QM/MM Investigation of the Spectroscopic Properties of the Fluorophore of Bacterial Luciferase

Germano Giuliani,

Department of Biotechnology, Chemistry and Pharmacy, University of Siena, 53100 Siena, Italy

Federico Melaccio,

Department of Biotechnology, Chemistry and Pharmacy, University of Siena, 53100 Siena, Italy

Samer Gozem,

Department of Chemistry, Georgia State University, Atlanta, Georgia 30302, United States

Andrea Cappelli,

Department of Biotechnology, Chemistry and Pharmacy, University of Siena, 53100 Siena, Italy

Massimo Olivucci

Department of Biotechnology, Chemistry and Pharmacy, University of Siena, 53100 Siena, Italy; Department of Chemistry, Bowling Green State University, Bowing Green, Ohio 43403, United States

Abstract

We employ replica-exchange molecular dynamics (REMD) and a hybrid *ab initio* multiconfigurational quantum mechanics/molecular mechanics (QM/MM) approach to model the absorption and fluorescence properties of bacterial luciferin-luciferase. Specifically, we employ complete active space perturbation theory (CASPT2) and study the effect of active space, basis set, and IPEA shift on the computed energies. We discuss the effect of the protein environment on the fluorophore's excited-state potential energy surface and the role that the protein plays in enhancing the fluorescence quantum yield in bacterial bioluminescence.

Graphical Abstract

Corresponding Authors Samer Gozem – sgozem@gsu.edu.

FMNHOH distortion hindered

+ 85°   + 175°

Bioluminescent enzymes such as bacterial luciferases are used as gene reporters in cell culture and small-animal imaging.[1] They have also become important tools for life science research and drug discovery.[2] Bacterial luciferases are found naturally in bioluminescent bacteria.[3] Their fluorescence wavelength is typically in the range 460–500 nm, with a peak near 490 nm[4] and a fluorescent quantum yield ranging from 0.1 to 0.3.[5] There is strong experimental[6] and computational[7] evidence implicating the excited state of intermediate **2** (see Scheme 1) in the fluorescence.

Bacterial luciferases generate bioluminescence by catalyzing a reaction involving the reduced flavin mononucleotide (FMNH$_2$; **1** in Scheme 1), O$_2$, and an aliphatic aldehyde. The reaction produces the corresponding aliphatic carboxylic acid, flavin mononucleotide (FMN, **3**), and blue-green light (~490 nm). As reported in Scheme 1B, the reaction proceeds via three intermediates formed by flavin derivatives inside the luciferase binding pocket. First, the bacterial luciferase apoprotein binds with FMNH$_2$ (**1**), which rapidly reacts with atmospheric oxygen to give the corresponding oxidized intermediate (FMNHOOH). Another reductive enzyme then provides a long-chain aldehyde (by reducing a carboxylic acid)[8] which reacts with FMNHOOH to give a peroxyhemiacetal intermediate. This intermediate decomposes to generate the 4a-hydroxy flavin intermediate FMNHOH (**2** in Scheme 1A) in an electronically excited state. Finally, **2** undergoes radiative decay to the ground state, yielding FMN (**3**) as a final product.

Although intermediate **2** displays weak fluorescence in solution,[10] it exhibits strong fluorescence when bound to the enzymatic binding pocket.[11] We previously investigated the fluorescence properties of the *N*(5)-alkylated derivative of **2** in a joint computational and experimental study.[12] The results (see Figure S1 in the Supporting Information) indicated that a large distortion of the terminal pyrimidine ring of the molecule leads to an excited-state deactivation via a conical intersection (CI)-mediated internal conversion. When the geometric deformation of the molecule was inhibited by some constraint (e.g., in a frozen solvent that restrains the pyrimidine ring distortion), an enhanced fluorescence emission was observed. It is assumed that the surrounding environment within the enzymatic binding pocket can behave in the same manner as a frozen solvent. In fact, the crystal structure of the *Vibrio harveyi* luciferase FMN complex (PDB ID: 3FGC)[13] reveals that the fluorophore is hosted in a tight protein cavity.

To investigate the fluorescent properties of this system, we employ a hybrid quantum mechanical/molecular mechanical (QM/MM) model constructed from the crystal structure.

METHODOLOGY

Preparing the Protein Model.

The crystal structure of the luciferase/FMN complex (PDB ID 3FGC) was used as a starting point for the simulations.[13] Specifically, we started with the coordinates of heterodimer **2**, since the mobile loop from heterodimer **2** had stronger electron density and therefore higher resolution.

Bacterial luciferase comprises two homologous subunits, designated α and β, both of which assume the TIM barrel fold.[14] Although the β-subunit is required for activity, the catalytic site resides exclusively in the α-subunit.[15] The crystallographic structure of the complex contains these two β/α-heterodimers,[13] with the α-subunit binding the FMN. The primary difference between the two subunits corresponds to a secondary structural element comprising two antiparallel β-strands near the interface with the β-subunit.[16] For computational efficiency, only the α-chain was included in the simulations. The residues between sequence positions 283 and 291 not observed in the crystal structure were added using homology-based loop modeling (ModLoop).[17] This disordered segment (residues 283–291) is directly adjacent to the FMN-binding cavity, so to build an accurate QM/MM model would require this segment to be modeled carefully. Indeed, as observed by Campbell et al.,[18] the presence of the fluorophore has a significant influence on the conformation of this loop. Therefore, we chose to model these large loop movements in the α-subunit in the presence of the fluorophore (**2**) within the catalytic binding site.

To generate a model of the fluorophore-bound luciferase, we first perform a gas-phase energy minimization of **2** in the *trans* configuration.[7a] Next, FMN from the crystal structure was replaced by this optimized structure of **2**. The entire system was then solvated with explicit water molecules and energy minimized. The movements of residues within the subunit interface were weakly constrained using a harmonic potential to prevent complete unfolding of the α-subunit over the course of simulation. No constraints were applied to the FMNHOH or the mobile loop.

Replica-Exchange Molecular Dynamics (REMD).

REMD differs from standard molecular dynamics (MD) due to the parallel nature of the simulation.[19] Instead of a single simulation at a single temperature, multiple parallel simulations are conducted over a range of temperatures. Periodically, conformations of simulation replicas running at different temperatures are exchanged according to a Monte Carlo procedure. As a result, the conformational space sampled over the simulation is substantially increased. We employed REMD simulations to obtain better sampling of loop conformations. MD and REMD simulations were executed using the GROMACS 4.6.3 program[20] using the AMBER99SB force field.[21] The loop and all atoms within 6 Å of any atom of the loop were allowed to move during REMD, while the rest was kept fixed. The trajectories obtained and the operative conditions are reported in the Supporting Information.

The cluster analysis was accomplished using the single linkage algorithm as implemented in the GROMACS 4.6.3[20] *g_cluster* routine. The five most populated clusters (Figure 1) were obtained by statistical analysis of the REMD trajectory. The setup of the procedure employed 0.1 nm as the RMSD cutoff option to obtain similar structures within the ensembles, and subsequently, the structure with the smallest average distance from the others in each cluster (populated by a minimum of 500 different states) was selected.

In Figure 1A, the entire α-subunit/FMNHOH complex is shown. The final structure of the mobile loop (in red), which was subsequently used for QM/MM calculations, was chosen as the average of the most populated cluster. In Figure 1B, the structures of the other four major configurations obtained by the clustering analysis are displayed in different colors. Each one of those configurations was extrapolated from a cluster of structures collected in the presence of the ligand. A comparison of the RMSD values for each representation suggests that the closed conformation of the mobile loop is predominant within the ensembles.

QM/MM Calculations.

Simulating the spectroscopy of luciferin-luciferase requires the use of a quantum chemical method capable of describing both the ground and excited states of FMNHOH on an equal footing. This means having a balanced description of static and dynamical electron correlation for both states. We employ the complete active space self-consistent field (CASSCF)[22] for geometry optimizations and complete active space second order perturbation theory (CASPT2)[23] for correcting the electronic energies of the CASSCF-optimized structures.

This CASPT2//CASSCF protocol is well-established for photochemistry and photobiology.[24] However, while it has been widely used for rhodopsins and other photoreceptor proteins, far fewer studies have applied it to model bacterial luciferase fluorescence. It is therefore not immediately clear which active space and basis set to use. Similarly, an ionization potential electron affinity (IPEA) shift of 0.25 is often applied to improve CASPT2 excitation energy calculations,[25] but the use of IPEA and the magnitude of the IPEA shift that should be used has been discussed multiple times in the literature.[26] We therefore compute the absorption wavelength and fluorescence wavelength of FMNHOH using several different active spaces, basis sets, and testing both IPEA = 0 and IPEA = 0.25.

Note that the full π active space of 2 would comprise 18 electrons in 15 orbitals. However, we might expect that a reasonable reduction in active space would be the removal of two π and π* orbitals from the active space, resulting in a 14-electron, 11-orbital active space. This is the active space used in our reference solvent-phase study.[12]

While relative energies (e.g., across a series of similar structures or protein mutants) are typically not very sensitive to the details of the CASPT2//CASSCF method used, absolute energy differences(such as computed absorption and fluorescence wavelengths) are sensitive to details such as the active space, basis set, and IPEA shift.[24] We will test both the 18,15 and 14,11 active spaces. We will also test the effect of using single-state vs multistate CASPT2, the effect of the IPEA shift set to 0 vs 0.25, and the effect of increasing the basis set from 6–31G* to ANO-L-VTZP.

Finally, we note that there is a cysteine (Cys106) residue that is near the pyrimidine ring of the fluorophore (within ca. 2.7 Å). This cysteine is highly conserved in many proteins associated with the bacterial luciferase-like superfamily.[27] Site-directed mutagenesis studies of the luciferase have shown that mutation of the polar thiol group of cysteine to the nonpolar isopropyl side chain of valine destabilizes the FMNHOH emitting intermediate,[28] indicating that the electrostatic effect of cysteine may also be important for modulating the emission of the fluorophore. Therefore, we have generated two QM/MM models: one where the −CH_2SH atoms of Cys106 are treated at the QM level of theory (Figure 2) and one where the Cys106 is entirely treated at the MM level of theory.

Optimizations and minimum energy path (MEP) calculations for FMNHOH were performed at the three-root state-averaged CASSCF level of theory[22] with a 14,11 active space and 6–31G* basis set (see the Supporting Information for more details on the active space). The MEPs both with and without the cysteine in the QM subsystem were computed with the intrinsic reaction coordinate method employing a step constraint of 0.03 b·$amu^{1/2}$ for the first eight steps, followed by 0.05 b·$amu^{1/2}$ for the remainder of the MEP. Single-point CASPT2 calculations[23] were then performed for stationary points and for each point along the MEP path. The Mulliken charges and oscillator strengths reported are computed at the CASPT2 level of theory. All QM and QM/MM calculations were performed with Molcas 7.8 interfaced with Tinker.[29] MM atoms were treated using the AMBER99SB force field.[21] More details regarding the computational protocol and the QM/MM model are provided in the Supporting Information.

RESULTS AND DISCUSSION

Luciferase-bound FMNHOH (**2**) was first optimized in the ground state (GS). The structure and relevant bond lengths and dihedrals are shown in Figure 3, top left. We also optimized the structure on the first singlet excited state, to obtain the excited-state minimum (EM) from which fluorescence occurs.

At the GS and EM structures, respectively, we computed the vertical excitation wavelength (λ_{abs}) and fluorescence wavelength (λ_{em}) using different CASPT2 approaches. The data is shown in Table 1.

A number of trends are clear in the computed wavelengths. Specifically, we find the following:

- The IPEA shift of 0.25 invariably blue-shifts the absorption and fluorescence wavelengths (by up to 100 nm). The IPEA shift has originally been introduced to remedy a systematic underestimation of excitation energies computed using the CASPT2 method,[25] and here, we find that the IPEA shift has a similar effect as its originally intended use. For instance, in rhodopsin models, it has been shown that using no IPEA shift benefits from a cancelation of errors in combination with smaller basis sets and that IPEA = 0.25 works well in combination with large basis sets.[26a] More recently, Gonzalez and co-workers indicated that the IPEA shift simply is a way to recover missing dynamical electron correlation.[26b] Therefore, in this case, the IPEA shift may be considered a computationally

inexpensive way to approximately account for missing dynamical electron correlation.

- Increasing the basis set invariably red-shifts the absorption and fluorescence wavelengths. The change from 6–31G* to ANO-L-VDZP is usually more substantial (8–39 nm, depending on method) than the change from ANO-L-VDZP to ANO-L-VTZP (2–13 nm). Again, this is consistent with results found for rhodopsin models[26a] and is likely due to the excited state being more diffuse than the ground state.

- The red-shift from the basis set increase is not enough to offset the effect of the blue-shift from the IPEA shift. Therefore, unlike for other biological chromophores where this method provides a favorable cancelation of errors,[26a] the IPEA = 0 and 6–31G* basis set combination does not provide quantitative results in this system.

- It is difficult to predict the effect of moving from single-state (SS) CASPT2 to multistate (MS) CASPT2. This sometimes has no effect at all and sometimes may have a large effect (as large as 67 nm). Interestingly, with the 14,11 active space, the absorption changes only slightly while the fluorescence wavelength changes a lot more when using SS vs MS CASPT2. However, with the 18,15 active space, the opposite is observed; the SS and MS CASPT2 are in good agreement for the fluorescence, while they are very different for the absorption. This, likely, stems from state crossing occurring between higher states that may change the character of the third excited state used in state averaging. Indeed, SS and MS CASPT2 do suffer from artifacts near state crossings (see refs 26a and 31). Methods to address this have been recently developed but have not been tested in this work.

Due to the computational cost of running MS-CASPT2-(18,15) with the ANO-L-VTZP, we obtained an estimate by extrapolating other data. Specifically, the basis set effect of moving from ANO-L-VDZP to ANO-L-VTZP for MS-CASPT2(14,11)/ IPEA = 0.25 was used to correct the MS-CASPT2(18,15)/IPEA = 0.25 ANO-L-VDZP data. These resulting computed vertical absorption and emission wavelengths are not in quantitative agreement with experiment, but they are within 0.3 eV. Specifically, the vertical emission wavelength computed at the EM (2.24 eV, or 554 nm) is red-shifted with respect to the experimentally observed wavelength in the luciferase of *Vibrio harveyi* (2.53 eV, 490 nm),[9] while, the computed absorption wavelength (3.62 eV, 342 nm) is blue-shifted relative to the experimentally determined absorption spectrum in *Vibrio harveyi* (360 nm).[4] Our results are similar to calculations by Luo et al. for this system.[7b] We also note that calculations of the absorption and emission maxima of reduced flavin have also been red-shifted to a similar extent.[32] Indeed, in both oxidized and reduced flavins, it has also been shown that vertical excitation and emission energies are often shifted with respect to experiments, with ref 32 in particular pointing to the need for computing adiabatic excitation energies and Franck–Condon factors to get a better agreement between computations and experimental λ_{max} for these systems. Our computed adiabatic excitation energy at the MS-CASPT2(18,15)/IPEA = 0.25/ANO-L-VDZP level of theory is 2.82 eV (440 nm). Notice that the experimental

absorption and emission wavelengths lie somewhere between the computed vertical and adiabatic excitation energies. We also cannot rule out that other details of the QM/MM model, such as the polarizability of the MM subsystem, may be important to take into account.[33]

The oscillator strength at the GS structure is 0.8. At the EM, it decreases to 0.6. Both strengths are indicative of strong absorption and fluorescence intensities, consistent with a strong radiative $S_1 \rightarrow S_0$ decay (i.e., fluorescence). The radiative lifetime can be roughly estimated from the oscillator strength (in the absence of competing photophysical processes) using a simplified version of the Strickler–Berg equation[34]

$$1/\tau^0_{\mathrm{f}} \approx 0.7\tilde{v}_0{}^2 f$$

where τ^0_{f} is the fluorescence lifetime, v_0 is the adiabatic excitation in units of cm^{-1}, and f is the oscillator strength. If we assume that the adiabatic absorption corresponds to a wavelength of ~500 nm, this yields a fluorescence lifetime of ~6 ns, in good agreement with the measured fluorescence lifetime of ~10 ns.[35] This indicates that competing processes that may quench fluorescence, such as internal conversion or intersystem crossing, are minimized in the protein, which is different from the behavior of the same system in solution where internal conversion is efficient.[12]

Next, we connect the vertical excited GS structure, corresponding to the Franck–Condon (FC) point, to the EM, as done in ref 12. This is done using a minimum energy path (MEP) starting from the FC structure. The S_0 and S_1 energy profiles along the MEP are shown in Figure 3. The purpose of the MEP is to explore the excited-state potential energy surface of FMNHOH near the FC and EM region and to compare with the potential energy surface in Figure S1 from the Supporting Information.[12]

Mutagenesis studies on bacterial luciferase have shown that mutation of Cys106 destabilizes the FMNHOH emitting intermediate.[28] Therefore, to investigate a possible role of the Cys106 side chain, we compute this MEP both in the presence and absence of the cysteine in the QM region. Structures along the MEP were obtained by CASSCF geometry optimization, while their single-point energies were computed at the CASPT2 level of theory. In both MEP calculations (with and without the QM Cys106), we use SS-CASPT2(14,11)/IPEA = 0/6–31G*, since this method benefits from some cancelation of error, is affordable, and still gives us an idea about the potential energy surface and the modes involved in reaching the EM.

Both the calculations with and without the Cys106 in the QM region indicate that there is no CI in the vicinity of the FC and EM regions of the S1 potential energy surface. Indeed, at the EM, the S_0–S_1 energy gap is over 2 eV. This is in contrast to calculations on the same fluorophore in solution, where the MEP calculations lead to a CI instead of an EM, indicating relatively efficient internal conversion in solution (ref 12 and Figure S1). The following changes upon including Cys106 in the QM region (see Figure 3 left):

- The first singlet excited state decreases in energy relative to the ground state.

- The excited-state potential becomes slightly flatter. Indeed, while the MEP with Cys106 in the QM region shows a decrease in energy of 0.81 eV between the FC and EM points on the excited state, the calculation with Cys106 in the MM region shows a larger decrease, of 0.87 eV. It also takes additional MEP points for the Cys106 in QM calculation to reach the EM.

The second point indicates that the Cys106 may play a role in modifying the shape of the S_1 potential energy surface. We speculate that this affects vibronic coupling between the S_1 and S_0 states and may explain why mutation of Cys106 affects the fluorescence of bacterial luciferase.[28]

We also note that including Cys106 in the QM region further worsens the agreement of the absorption and fluorescence wavelengths with experiment. However, including only one polarizing amino acid in the QM region without including other polar amino acids in different positions is biasing the calculation, which may explain the bad agreement (almost 100 nm red-shifted) with respect to the experiment.

The analysis of the MEP (Figure 3) shows that **2** follows multiple molecular relaxation modes to reach the EM. First, a bond length alternation (BLA) inversion occurs within the isoalloxazine ring, where alternating double and single bonds become elongated and shortened, respectively. Specifically, the N10–C10a bond evolves from having double bond character (1.32 Å) on S_0 to a single bond (1.42 Å) on S_1. On the other hand, the adjacent N10–C9a bond changes from a single bond on S_0 (1.42 Å) to gain double-bond character (1.38 Å). This is a fast relaxation typically associated with the π,π^* excitation of π-conjugated systems.[36] A similar excited-state relaxation was reported for lumiflavin using computations of comparable quality.[37] At the same time, since the out-of-plane deformation of the pyrimidine ring with respect to the original flavin plane does not occur in the protein, a different geometrical distortion becomes more prominent here instead. More specifically, following the S_1 energy profile reported in Figure 3, we note an increase in the pyramidalization of carbon C4a indicating an increased tendency to adopt a tetrahedral geometry. Note that C4a is originally sp2-hybridized in FMN but is sp3-hybridized in FMNHOH, where the C4a–OH bond in FMNHOH is nearly perpendicular to the isoalloxazine plane. According to these observations, the calculations revealed a change of the N1–C10a–C4a–C4 dihedral angle from −21° at the FC structure to −43° at the energy minimum on S_1.

Other changes in bond lengths are directly consistent with the BLA previously described;[12] the C4a–N5 bond becomes more elongated (1.41 to 1.44 Å), while the adjacent N5–C5a reduces its length (1.40 to 1.33 Å). These bond stretching modes are coupled with the pyramidalization at C4a. As a result of these structural changes, part of the electronic charge on N5 is transferred to the contiguous benzene ring, resulting in a shorter N5–C5a bond. The N5 lone pair electrons do not conjugate with the aromatic ring because they still lie in an orthogonal plane (Figure S7 of the Supporting Information).

Such a coupling between the charge transfer and excited-state distortion is similar to that observed in other π-conjugated biological chromophores, such as the retinal protonated Schiff base.[36] One explanation for the increased pyramidalization of C4a is the propensity

of the fluorophore to distort out of plane, as observed in our previous study for the isolated fluorophore.[12] In the protein cavity, since the large out-of-plane distortion is prevented by the tight protein binding cavity, the molecule instead relaxes through a localized pyramidalization.

An analysis of the Mulliken charges along the MEP is reported in the Supporting Information. As anticipated earlier, a charge transfer process accompanies the S_1 geometrical relaxation of **2**. The $S_0 \rightarrow S_1$ vertical excitation gives a small charge transfer of ca. -0.1 e$^-$ from the benzene and pyrazine rings to the pyrimidine ring, but as the molecule relaxes on the excited state, this charge transfer magnitude increases to -0.3 e$^-$.

To better understand the photophysics and spectroscopy of this bacterial luciferase, it is important to analyze both the steric and electrostatic effects of the protein on the FMNHOH fluorophore. A Poisson–Boltzmann analysis[38] that was performed during the PROPKA (pH 7) determination of side-chain ionization states indicates that the cavity region near the pyrimidine moiety has a negative electrostatic potential. These electrostatic effects have differential effects on the different states and geometries of FMNHOH. For example, since excitation is accompanied by a transfer of negative charge to the pyrimidine ring (fast BLA relaxation), the electrostatics of the proximal residues would play a relevant role to stabilize the emitting state of the intermediate, therefore increasing both the S_1 lifetime and fluorescence quantum yield.

Since the fluorescence of FMNHOH is bright and long-lived, that means that the CI observed in the gas-phase study[12] (reached by distortion of the C9a–N10–C10a–N1 dihedral in Figure S1) must be inaccessible in the protein. In an attempt to map the energy profile connecting the EM to the CI in the protein, we performed a relaxed scan to mimic the same geometrical distortion within the protein binding pocket by constraining the C9a–N10–C10a–N1 up to a 90° out-of-plane distortion. The resulting potential energy scan (shown on the right side of Figure 3) has a very different shape compared to the gas-phase calculation.[12] Specifically, as expected, the geometrical distortions result in an increase rather than a decrease in the potential energy in the presence of the protein. No CI is observed in the protein, even when the pyrimidine ring is constrained to be 90° with respect to the rest of the fluorophore. This is in contrast to the solution phase where the same deformation results in easy access to a low-lying CI. The same CI in the protein is sloped and must be at least 1 eV above the EM (Figure 3, right). This is largely due to the steric effect of the binding pocket, as hypothesized in ref 12, which constrains the fluorophore in a planar configuration and maximizes the fluorescent quantum yield of the EM.

In conclusion, this work computationally investigates the spectroscopy and photophysics of the FMNHOH fluorophore of bacterial luciferase. It is found that the protein blocks an internal conversion channel that was accessible in solution by significantly modifying the shape of the excited-state potential energy surface of the fluorophore. By comparing different QM methods, we test the ability of CASPT2//CASSCF to model the spectroscopic properties of the protein-bound flavin derivative and provide additional insight into the effect of the protein on the fluorophore's energetics. These effects stabilize the emitting state of the fluorophore to maximize its fluorescence. We hypothesize that steric effects play the

main role in enhancing fluorescence by keeping the chromophore in its planar fluorescent state. We have also studied the effect of placing one of the nearby cysteine side chains, Cys106, in the QM region. This test indicates that it is unlikely that the Cys106 plays a role in controlling access to the CI, but it may play a role in modulating the curvature of the excited-state potential energy surface. Further studies are needed to understand how the electrostatic environment around the fluorophore modulates its fluorescence properties.

The investigations of the effect of protein environment on the fluorescent properties of FMNHOH may lead to the further development of FMNHOH as a tool to probe dynamics or rigidity of protein active sites during processes such as catalysis, similar to studies with FMN.[39] Furthermore, the insight revealed in this work may aid in the synthesis of novel flavin adducts as novel fluorescent probes.[40]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES

(1). (a) Xu T; Close D; Handagama W; Marr E; Sayler G; Ripp S The Expanding Toolbox of In Vivo Bioluminescent Imaging. Front. Oncol. 2016, 6, 150. [PubMed: 27446798] (b) Contag CH; Bachmann MH Advances in In Vivo Bioluminescence Imaging of Gene Expression. Annu. Rev. Biomed. Eng. 2002, 4, 235–260. [PubMed: 12117758]

(2). (a) Shimomura O Bioluminescence: chemical principles and methods; World Scientific Publishing Co. Ltd: Singapore, 2005; Chapter 2.(b) Roda A; Guardigli M; Michelini E; Mirasoli M Bioluminescence in Analytical Chemistry and In Vivo Imaging. TrAC, Trends Anal. Chem. 2009, 28, 307–322.

(3). Campbell AK Chemiluminescence, Principles and Applications in Biology and Medicine; Ellis Horwood Ltd.: Chichester, England, 1988; pp 189–218.

(4). Kurfurst M; Ghisla S; Hastings JW Characterization and Postulated Structure of the Primary Emitter in the Bacterial Luciferase Reaction. Proc. Natl. Acad. Sci. U. S. A. 1984, 81, 2990–2994. [PubMed: 16593462]

(5). Lee J; Matheson IBC; O'Kane DJ; Vervoort J; Visser AJWG Chemistry and Biochemistry of Flavin and Flavoenzymes; CRC: Orlando, FL, 1999; Vol. 2.

(6). (a) Hastings JW; Potrikusv CJ; Gupta SC; Kurfurst M; Makemson JC Biochemistry and Physiology of Bioluminescent Bacteria. Adv. Microb. Physiol. 1985, 26, 235–291. [PubMed: 3913293] (b) Vervoort J; Miller F; Okane DJ; Lee J; Bacher A Bacterial Luciferase: a Carbon-13, Nitrogen-15, and Phosphorus-31 Nuclear Magnetic Resonance Investigation. Biochemistry 1986, 25, 8067–8075.(c) Macheroux P; Ghisla S; Hastings JW Spectral Detection of an Intermediate Preceding the Excited State in the Bacterial Luciferase Reaction. Biochemistry 1993, 32, 14183–14186. [PubMed: 8260504]

(7). (a) Hou C; Liu Y-J; Ferre N; Fang W-H Understanding Bacterial Bioluminescence: A Theoretical Study of the Entire Process, from Reduced Flavin to Light Emission. Chem. - Eur. J. 2014, 20, 7979–7986. [PubMed: 24825310] (b) Luo Y; Liu YJ Bioluminophore and Flavin Mononucleotide

Fluorescence Quenching of Bacterial Bioluminescence—A Theoretical Study. Chem. - Eur. J. 2016, 22, 16243–16249. [PubMed: 27665749]

(8). (a) Ulitzur S; Hastings JW Evidence for Tetradecanal as the Natural Aldehyde in Bacterial Bioluminescence. Proc. Natl. Acad. Sci. U. S. A. 1979, 76, 265–267. [PubMed: 311472] (b) Meighen EA Bacterial Bioluminescence: Organization, Regulation, and Application of the Lux Genes. FASEB J. 1993, 7, 1016–1022. [PubMed: 8370470]

(9). Lei B; Ding Q; Tu SC Identity of the Emitter in the Bacterial Luciferase Luminescence reaction: Binding and Fluorescence Quantum Yield studies of 5-Decyl-4a-hydroxy-4a,5-dihydroriboflavin-5′-phosphate as a model. Biochemistry 2004, 43, 15975–15982. [PubMed: 15595854]

(10). Moonen CT; Vervoort J; Müller, F. Carbon-13 Nuclear Magnetic Resonance study on the Dynamics of the Conformation of Reduced Flavin. Biochemistry 1984, 23, 4868–4872. [PubMed: 6541948]

(11). (a) Kao YT; Saxena C; He TF; Guo L; Wang L; Sancar A; Zhong D Ultrafast Dynamics of Flavins in Five Redox States. J. Am. Chem. Soc. 2008, 130, 13132–13139. [PubMed: 18767842] (b) Zhou D; Mirzakulova E; Khatmullin R; Schapiro I; Olivucci M; Glusac KD Fast Excited-state Deactivation in N(5)-Ethyl-4a-hydroxyflavin Pseudobase. J. Phys. Chem. B 2011, 115, 7136–7143. [PubMed: 21553832] (c) Li G; Sichula V; Glusac KD Role of Adenine in Thymine-Dimer Repair by Reduced Flavin-Adenine Dinucleotide. J. Phys. Chem. B 2008, 112, 10758–10764. [PubMed: 18681479] (d) Li J; Liu Z; Tan C; Guo X; Wang L; Sancar A; Zhong D Dynamics and Mechanism of Repair of Ultraviolet-Induced (6–4) Photoproduct by Photolyase. Nature 2010, 466, 887–890. [PubMed: 20657578]

(12). Gozem S; Mirzakulova E; Shapiro I; Melaccio F; Glusac KD; Olivucci M A Conical Intersection Controls the Deactivation of the Bacterial Luciferase Fluorophore. Angew. Chem., Int. Ed. 2014, 53, 9870–9875.

(13). Campbell ZT; Weichsel A; Montfort WR; Baldwin TO Crystal Structure of the Bacterial luciferase/Flavin Complex Provides Insight into the Function of the β Subunit. Biochemistry 2009, 48, 6085–6094. [PubMed: 19435287]

(14). Fisher AJ; Raushel FM; Baldwin TO; Rayment I Three-dimensional Structure of Bacterial Luciferase from Vibrio Harveyi at 2.4 Å Resolution. Biochemistry 1995, 34, 6581–6586. [PubMed: 7756289]

(15). Nicoli MZ; Meighen EA; Hastings JW Bacterial Luciferase, Chemistry of the Reactive Sulfhydryl. J. Biol. Chem. 1974, 249, 2385–2392. [PubMed: 4822498]

(16). Cohn DH; Mileham AJ; Simon MI; Nealson KH; Rausch SK; Bonam D; Baldwin TO Nucleotide Sequence of the LuxA Gene of Vibrio Harveyi and the Complete Amino Acid Sequence of the Alpha Subunit of Bacterial Luciferase. J. Biol. Chem. 1985, 260, 6139–6146. [PubMed: 3997817]

(17). Fiser A; Sali A ModLoop: Automated Modeling of Loops in Protein Structures. Bioinformatics 2003, 19, 2500–2501. [PubMed: 14668246]

(18). Campbell ZT; Baldwin TO; Miyashita O Analysis of the Bacterial Luciferase Mobile Loop by Replica-Exchange Molecular Dynamics. Biophys. J. 2010, 99, 4012–4019. [PubMed: 21156144]

(19). (a) Sugita Y; Okamoto Y Replica-Exchange Molecular Dynamics Method for Protein Folding. Chem. Phys. Lett. 1999, 314, 141–151.(b) Nymeyer H; Gnanakaran S; Garcia AE Atomic Simulations of Protein, Using the Replica Exchange Algorithm. Methods Enzymol. 2004, 383, 119–149. [PubMed: 15063649]

(20). (a) Torda AE; van Gunsteren WF Algorithms for Clustering Molecular Dynamics Configurations. J. Comput. Chem. 1994, 15 (12), 1331–1340.(b) Berendsen HJC; Van der Spoel D; Van Drunen R GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. Comput. Phys. Commun. 1995, 91, 43–56.(c) Lindahl E; Hess B; van der Spoel D GROMACS 3.0: a Package for Molecular Simulation and Trajectory Analysis. J. Mol. Model. 2001, 7, 306–317.(d) Abraham MJ; Van der Spoel DJ; Lindahl E; Hess B and the GROMACS development team. GROMACS: User Manual version 5.0.4, www.gromacs.org (2014).

(21). Hornak V; Abel R; Okur A; Strockbine B; Simmerling C Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. Proteins: Struct., Funct., Genet. 2006, 65, 712–725. [PubMed: 16981200]

(22). Roos BO Advances in Chemical Physics: Ab Initio Methods in Quantum Chemistry Part 2; Lawley KP, Ed.; Wiley Online Library: 1987; Vol. 69, pp 399–445.

(23). Andersson K; Malmqvist PA; Roos BO; Sadlej AJ; Wolinski K Second-Order Perturbation Theory with a CASSCF Reference Function. J. Phys. Chem. 1990, 94, 5483–5488.

(24). (a) Schapiro I; Ryazantsev MN; Ding WJ; Huntress MM; Melaccio F; Andruniow T; Olivucci M Computational Photobiology and Beyond. Aust. J. Chem. 2010, 63, 413–429.(b) El-Khoury PZ; Schapiro I; Huntress M; Melaccio F; Gozem S; Frutos LM; Olivucci M CRC Handbook of Organic Photochemistry and Photobiology; Griesbeck A, Oelgemoller M, Ghetti F, Eds.; CRC Press, Taylor & Francis Group: Boca Raton, FL, 2012; pp 1029–1056.(c) Gozem S; Melaccio F; Luk HL; Rinaldi S; Olivucci M Learning from Photobiology how to Design Molecular Devices Using a Computer. Chem. Soc. Rev. 2014, 43, 4019–4036. [PubMed: 24811294] (d) Roca-Sanjuan D; Aquilante F; Lindh R WIREs Comput. Mol. Sci. 2012, 2, 585–603.(e) Gozem S; Luk HL; Schapiro I; Olivucci M Theory and Simulation of the Ultrafast Double-Bond Isomerization of Biological Chromophores. Chem. Rev. 2017, 117, 13502–13565. [PubMed: 29083892]

(25). Ghigo G; Roos BO; Malmqvist P A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2). Chem. Phys. Lett. 2004, 396, 142–149.

(26). (a) Gozem S; Huntress M; Schapiro I; Lindh R; Granovsky AA; Angeli C; Olivucci M Dynamic Electron Correlation Effects on the Ground State Potential Energy Surface of a Retinal Chromophore Model. J. Chem. Theory Comput. 2012, 8, 4069–4080. [PubMed: 26605574] (b) Zobel JP; Nogueira JJ; Gonzalez L The IPEA dilemma in CASPT2. Chem. Sci. 2017, 8, 1482–1499. [PubMed: 28572908] (c) Kepenekian M; Robert V; Le Guennic B What zeroth-order Hamiltonian for CASPT2 adiabatic energetics of Fe(II)N6 architectures? J. Chem. Phys. 2009, 131, 114702. [PubMed: 19778137]

(27). http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.d.b.bh.html.

(28). Lin LY; Sulea T; Szitner R; Vassilyev V; Purisima EO; Meighen EA Modeling of the Bacterial Luciferase-Flavin Mononucleotide Complex Combining Flexible Docking with Structure-Activity Data. Protein Sci. 2001, 10, 1563–1571. [PubMed: 11468353]

(29). (a) Aquilante F; De Vico L; Ferré N; Ghigo G; Malmqvist PA; Neogrády P; Pedersen TB; Pitoná k M; Reiher M; Roos BO; Serrano-Andrés L; Urban M; Veryazov V; Lindh R MOLCAS 7: The Next Generation. J. Comput. Chem. 2010, 31, 224–247. [PubMed: 19499541] (b) Ponder JW; Richards FM An efficient newton-like method for molecular mechanics energy minimization of large molecules. J. Comput. Chem. 1987, 8, 1016.(c) Ferré N; Olivucci M Probing the rhodopsin cavity with reduced retinal models at the CASPT2//CASSCF/AMBER level of theory. J. Am. Chem. Soc. 2003, 125, 6868–6869. [PubMed: 12783530]

(30). (a) Koch H; Sanchez de Meras A; Pedersen TB Reduced scaling in electronic structure calculations using Cholesky decompositions. J. Chem. Phys. 2003, 118, 9481–9484.(b) Aquilante F; Malmqvist P; Pedersen TB; Ghosh A; Roos BO Cholesky Decomposition-Based Multiconfiguration Second-Order Perturbation Theory (CD-CASPT2): Application to the Spin-State Energetics of CoIII(diiminato)(NPh). J. Chem. Theory Comput. 2008, 4, 694–702. [PubMed: 26621084]

(31). (a) Granovsky AA Extended multi-configuration quasi-degenerate perturbation theory: The new approach to multi-state multi-reference perturbation theory. J. Chem. Phys. 2011, 134, 214113. [PubMed: 21663350] (b) Gozem S; Melaccio F; Lindh R; Krylov A; Granovsky AA; Angeli C; Olivucci M Mapping the excited state potential energy surface of a retinal chromophore model with multireference and EOM-CC methods. J. Chem. Theory Comput. 2013, 9, 4495–4506. [PubMed: 26589167] (c) Gozem S; Melaccio F; Valentini A; Filatov M; Huix-Rotllant M; Ferrè N; Frutos LM; Angeli C; Krylov A; Granovsky AA; Lindh R; Olivucci M On the Shape of Multireference, EOM-CC, and DFT Potential Energy Surfaces at a Conical Intersection. J. Chem. Theory Comput. 2014, 10, 3074–3084. [PubMed: 26588278]

(32). (a) Klaumü B; Kröner D; Saalfrank P (TD-)DFT Calculation of Vibrational and Vibronic Spectra of Riboflavin in Solution. J. Phys. Chem. B 2010, 114, 10826–10834. [PubMed: 20681576] (b) Kabir MP; Orozco-Gonzalez Y; Gozem S Electronic Spectra of Flavin in Different Redox and Protonation States: a Computational Perspective on the Effect of the Electrostatic Environment. Phys. Chem. Chem. Phys. 2019, 21, 16526–16537. [PubMed: 31312822] (c) Su D; Kabir MP; Orozco-Gonzalez Y; Gozem S; Gadda G Fluorescence Properties of Flavin

Semiquinone Radicals in Nitronate Monooxygenase. ChemBioChem 2019, 20, 1646–1652. [PubMed: 30748074]

(33). (a) Hedegård ED; Haugaard Olsen JM; Knecht S; Kongsted J; Aagaard Jensen HJ Polarizable embedding with a multiconfiguration short-range density functional theory linear response method. J. Chem. Phys. 2015, 142, 114113. [PubMed: 25796237] (b) Soderhjelm P; Husberg C; Strambi A; Olivucci M; Ryde U Protein Influence on Electronic Spectra Modeled by Multipoles and Polarizabilities. J. Chem. Theory Comput. 2009, 5, 649–658. [PubMed: 26610229] (c) Loco D; Polack E; Caprasecca S; Lagardère L; Lipparini F; Piquemal JP; Mennucci B A QM/MM approach using the AMOEBA polarizable embedding: from ground state energies to electronic excitations. J. Chem. Theory Comput. 2016, 12 (8), 3654–3651. [PubMed: 27340904]

(34). (a) Strickler SJ; Berg RA Relationship between Absorption Intensity and Fluorescence Lifetime of Molecules. J. Chem. Phys. 1962, 37, 814–822.(b) Turro NJ; Ramamurthy V; Scaiano JC Principles of molecular photochemistry: an introduction; University science books. 2019.

(35). Lee J; Wang Y; Gibson BG Electronic Excitation Transfer in the Complex of Lumazine Protein with Bacterial Bioluminescence Intermediates. Biochemistry 1991, 30, 6825–6835. [PubMed: 2069948]

(36). (a) González-Luque R; Garavelli M; Bernardi F; Merchán M; Robb MA; Olivucci M Computational Evidence in Favor of a Two-State, Two-Mode Model of the Retinal Chromophore Photoisomerization. Proc. Natl. Acad. Sci. U. S. A. 2000, 97, 9379–9384. [PubMed: 10944211] (b) Drobizhev M; Hughes TE; Stepanenko Y; Wnuk P; O'Donnell K; Scott JN; Callis PR; Mikhaylov A; Dokken L; Rebane A Primary Role of the Chromophore Bond Length Alternation in Reversible Photoconversion of Red Fluorescence Proteins. Sci. Rep. 2012, 2, 688. [PubMed: 23008753] (c) Bona i -Koutecký V; Koutecký J; Michl J Neutral and Charged Biradicals, Zwitterions, Funnels in S1, and Proton Translocation: Their Role in Photochemistry, Photophysics, and Vision. Angew. Chem., Int. Ed. Engl. 1987, 26, 170–189.(d) Gozem S; Schapiro I; Ferré N; Olivucci M The Molecular Mechanism of Thermal Noise in Rod Photoreceptors. Science 2012, 337, 1225–1228. [PubMed: 22955833]

(37). Hasegawa JY; Bureekaew S; Nakatsuji H SAC-CI Theoretical Study on the Excited States of Lumiflavin: Structure, Excitation Spectrum, and Solvation Effect. J. Photochem. Photobiol., A 2007, 189, 205–210.

(38). Baker NA; Sept D; Joseph S; Holst MJ; McCammon A Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. Proc. Natl. Acad. Sci. U. S. A. 2001, 98, 10037–41. [PubMed: 11517324]

(39). (a) Laptenok SP; Bouzhir-Sima L; Lambry JC; Myllykallio H; Liebl U; Vos MH Ultrafast Real-Time Visualization of Active Site Flexibility of Flavoenzyme Thymidylate Synthase ThyX. Proc. Natl. Acad. Sci. U. S. A. 2013, 110, 8924–8929. [PubMed: 23671075] (b) Yang H; Luo G; Karnchanaphanurach P; Louie TM; Rech I; Cova S; Xun L; Xie XS Protein Conformational Dynamics Probed by Single-Molecule Electron Transfer. Science 2003, 302, 262–266. [PubMed: 14551431] (c) Brazard J; Usman A; Lacombat F; Ley C; Martin MM; Plaza P; Mony L; Heijde M; Zabulon G; Bowler C Spectro–Temporal Characterization of the Photoactivation Mechanism of Two New Oxidized Cryptochrome/Photolyase Photoreceptors. J. Am. Chem. Soc. 2010, 132, 4935–4945. [PubMed: 20222748]

(40). Mukherjee A; Weyant KB; Walker J; Schroeder CM Directed Evolution of Bright Mutants of an Oxygen-Independent Flavin-Binding Fluorescent Protein from Pseudomonas Putida. J. Biol. Eng. 2012, 6, 20. [PubMed: 23095243]
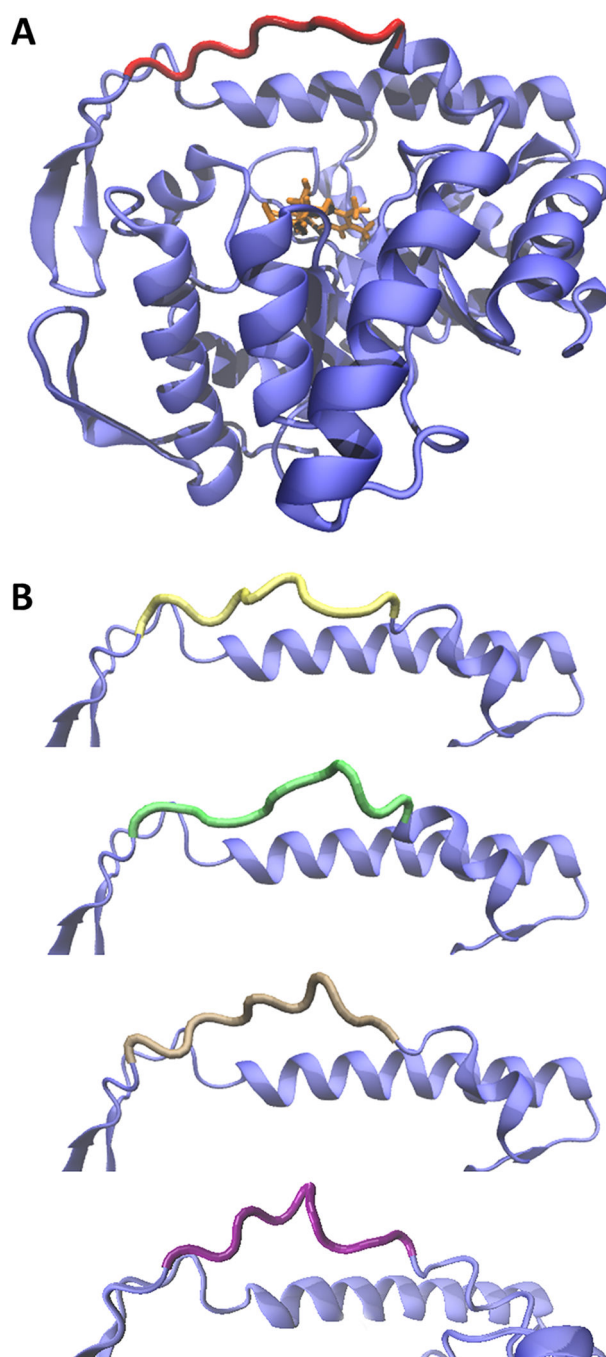
Figure 1.
(A) The structure obtained from REMD simulations of the a-subunit of bacterial luciferase in the presence of the emitting intermediate FMNHOH (**2**, in orange). The structure of the proximal mobile loop was extrapolated from the cluster analysis (in red). (B) Comparison of the conformations of the mobile loops obtained from the other four most populated clusters.
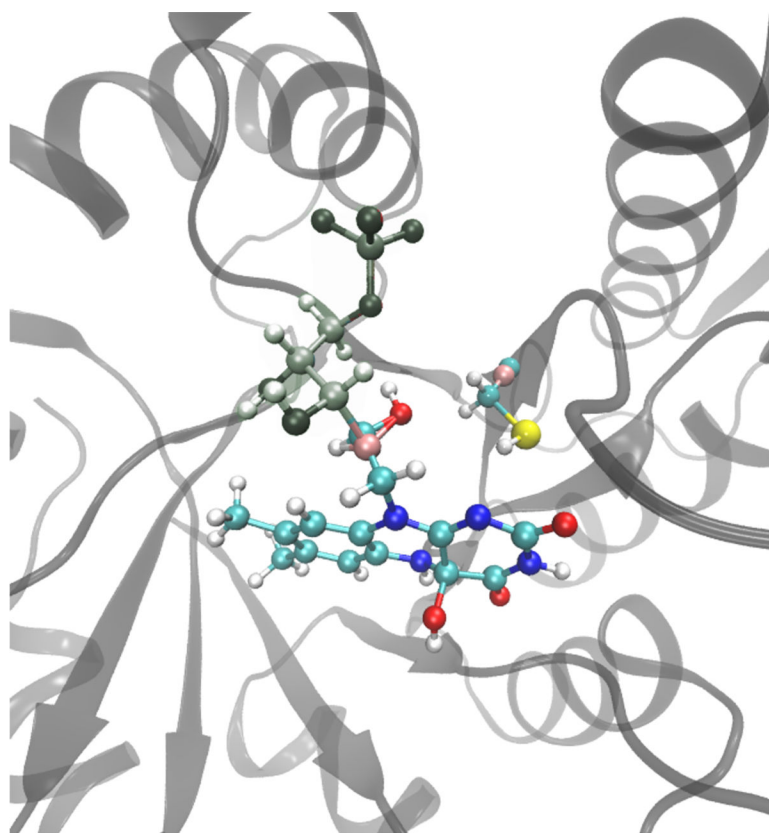
Figure 2.
Schematic representation of the QM/MM setup including Cys106 in the QM subsection. The colored ball-and-stick model represents the QM subsystem, containing the lumiflavin group of FMNHOH (**2**) and the –CH$_2$SH part of Cys106. The gray ribbons in the background represent the protein environment. The hydrogen link atoms, which cap the vacancies at the frontier of QM/MM regions, are highlighted in pink. Another model, that places the entirety of Cys106 in the MM region, is otherwise identical.
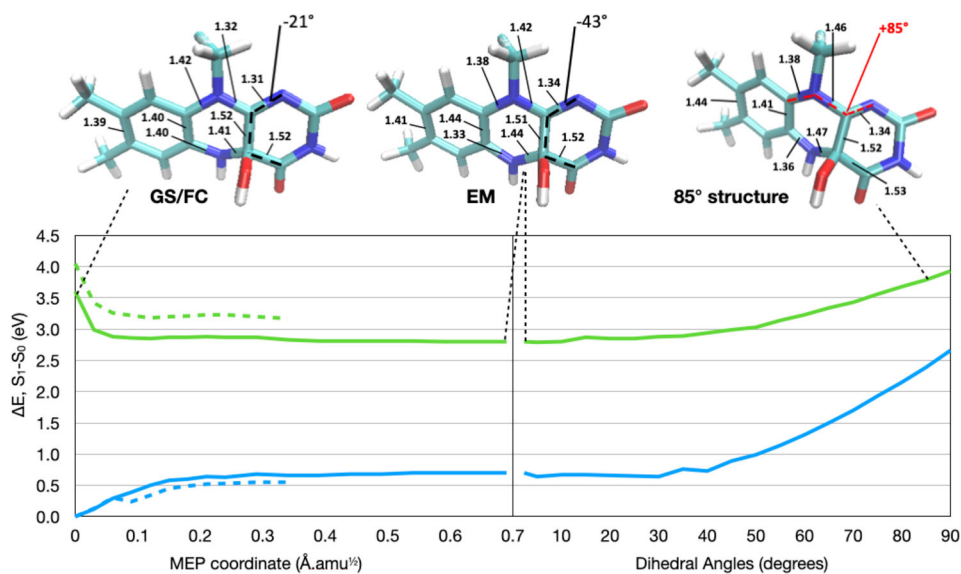
Figure 3.

Left: CASPT2//CASSCF S_0 (blue) and S1 (green) energy profiles along the MEP connecting the FC to the energy minimum (EM). Calculations obtained with Cys106 in the QM region are shown as solid lines, while calculations with Cys106 in the MM region are shown as dashed lines. Right: The results of the C9a–N10–C10a–N1 dihedral relaxed scan, performed in intervals of 5° starting from the EM. The energy gap between S_0 and S_1 decreases from 2.1 eV at the EM to 1.3 eV at 90°. Top: Structures of the GS/FC, EM, and 85° constrained structure. All bond lengths that change by more than 0.02 Å during the MEP are labeled for the structures. We also indicate the C4–C4a bond length, which changes minimally. Moreover, we label the C4–C4a–C10a–N1 dihedral, representing out-of-plane distortion of the fluorophore. For clarity, the Cys106 side chain is not included in the structure representations.
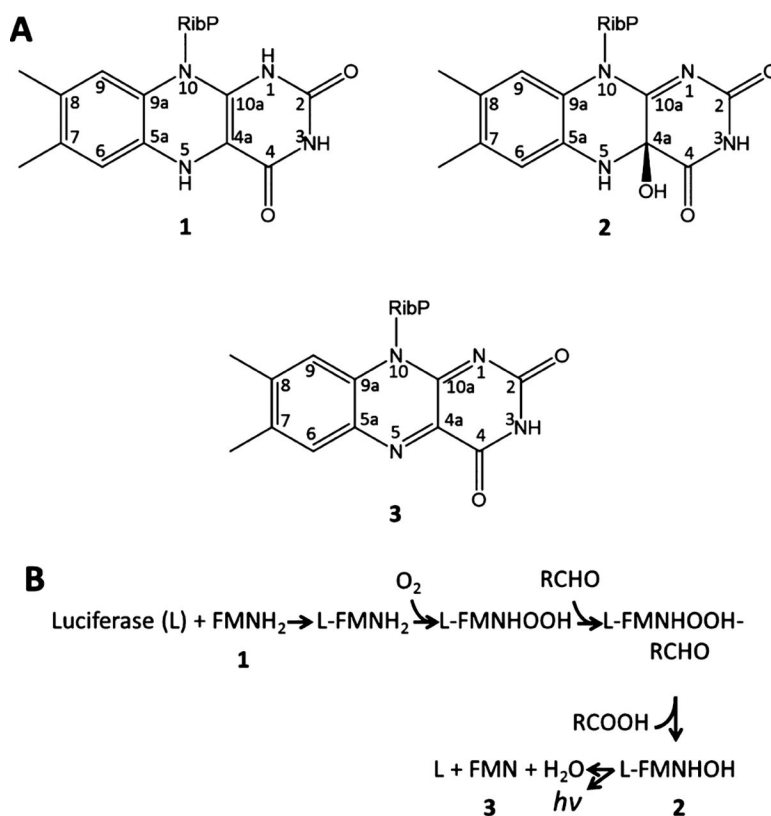
Scheme 1. Structure of the Isoalloxazine Moiety in FMNH$_2$ (1), FMNHOH (2), and FMN (3); (B) Proposed Mechanism of Bioluminescence in Bacterial Luciferase^a

. (A) ^{a}Adapted with permission from ref 9. Copyright 2004 American Chemical Society.

Table 1.

Computed Absorption and Emission Wavelengths with Different CASPT2 Methods and Basis Sets

SS or MS	active space	IPEA	basis set	λ_{abs}, eV (nm)	λ_{em}, eV (nm)
SS	14,11	0	6–31G*	3.67 (338)	2.20 (564)
MS	14,11	0	6–31G*	3.71 (334)	2.34 (530)
SS	14,11	0.25	6–31G*	4.13 (300)	2.54 (488)
MS	14,11	0.25	6–31G*	4.16 (298)	2.67 (464)
SS	14,11	0	ANO-L-VDZP	3.50 (354)	2.06 (603)
MS	14,11	0	ANO-L-VDZP	3.57 (347)	2.29 (542)
SS	14,11	0.25	ANO-L-VDZP	4.00 (310)	2.43 (511)
MS	14,11	0.25	ANO-L-VDZP	4.05 (306)	2.62 (474)
SS	14,11	0	ANO-L-VTZP	3.42 (362)	2.01 (616)
MS	14,11	0	ANO-L-VTZP	3.49 (355)	2.27 (547)
SS	14,11	0.25	ANO-L-VTZP	3.94 (314)	2.40 (517)
MS	14,11	0.25	ANO-L-VTZP	3.99 (311)	2.60 (476)
SS	18,15	0	6–31G*	3.93 (316)	2.08 (595)
MS	18,15	0	6–31G*	3.36 (369)	2.08 (597)
SS	18,15	0.25	6–31G*	4.35 (285)	2.37 (524)
MS	18,15	0.25	6–31G*	3.87 (321)	2.36 (525)
SS	18,15	0	ANO-L-VDZP	3.74(331)[a]	1.95(636)[a]
MS	18,15	0	ANO-L-VDZP	3.12(398)[a]	1.94(638)[a]
SS	18,15	0.25	ANO-L-VDZP	4.21(295)[a]	2.25(551)[a]
MS	18,15	0.25	ANO-L-VDZP	3.68(337)[a]	2.25(551)[a]
MS	18,15	0.25	ANO-L-VTZP	3.62(342)[b]	2.24(554)[b]
Experiment				(~360)[b]	(~490)[d]

[a]This calculation was performed with Cholesky decomposition to reduce computational cost.[30]

[b]This data is obtained using an extrapolation from the effect of triple-ζ relative to double-ζ for the 14,11 active space result.

[c]Reference 4.

[d]Reference 9.