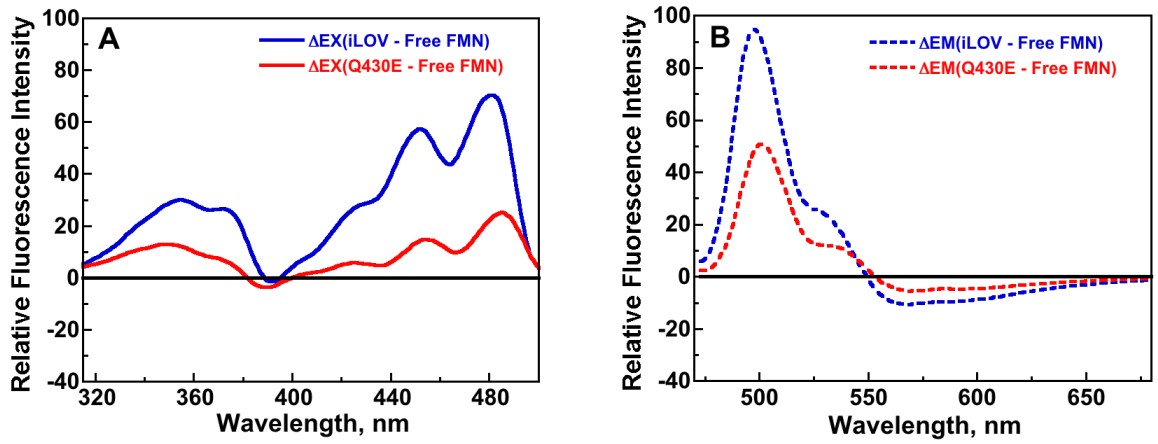

Figure 7. A. Difference excitation spectra of iLOV and iLOV-Q430E minus that of free FMN. **B.** Difference emission spectra of the iLOV and iLOV-Q430E minus that of free FMN. The reference line in panels A and B represents the ΔEX (free FMN-Free FMN) and ΔEM (free FMN-Free FMN), respectively.

Table 1: Experimental absorption and fluorescence properties of iLOV proteins.

	iLOV	iLOV-Q430E	Free FMN
Maximal absorbance, nm	370, 448	366, 450	372, 445
Extinction coefficient, $mM^{-1}cm^{-1}$	11.9, 15.6	11.8, 12.8	10.0, 12.5
Fluorescence emission, nm[b]	500 ± 1, 526 ± 1[c]	508 ± 1, 530 ± 1	530 ± 1
Fluorescence intensity	113 ± 1, 86 ± 1[c]	78 ± 1, 74 ± 1	61 ± 1
UV-visible and fluorescence data were recorded in 20 mM pH 8.0 phosphate buffer solution and 15 °C. [b] Excited at the low energy peak of the flavin. [c] Standard errors refer to the average of three independent measurements.			

Next, we investigated the effect of the Q430E mutation on the fluorescence brightness by measuring the relative absorbance and fluorescence intensity of iLOV and iLOV-Q430E at the same concentration (**Fig. 7**). We found that the absorbance and fluorescence intensities in iLOV-Q430E are reduced to about 50% compared to iLOV.

The experimental spectra indicate that the fluorescence wavelength is more sensitive to the Q430E point mutation than the absorption. This may be partly explained by comparing the excitation and emission ESTMs in **Fig. 2A** and **Fig. 2B**, respectively. The emission ESTM indicates a higher sensitivity of the emission energy compared to the absorption. However, to

check if we can reproduce this effect in the QM/MM calculations, we repeated the ASEC-FEG calculations for both iLOV and iLOV-Q430E using the charges and gradient of the S_1 excited state of flavin instead of the S_0 ground state. We computed the $S_1 - S_0$ vertical emission energy at the excited state S_1 geometry. Note that, due to the use of excited-state charges for flavin during the MD calculations, the protein adapts to the excited-state charge distribution. The 5 ns MD calculations have a similar timescale as a typical fluorescence lifetime, giving the protein a reasonable time to rearrange around the excited state configuration. In this case, it took slightly longer to achieve self-consistency of the ASEC-FEG calculations, so the first three steps were discarded, and the excitation energy was averaged over the next four ASEC-FEG steps. The ASEC-FEG calculations indicate that iLOV-Q430E has a 9 nm red-shifted vertical emission wavelength compared to iLOV. Inspecting the MD simulations revealed that, on average, the E430 moves closer to the flavin in the excited state compared to the ground state. The average distance between flavin's N1 atom and the C-atom of the side chain carboxylate ion is 8.4 Å for the excited state, compared to 8.9 Å in the ground state geometry ($E430^{abs}$ and $E430^{fl}$ in **Fig. 4**). Therefore, the more significant Stokes shift in iLOV-Q430E compared to iLOV can be attributed to two factors; one is electronic since the ESTM already shows a higher sensitivity of flavin's S_0–S_1 energy difference to charges near the N1 atom after flavin relaxes on its S_1 potential energy surface to the fluorescent minimum (**Fig. 2**). The second effect comes from the protein; the deprotonated E430 is less repelled by the flavin excited state than the ground state and moves slightly closer, on average, to the flavin chromophore after it is excited to S_1.

Conclusions

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

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

experiments to verify the results of the calculations. Conversely, the experiments often bring up new observations and questions for the calculations to answer, as was the case here for the more significant Stokes shift observed in iLOV-Q430E compared to iLOV.

Experimental Procedures

Bacterial strains and plasmids. The iLOV and iLOV-Q430E were obtained as synthetic genes from GenScript (Inc., Piscataway, NJ, USA) and were flanked with a 5'- NdeI and a 3'- XhoI restriction endonuclease recognition sites in a pET20b(+) plasmid. The pET plasmid harboring the iLOV and iLOV-Q430E genes contains an N-terminal His_6-tag fused to the target proteins to facilitate heterologous expression in *E. coli* and affinity chromatography purification. The genes were transformed into *E. coli* strain DH5α and Rosetta(DE3)pLysS competent cells for storage and expression, respectively. The resulting plasmids were verified by sequencing (Psomagen, Inc., Rockville, MD, USA), and permanent stocks of the cells were prepared and stored at -80 °C.

Protein expression and purification. Permanently frozen stocks of *E. coli* cells Rosetta(DE3)pLysS harboring iLOV or iLOV-Q430E genes were used to inoculate 100 mL of Luria-Bertani broth medium containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol, and cultures were grown at 37 ºC overnight to be used as a preculture. A 10 mL portion of preculture was used to inoculate 1.0 L of Luria-Bertani broth medium containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol. When the cultures reached optical densities of ~0.6 at 600 nm, the temperature was lowered to 18 °C, and Isopropyl-thio-galactoside (IPTG) was added to a final concentration of 0.1 mM. After 18 h, the cells were harvested by centrifugation at 5000g for 20 min at 4°C. All purification steps were carried out at 4 °C. The wet cell paste was suspended in 0.1 mM PMSF, 0.2 mg/mL lysozyme, 10 % glycerol, and 50 mM pH 8.0 phosphate buffer

solution, containing 300 mM NaCl, 10 mM Imidazole, and 10 % glycerol in a ratio of 1 g of wet cell paste to 4 mL of lysis buffer. The suspended cells were then allowed to incubate with stirring for 30 min on ice with 5 μg/mL RNase and 5 μg/mL DNase in the presence of 10 mM $MgCl_2$. The resulting slurry was sonicated for 60 cycles of 20 seconds with the pulse on and 10 seconds with the pulse off for 20 min. The cell debris was removed by centrifugation at 10000 x g for 20 min. The supernatant was loaded onto a 5 mL Ni-NTA column (GE Healthcare), equilibrated with 50 mM pH 8.0 phosphate buffer solution, 300 mM NaCl, 10 mM Imidazole, and 10 % glycerol. The proteins were purified with gradient elution from 10 to 250 mM imidazole in 50 mM pH 8.0 phosphate buffer solution, 300 mM NaCl, and 10 % glycerol buffer. The eluted fractions containing the iLOV protein were dialyzed against four changes of 10 mM pH 8.0 phosphate buffer solution, 10 mM NaCl, and 10 % glycerol. After the dialysis, the proteins were centrifuged at 10000 x g for 20 min to remove any precipitated protein. The iLOV and the iLOV-Q430E were then stored at -20 ºC.

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

UV-Visible Absorption and Fluorescence Spectroscopy. The UV-visible absorption spectra of iLOV and iLOV-Q430E were recorded with an Agilent Technologies model HP 8453 PC diode-array spectrophotometer equipped with a thermostated water bath. The proteins were prepared fresh by gel filtration through PD-10 desalting columns (General Electric, Fairfield, CT) just before being used. The extinction coefficients of the enzyme-bound FMN to the iLOV protein

were determined in 20 mM pH 7.0 phosphate buffer solution after incubation of the protein with 4 M urea at 40 °C for 1 h, based upon an ε450 value of 12.2 mM^{-1} cm^{-1} for free FMN and the method published by Whitby et al.(57) The fluorescence emission spectra of the iLOV and iLOV-Q430E variant protein were recorded in 20 mM pH 8.0 phosphate buffer solution at 15 °C, with a Shimadzu model RF-5301 PC spectrofluorometer using a 1 cm path length quartz cuvette. All fluorescence spectra were corrected by subtracting the corresponding blanks to account for Rayleigh and Raman scattering. The samples at a concentration of 10 μM protein-bound flavin were excited at the low-energy peak of the UV-visible absorption spectrum, and emission scans were determined from 475 to 600 nm.

Computational approach. The ASEC-FEG method builds on the average solvent electrostatic configuration (ASEC) approach developed by Canuto and coworkers(58) and approximates the FEG approach from Okuyama-Yoshida et al.,(59) which is rooted in the free energy perturbation theory.(60) The ASEC-FEG approach was first extended to proteins by Orozco-Gonzalez et al. for rhodopsins.(61) We recently developed ASEC-FEG for flavoproteins.(56,62) With ASEC, the quantum chemical calculations are performed in the field of a time-averaged electrostatic potential environment of the protein and solution (collectively referred to as a "solvent" in the ASEC acronym). Effectively, the protein and solution are represented as a "superposition" of structures obtained from MD simulations. This approach leaves the representation of rigid atoms intact while flexible atoms are replaced by a cloud of charges over the space sampled during the dynamics. Flexible atoms also have a broader and shallower Lennard Jones potential (see ref. (56) for more details). The optimization of the QM system within the ASEC configuration is done self-consistently. ASEC is well suited for averaging the effect of long-range charge interactions that are difficult to capture with more traditional QM/MM methods.(56)

The ASEC-FEG protocol is a series of scripts building on an existing QM/MM interface between the OpenMolcas(63) and Tinker(64) software packages.(65,66) The protocol guides users through the model construction starting from the PDB file of the protein and culminating in the generation of an ASEC QM/MM model. The protocol calls on several other software: PropKa 3.1(67), Dowser(68), SCWRL4.0(69), and Gromacs.(70) Gromacs is used to add hydrogen atoms to the PDB, solvate the protein, and run MD simulations to equilibrate the system and sample the protein around the cofactor. The OpenMolcas(63)/Tinker(64) interface uses an additive QM/MM scheme that includes Lennard Jones and electrostatic interactions through the ElectroStatic Potential Fitted (ESPF) approach.(71) The automation of this protocol, done in the same vein as efforts to automate the construction of QM/MM models for rhodopsins by Olivucci and coworkers,(72-74) mitigates problems with reproducibility of QM/MM calculations and allows the systematic investigation of closely related proteins using a consistent approach.

The initial coordinates of iLOV were taken from the X-ray structure PDB 4EES (resolution: 1.8 A).(22) Parameters for FMN were initially retrieved from the AMBER parameter database maintained by the University of Manchester.(75) The Q430E mutation was introduced in the protein structure and modeled using SCWRL4.(69) Dowser was used to remove non-bonded water molecules from the crystallographic structure.(68) The total charge of the systems was neutralized by adding solution counterions. This model was used as a starting point for MD simulations and the generation of the ASEC environment. The MD calculations for iLOV and iLOV-Q430E were performed with periodic boundary conditions in a 7.0 nm x 7.0 nm x 7.0 nm cubic solvent box. Geometry minimization and MD simulations were carried out using GROMACS.(70) The AMBER99SB(76,77) and TIP3P(78) force fields were used for protein and water, respectively. During each step of the ASEC-FEG cycle, the MD calculations were

performed in three phases: the system was first gradually heated from 0 to 300 K at 1 atm pressure over 300 ps. This was followed by 4700 ps of equilibration and 5000 ps of production simulations carried out with the NPT ensemble under standard ambient temperature and pressure. The ASEC configuration of the protein was formed by sampling 100 configurations from the production part of the MD. The ASEC GROMACS file is then converted to Tinker format for QM/MM calculations.

For QM/MM calculations, the protein was divided into two subsystems (**Fig. 8**): (i) the QM region, comprising the lumiflavin (structure shown in **Fig. 1**), and (ii) the MM region, which includes all other atoms in the simulation (the ribose-5′-phosphate group, the protein, the solvent, and solution ions). The frontier between the QM and the MM parts is treated using the hydrogen link atom (LA, **Fig. 8**). The charges for the MM atoms near the LA are set to zero and distributed over other MM atoms to avoid over polarizing the QM wavefunction. The QM subsystem is then optimized in the presence of a frozen ASEC MM environment with electrostatic embedding. Using the updated geometry and updated ESPF charges of the QM subsystem, another MD calculation is run for 5 ns to generate a new ASEC configuration. This process is repeated for several steps until the computed excitation energies stay consistent for four consecutive steps (i.e., within 0.02 eV nm of the four-step moving average). We then took the average excitation energies from those four steps and used them to compute the wavelength shift.

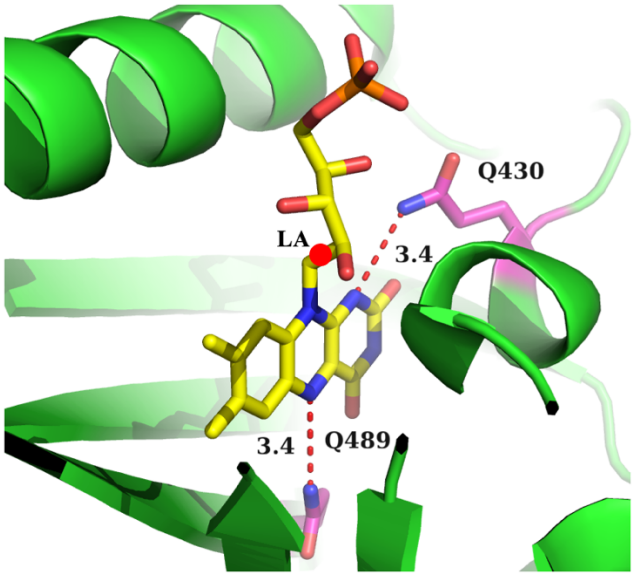

Figure 8. FMN inside the binding pocket of iLOV from PDB 4EES. The dash lines indicate the distance in angstrom between the Q489 and Q430 glutamine side chain nitrogen atoms and the closest FMN nitrogen atom. The red circle indicates the hydrogen Link Atom (LA), which separates the QM subsystem (the lumiflavin) and the MM subsystem (the ribose-5′-phosphate group, protein, and solvent). The figure was prepared using PyMol.(55)

The geometry optimization of the QM subsystem was performed using the complete-active-space self-consistent field (CASSCF) level of theory and the ANO-L-VDZP basis set. We tested the effect of the active space on the first excited state (π-π*) excitation energy of flavin using gas-phase benchmark calculations; we found that there is a limited benefit to increasing the active space beyond 10 electrons and 10 orbitals (five π, five π*). Therefore, QM/MM geometry optimizations were performed using CASSCF (10,10). State averaging was not used for ground state optimizations, while 2-root state averaging was used for excited-state optimizations. Excitation energies were computed using the complete active space second-order perturbation theory (CASPT2) with the ANO-L-VDZP basis set. CASPT2 calculations were performed using the Cholesky decomposition(79) and applying an imaginary level shift(80) of 0.2. An IPEA shift,(81) sometimes used for flavins for the purpose of error cancellation,(82,83) was not used

here. It was recently shown that the IPEA shift is not needed in cases where the dynamical electron correlation is adequately accounted for in CASPT2 calculations.(84) We also note that vertical excitation energy calculations typically underestimate the absorption wavelength of flavins compared to the experimental wavelength of maximal absorption; the calculation of the vibronic progression from Franck-Condon factors is needed for better quantitative agreement between theory and experimental spectra.(48,85-87)

Supporting Information

This article contains supporting information.

Acknowledgements

M.P.K. and D.O. acknowledge a fellowship from the Molecular Basis of Disease Program at Georgia State University. M.P.K. also acknowledges support from the Provost Dissertation Fellowship at Georgia State University. This material is based upon work supported by the National Science Foundation (NSF) under Grant No. CHE-2047667 (to S.G.) and CHE-1506518 (to G.G.). We acknowledge NSF XSEDE for computational resources through Research Allocation CHE180027, and the Advanced Research Computing Technology and Innovation Core (ARCTIC) resources, which are supported by the NSF Major Research Instrumentation (MRI) grant number CNS-1920024.

Conflict of Interest

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Baulcombe, D. C., Chapman, S., and Santa Cruz, S. (1995) Jellyfish green fluorescent protein as a reporter for virus infections. *The Plant Journal* **7**, 1045-1053
2. Shaner, N. C., Patterson, G. H., and Davidson, M. W. (2007) Advances in fluorescent protein technology. *Journal of cell science* **120**, 4247-4260
3. Zacharias, D. A., Baird, G. S., and Tsien, R. Y. (2000) Recent advances in technology for measuring and manipulating cell signals. *Current opinion in neurobiology* **10**, 416-421
4. Chudakov, D. M., Matz, M. V., Lukyanov, S., and Lukyanov, K. A. (2010) Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological reviews* **90**, 1103-1163
5. Misteli, T., and Spector, D. L. (1997) Applications of the green fluorescent protein in cell biology and biotechnology. *Nature biotechnology* **15**, 961-964
6. Remington, S. J. (2011) Green fluorescent protein: a perspective. *Protein Science* **20**, 1509-1519
7. Zimmer, M. (2002) Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chemical reviews* **102**, 759-782
8. Acharya, A., Bogdanov, A. M., Grigorenko, B. L., Bravaya, K. B., Nemukhin, A. V., Lukyanov, K. A., and Krylov, A. I. (2017) Photoinduced chemistry in fluorescent proteins: Curse or blessing? *Chemical reviews* **117**, 758-795
9. Delagrave, S., Hawtin, R. E., Silva, C. M., Yang, M. M., and Youvan, D. C. (1995) Red-shifted excitation mutants of the green fluorescent protein. *Bio/technology* **13**, 151-154
10. Molina, R. S., Tran, T. M., Campbell, R. E., Lambert, G. G., Salih, A., Shaner, N. C., Hughes, T. E., and Drobizhev, M. (2017) Blue-shifted green fluorescent protein homologues are brighter than enhanced green fluorescent protein under two-photon excitation. *The journal of physical chemistry letters* **8**, 2548-2554
11. Kamarchik, E., and Krylov, A. I. (2011) Non-Condon effects in the one-and two-photon absorption spectra of the green fluorescent protein. *The Journal of Physical Chemistry Letters* **2**, 488-492
12. Filippi, C., Buda, F., Guidoni, L., and Sinicropi, A. (2012) Bathochromic shift in green fluorescent protein: a puzzle for QM/MM approaches. *Journal of chemical theory and computation* **8**, 112-124
13. Kaila, V. R., Send, R., and Sundholm, D. (2013) Electrostatic spectral tuning mechanism of the green fluorescent protein. *Physical Chemistry Chemical Physics* **15**, 4491-4495
14. Sinicropi, A., Andruniow, T., Ferré, N., Basosi, R., and Olivucci, M. (2005) Properties of the emitting state of the green fluorescent protein resolved at the CASPT2//CASSCF/CHARMM level. *Journal of the American Chemical Society* **127**, 11534-11535
15. Tang, S., Wong, H. C., Wang, Z. M., Huang, Y., Zou, J., Zhuo, Y., Pennati, A., Gadda, G., Delbono, O., and Yang, J. J. (2011) Design and application of a class of sensors to monitor Ca2+ dynamics in high Ca2+ concentration cellular compartments. *Proc Natl Acad Sci U S A* **108**, 16265-16270
16. Deng, X., Yao, X. Q., Berglund, K., Dong, B., Ouedraogo, D., Ghane, M. A., Zhuo, Y., McBean, C., Wei, Z. Z., Gozem, S., Yu, S. P., Wei, L., Fang, N., Mabb, A. M., Gadda, G., Hamelberg, D., and Yang, J. J. (2021) Tuning Protein Dynamics to Sense Rapid Endoplasmic-Reticulum Calcium Dynamics. *Angew Chem Int Ed Engl* **60**, 23289-23298

17. Zhang, L., Patel, H. N., Lappe, J. W., and Wachter, R. M. (2006) Reaction progress of chromophore biogenesis in green fluorescent protein. *J Am Chem Soc* **128**, 4766-4772
18. Ganini, D., Leinisch, F., Kumar, A., Jiang, J., Tokar, E. J., Malone, C. C., Petrovich, R. M., and Mason, R. P. (2017) Fluorescent proteins such as eGFP lead to catalytic oxidative stress in cells. *Redox biology* **12**, 462-468
19. Kalyanaraman, B., and Zielonka, J. (2017) Green fluorescent proteins induce oxidative stress in cells: A worrisome new wrinkle in the application of the GFP reporter system to biological systems? *Redox Biol* **12**, 755-757
20. Drepper, T., Eggert, T., Circolone, F., Heck, A., Krauß, U., Guterl, J.-K., Wendorff, M., Losi, A., Gärtner, W., and Jaeger, K.-E. (2007) Reporter proteins for in vivo fluorescence without oxygen. *Nature biotechnology* **25**, 443-445
21. Chapman, S., Faulkner, C., Kaiserli, E., Garcia-Mata, C., Savenkov, E. I., Roberts, A. G., Oparka, K. J., and Christie, J. M. (2008) The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proceedings of the National Academy of Sciences* **105**, 20038-20043
22. Christie, J. M., Hitomi, K., Arvai, A. S., Hartfield, K. A., Mettlen, M., Pratt, A. J., Tainer, J. A., and Getzoff, E. D. (2012) Structural tuning of the fluorescent protein iLOV for improved photostability. *Journal of Biological Chemistry* **287**, 22295-22304
23. Röllen, K., Granzin, J., Remeeva, A., Davari, M. D., Gensch, T., Nazarenko, V. V., Kovalev, K., Bogorodskiy, A., Borshchevskiy, V., and Hemmer, S. (2021) The molecular basis of spectral tuning in blue-and red-shifted flavin-binding fluorescent proteins. *Journal of Biological Chemistry* **296**
24. Ko, S., Jeon, H., Yoon, S., Kyung, M., Yun, H., Na, J.-H., and Jung, S. T. (2020) Discovery of novel Pseudomonas putida flavin-binding fluorescent protein variants with significantly improved quantum yield. *Journal of agricultural and food chemistry* **68**, 5873-5879
25. Mukherjee, A., and Schroeder, C. M. (2015) Flavin-based fluorescent proteins: emerging paradigms in biological imaging. *Curr Opin Biotechnol* **31**, 16-23
26. Mukherjee, A., Walker, J., Weyant, K. B., and Schroeder, C. M. (2013) Characterization of flavin-based fluorescent proteins: an emerging class of fluorescent reporters. *PLoS One* **8**, e64753
27. Buckley, A. M., Petersen, J., Roe, A. J., Douce, G. R., and Christie, J. M. (2015) LOV-based reporters for fluorescence imaging. *Curr Opin Chem Biol* **27**, 39-45
28. Mishin, A. S., Belousov, V. V., Solntsev, K. M., and Lukyanov, K. A. (2015) Novel uses of fluorescent proteins. *Curr Opin Chem Biol* **27**, 1-9
29. Wingen, M., Jaeger, K. E., Gensch, T., and Drepper, T. (2017) Novel Thermostable Flavin-binding Fluorescent Proteins from Thermophilic Organisms. *Photochem Photobiol* **93**, 849-856
30. Christie, J. M., Salomon, M., Nozue, K., Wada, M., and Briggs, W. R. (1999) LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): binding sites for the chromophore flavin mononucleotide. *Proceedings of the National Academy of Sciences* **96**, 8779-8783
31. Potzkei, J., Kunze, M., Drepper, T., Gensch, T., Jaeger, K.-E., and Büchs, J. (2012) Real-time determination of intracellular oxygen in bacteria using a genetically encoded FRET-based biosensor. *BMC biology* **10**, 1-13
32. Liu, X., Jiang, L., Li, J., Wang, L., Yu, Y., Zhou, Q., Lv, X., Gong, W., Lu, Y., and Wang, J. (2014) Significant expansion of fluorescent protein sensing ability through the genetic

incorporation of superior photo-induced electron-transfer quenchers. *Journal of the American Chemical Society* **136**, 13094-13097
33. Shcherbakova, D. M., Shemetov, A. A., Kaberniuk, A. A., and Verkhusha, V. V. (2015) Natural photoreceptors as a source of fluorescent proteins, biosensors, and optogenetic tools. *Annual review of biochemistry* **84**, 519
34. Ravikumar, Y., Nadarajan, S. P., Lee, C. S., and Yun, H. (2017) Engineering an FMN-based iLOV protein for the detection of arsenic ions. *Anal Biochem* **525**, 38-43
35. Zhao, H., Zhang, Y., Pan, M., Song, Y., Bai, L., Miao, Y., Huang, Y., Zhu, X., and Song, C. P. (2019) Dynamic imaging of cellular pH and redox homeostasis with a genetically encoded dual-functional biosensor, pHaROS, in yeast. *J Biol Chem* **294**, 15768-15780
36. Wingen, M., Potzkei, J., Endres, S., Casini, G., Rupprecht, C., Fahlke, C., Krauss, U., Jaeger, K.-E., Drepper, T., and Gensch, T. (2014) The photophysics of LOV-based fluorescent proteins—new tools for cell biology. *Photochemical & photobiological sciences* **13**, 875-883
37. Salzmann, S., Martinez-Junza, V., Zorn, B. r., Braslavsky, S. E., Mansurova, M., Marian, C. M., and Gärtner, W. (2009) Photophysical properties of structurally and electronically modified flavin derivatives determined by spectroscopy and theoretical calculations. *The Journal of Physical Chemistry A* **113**, 9365-9375
38. Silva, M. R., Jr., Mansurova, M., Gartner, W., and Thiel, W. (2013) Photophysics of structurally modified flavin derivatives in the blue-light photoreceptor YtvA: a combined experimental and theoretical study. *Chembiochem* **14**, 1648-1661
39. Mansurova, M., Simon, J., Salzmann, S., Marian, C. M., and Gartner, W. (2013) Spectroscopic and theoretical study on electronically modified chromophores in LOV domains: 8-bromo- and 8-trifluoromethyl-substituted flavins. *Chembiochem* **14**, 645-654
40. Kar, R. K., Chasen, S., Mroginski, M.-A., and Miller, A.-F. (2021) Tuning the quantum chemical properties of flavins via modification at C8. *The Journal of Physical Chemistry B* **125**, 12654-12669
41. Meteleshko, Y. I., Nemukhin, A. V., and Khrenova, M. G. (2019) Novel flavin-based fluorescent proteins with red-shifted emission bands: a computational study. *Photochemical & Photobiological Sciences* **18**, 177-189
42. Khrenova, M. G., Nemukhin, A. V., and Domratcheva, T. (2015) Theoretical characterization of the flavin-based fluorescent protein iLOV and its Q489K mutant. *The Journal of Physical Chemistry B* **119**, 5176-5183
43. Davari, M. D., Kopka, B., Wingen, M., Bocola, M., Drepper, T., Jaeger, K.-E., Schwaneberg, U., and Krauss, U. (2016) Photophysics of the LOV-based fluorescent protein variant iLOV-Q489K determined by simulation and experiment. *The Journal of Physical Chemistry B* **120**, 3344-3352
44. Khrenova, M. G., Meteleshko, Y. I., and Nemukhin, A. V. (2017) Mutants of the Flavoprotein iLOV as Prospective Red-Shifted Fluorescent Markers. *The Journal of Physical Chemistry B* **121**, 10018-10025
45. Wehler, P., Armbruster, D., Günter, A., Schleicher, E., Di Ventura, B., and Öztürk, M. A. (2021) Experimental Characterization of In Silico Red-Shift-Predicted iLOVL470T/Q489K and iLOVV392K/F410V/A426S Mutants. *ACS Omega*
46. Goncharov, I. M., Smolentseva, A., Semenov, O., Natarov, I., Nazarenko, V. V., Yudenko, A., Remeeva, A., and Gushchin, I. (2021) High-resolution structure of a naturally red-shifted LOV domain. *Biochem Biophys Res Commun* **567**, 143-147

47. Orozco-Gonzalez, Y., Kabir, M. P., and Gozem, S. (2019) Electrostatic spectral tuning maps for biological chromophores. *The Journal of Physical Chemistry B* **123**, 4813-4824
48. Kabir, M. P., Orozco-Gonzalez, Y., and Gozem, S. (2019) Electronic spectra of flavin in different redox and protonation states: a computational perspective on the effect of the electrostatic environment. *Physical Chemistry Chemical Physics* **21**, 16526-16537
49. Kabir, M. P., Orozco-Gonzalez, Y., Hastings, G., and Gozem, S. (2021) The effect of hydrogen-bonding on flavin's infrared absorption spectrum. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **262**, 120110
50. Stephens, P. J., Devlin, F. J., Chabalowski, C. F., and Frisch, M. J. (1994) Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of physical chemistry* **98**, 11623-11627
51. Kendall, R. A., Dunning Jr, T. H., and Harrison, R. J. (1992) Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of chemical physics* **96**, 6796-6806
52. Matsuoka, Y., and Norden, B. (1983) Linear dichroism studies of flavins in stretched poly (vinyl alcohol) films. Molecular orientation and electronic transition moment directions. *The Journal of Physical Chemistry* **87**, 220-225
53. Siddiqui, M. S. U., Kodali, G., and Stanley, R. J. (2008) Electronic transition dipole moment directions of reduced anionic flavin in stretched poly (vinyl alcohol) films. *The Journal of Physical Chemistry B* **112**, 119-126
54. Gagliardi, L., Lindh, R., and Karlstrom, G. (2004) Local properties of quantum chemical systems: the LoProp approach. *J Chem Phys* **121**, 4494-4500
55. Schrodinger, LLC. (2015) The PyMOL Molecular Graphics System, Version 1.8.
56. Dratch, B. D., Orozco-Gonzalez, Y., Gadda, G., and Gozem, S. (2021) Ionic Atmosphere Effect on the Absorption Spectrum of a Flavoprotein: A Reminder to Consider Solution Ions. *The Journal of Physical Chemistry Letters* **12**, 8384-8396
57. Whitby, L. G. (1953) A new method for preparing flavin-adenine dinucleotide. *Biochem J* **54**, 437-442
58. Orozco-Gonzalez, Y., Manathunga, M., Marín, M. D. C., Agathangelou, D., Jung, K.-H., Melaccio, F., Ferré, N., Haacke, S., Coutinho, K., and Canuto, S. (2017) An average solvent electrostatic configuration protocol for Qm/mm free energy optimization: Implementation and application to rhodopsin systems. *Journal of chemical theory and computation* **13**, 6391-6404
59. Okuyama-Yoshida, N., Kataoka, K., Nagaoka, M., and Yamabe, T. (2000) Structure optimization via free energy gradient method: Application to glycine zwitterion in aqueous solution. *The Journal of Chemical Physics* **113**, 3519-3524
60. Zwanzig, R. W. (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. *The Journal of Chemical Physics* **22**, 1420-1426
61. Orozco-Gonzalez, Y., Manathunga, M., Marin, M. d. C., Agathangelou, D., Jung, K.-H., Melaccio, F., Ferré, N., Haacke, S., Coutinho, K., and Canuto, S. (2017) An average solvent electrostatic configuration protocol for Qm/mm free energy optimization: Implementation and application to rhodopsin systems. *Journal of chemical theory and computation* **13**, 6391-6404
62. Iyer, A., Reis, R. A. G., Gannavaram, S., Momin, M., Spring-Connell, A. M., Orozco-Gonzalez, Y., Agniswamy, J., Hamelberg, D., Weber, I. T., Gozem, S., Wang, S., Germann, M. W., and Gadda, G. (2021) A Single-Point Mutation in d-Arginine

Dehydrogenase Unlocks a Transient Conformational State Resulting in Altered Cofactor Reactivity. *Biochemistry* **60**, 711-724

63. Fdez Galvan, I., Vacher, M., Alavi, A., Angeli, C., Aquilante, F., Autschbach, J., Bao, J. J., Bokarev, S. I., Bogdanov, N. A., Carlson, R. K., Chibotaru, L. F., Creutzberg, J., Dattani, N., Delcey, M. G., Dong, S. S., Dreuw, A., Freitag, L., Frutos, L. M., Gagliardi, L., Gendron, F., Giussani, A., Gonzalez, L., Grell, G., Guo, M., Hoyer, C. E., Johansson, M., Keller, S., Knecht, S., Kovacevic, G., Kallman, E., Li Manni, G., Lundberg, M., Ma, Y., Mai, S., Malhado, J. P., Malmqvist, P. A., Marquetand, P., Mewes, S. A., Norell, J., Olivucci, M., Oppel, M., Phung, Q. M., Pierloot, K., Plasser, F., Reiher, M., Sand, A. M., Schapiro, I., Sharma, P., Stein, C. J., Sorensen, L. K., Truhlar, D. G., Ugandi, M., Ungur, L., Valentini, A., Vancoillie, S., Veryazov, V., Weser, O., Wesolowski, T. A., Widmark, P. O., Wouters, S., Zech, A., Zobel, J. P., and Lindh, R. (2019) OpenMolcas: From Source Code to Insight. *J Chem Theory Comput* **15**, 5925-5964
64. Rackers, J. A., Wang, Z., Lu, C., Laury, M. L., Lagardere, L., Schnieders, M. J., Piquemal, J. P., Ren, P., and Ponder, J. W. (2018) Tinker 8: Software Tools for Molecular Design. *J Chem Theory Comput* **14**, 5273-5289
65. Andruniów, T., Ferré, N., and Olivucci, M. (2004) Structure, initial excited-state relaxation, and energy storage of rhodopsin resolved at the multiconfigurational perturbation theory level. *Proc Natl Acad Sci U S A* **101**, 17908-17913
66. Ferré, N., and Olivucci, M. (2003) Probing the rhodopsin cavity with reduced retinal models at the CASPT2//CASSCF/AMBER level of theory. *Journal of the American Chemical Society* **125**, 6868-6869
67. Olsson, M. H., Sondergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* **7**, 525-537
68. Zhang, L., and Hermans, J. (1996) Hydrophilicity of cavities in proteins. *Proteins: Structure, Function, and Bioinformatics* **24**, 433-438
69. Krivov, G. G., Shapovalov, M. V., and Dunbrack Jr, R. L. (2009) Improved prediction of protein side-chain conformations with SCWRL4. *Proteins: Structure, Function, and Bioinformatics* **77**, 778-795
70. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19-25
71. Ferré, N., and Ángyán, J. G. (2002) Approximate electrostatic interaction operator for QM/MM calculations. *Chemical physics letters* **356**, 331-339
72. Melaccio, F., Del Carmen Marín, M., Valentini, A., Montisci, F., Rinaldi, S., Cherubini, M., Yang, X., Kato, Y., Stenrup, M., and Orozco-Gonzalez, Y. (2016) Toward automatic rhodopsin modeling as a tool for high-throughput computational photobiology. *Journal of Chemical Theory and Computation* **12**, 6020-6034
73. Pedraza-Gonzalez, L., De Vico, L., Mari, N. M., Fanelli, F., and Olivucci, M. (2019) a-ARM: Automatic Rhodopsin Modeling with Chromophore Cavity Generation, Ionization State Selection, and External Counterion Placement. *J Chem Theory Comput* **15**, 3134-3152
74. Pedraza-Gonzalez, L., Barneschi, L., Padula, D., De Vico, L., and Olivucci, M. (2022) Evolution of the Automatic Rhodopsin Modeling (ARM) Protocol. *Top Curr Chem (Cham)* **380**, 21

75. Schneider, C., and Sühnel, J. (1999) A molecular dynamics simulation of the flavin mononucleotide–RNA aptamer complex. *Biopolymers: Original Research on Biomolecules* **50**, 287-302
76. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **117**, 5179-5197
77. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **65**, 712-725
78. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926-935
79. Aquilante, F., Malmqvist, P.-Å., Pedersen, T. B., Ghosh, A., and Roos, B. O. (2008) Cholesky decomposition-based multiconfiguration second-order perturbation theory (CD-CASPT2): application to the spin-state energetics of CoIII (diiminato)(NPh). *Journal of chemical theory and computation* **4**, 694-702
80. Forsberg, N., and Malmqvist, P.-Å. (1997) Multiconfiguration perturbation theory with imaginary level shift. *Chemical Physics Letters* **274**, 196-204
81. Ghigo, G., Roos, B. O., and Malmqvist, P.-Å. (2004) A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2). *Chemical physics letters* **396**, 142-149
82. Gozem, S., Mirzakulova, E., Schapiro, I., Melaccio, F., Glusac, K. D., and Olivucci, M. (2014) A conical intersection controls the deactivation of the bacterial luciferase fluorophore. *Angew Chem Int Ed Engl* **53**, 9870-9875
83. Giuliani, G., Melaccio, F., Gozem, S., Cappelli, A., and Olivucci, M. (2021) QM/MM Investigation of the Spectroscopic Properties of the Fluorophore of Bacterial Luciferase. *J Chem Theory Comput* **17**, 605-613
84. Zobel, J. P., Nogueira, J. J., and Gonzalez, L. (2017) The IPEA dilemma in CASPT2. *Chem Sci* **8**, 1482-1499
85. Wu, M., and Eriksson, L. A. (2010) Absorption Spectra of Riboflavin-A Difficult Case for Computational Chemistry. *The Journal of Physical Chemistry A* **114**, 10234-10242
86. Klaumunzer, B., Kroner, D., and Saalfrank, P. (2010) (TD-)DFT calculation of vibrational and vibronic spectra of riboflavin in solution. *J Phys Chem B* **114**, 10826-10834
87. Gozem, S., and Krylov, A. I. (2022) The ezSpectra suite: An easy-to-use toolkit for spectroscopy modeling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **12**, e1546