Article

Computational Investigation of Structural and Spectroscopic Properties of LOV-Based Proteins with Improved Fluorescence

Felipe Cardoso Ramos,* Lorenzo Cupellini, and Benedetta Mennucci*

Read Online

ACCESS | 📊 Metrics & More | 📰 Article Recommendations | 🆔 Supporting Information

ABSTRACT: Flavin-based fluorescent proteins are a class of fluorescent reporters derived from light, oxygen, and voltage (LOV) sensing proteins. Through mutagenesis, natural LOV proteins have been engineered to obtain improved fluorescence properties. In this study, we combined extended classical Molecular Dynamics simulations and multiscale Quantum Mechanics/Molecular Mechanics methods to clarify the relationship between structural and dynamic changes induced by specific mutations and the spectroscopic response. To reach this goal we compared two LOV variants, one obtained by the single mutation needed to photochemically inactivate the natural system, and the other (iLOV) obtained through additional mutations and characterized by a significantly improved fluorescence. Our simulations confirmed the "flipping and crowding" effect induced in iLOV by the additional mutations and revealed its mechanism of action. We also showed that these mutations, and the resulting differences in the composition and flexibility of the binding pockets, are not reflected in significant shifts of the excitation and emission energies, in agreement with the similarity of the spectra measured for the two systems. However, a small but consistent reduction was found in the Stokes shift of iLOV, suggesting a reduction of the intermolecular reorganization experienced by the chromophore after excitation, which could slow down its internal conversion to the ground state and improve the fluorescence.
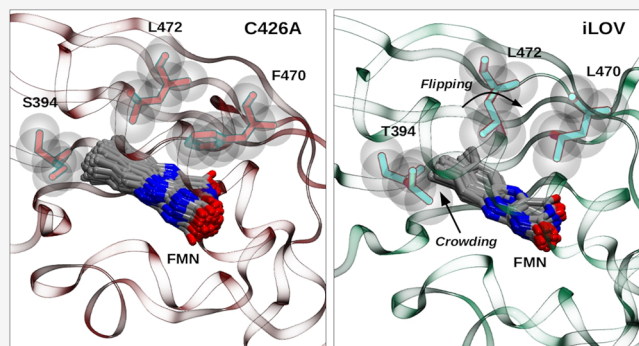
1. INTRODUCTION

Green fluorescent protein (GFP) has revolutionized the imaging of dynamic processes within living cells.[1] However, the use of GFP as *in vivo* reporters is limited by some environmental and cellular factors impeding either chromophore formation or fluorescence activity.[2,3] In this context, the so-called flavin-based fluorescent proteins (FbFPs) emerged as an alternative class of FPs.[4−7] FbFPs are derived from a highly conserved family of blue light photoreceptors known as light, oxygen, and voltage (LOV) sensing proteins. In nature, LOV proteins typically associate with flavin mononucleotide (FMN) to function as blue-light photoreceptors and regulate a serie of cellular processes in both bacteria (stress response and virulence) and plants (phototaxis).[8,9]

Natural LOV domains bind FMN noncovalently and, upon UVA/blue-light excitation, undergo a reversible photocycle involving the formation of a covalent bond between the chromophore and a conserved cysteine residue with complete loss of fluorescence.[10] The substitution of the cysteine residue abolishes LOV domain photochemistry and recovers the fluorescence of the bound FMN.[11] LOV domains photochemically inactivated in this way are inherently fluorescent; however, additional mutagenesis was needed to further improve their fluorescence and photostability.[11−13] Random and structure-based engineering methods have been combined to generate a large pool of mutants, which are expressed and then selected according to specific properties, such as fluorescence quantum yield, thermal stability, and photobleaching reversibility.[11−15] In particular, large attention has been given to a specific class of LOV-based reporter variants with improved properties obtained from the LOV2 domain of *Arabidopsis thaliana* phototropin 2 (phot2) through a directed evolution approach based on the DNA shuffling technique and screening toward enhanced fluorescence.[11]

As a first mutation, the photoactive cysteine (Cys426 of *Arabidopsis* phot2) was replaced with alanine to achieve the photochemically inactivated derivative C426A. Five additional mutations (S394T, S409G, I452T, F470L, and M475V, see Figure 1) were introduced to such a derivative leading to an improved LOV (iLOV) showing a substantial increase in fluorescence intensity and emission quantum yield with respect to the single mutant C426A.[11,15]
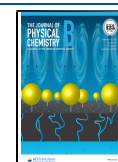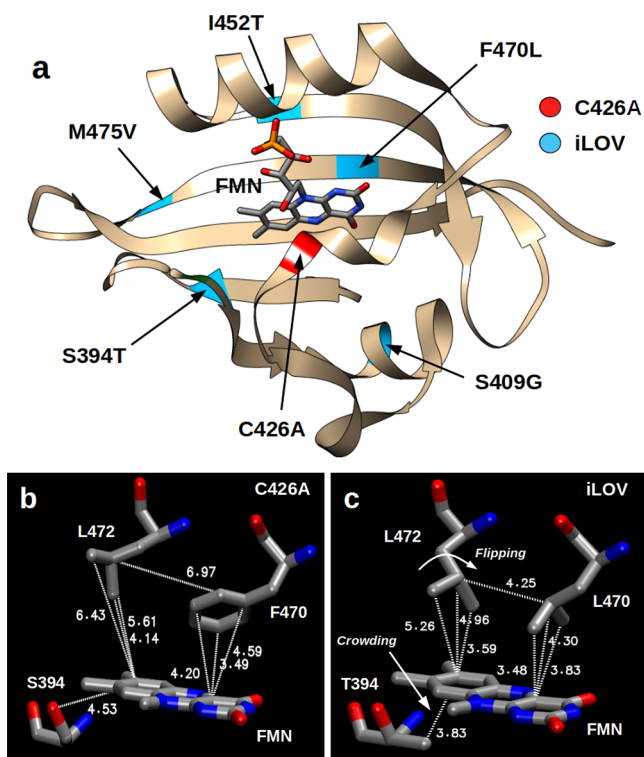
Figure 1. (a) Mutations present in C426A and iLOV with respect to the wild-type LOV2 domain. In the former, the photoactive cysteine was replaced with alanine and in the latter, five additional mutations (S394T, S409G, I452T, F470L, and M475V) were introduced. (b,c) Network of residues involved in the "crowding and flipping" model. The distance values are in angstroms and were obtained from 4EER (C426A) and 4EES (iLOV) crystal structures. Residue numbering is based on the phot2 protein sequence, which is the same found in the PDB structures.

The iLOV domain was functionally and structurally characterized by Christie et al.[12] From the crystallographic structures, the authors proposed that the four additional mutations of iLOV stabilize its structure by increasing packing interactions, especially in the flavin-binding cavity. According to this model, the FMN isoalloxazine ring would be rigidified through a "crowding and flipping" process undergone by the T394 and L472 side chains. The crowding is related to the replacement of a serine residue with a threonine residue in the flavin-binding pocket of iLOV (S394T mutation, see Figure 1) while the flipping effect is the rotation of a leucine (L472) side chain in the direction of the chromophore induced by the F470L mutation in iLOV. Both modifications are supposed to increase the stability of flavin in the pocket.

Since all the experimentally obtained FbFPs have similar spectral features (maximum absorption and fluorescence at 447 and 493 nm, respectively),[11,12,15−18] further optimization efforts have been focused on enhancing the emission color range by introducing single-point mutations in the FMN binding site. The goal in this case is to reach the "biotransparent" window (650−900 nm).[19,20] Along this research line, a fundamental role has been played by computational studies.[17,21−26] In particular, these studies have clearly shown that a detailed conformational analysis of the protein is the necessary prerequisite for any research efforts aiming to design variants with further improved fluorescence properties.[26]

Following these findings, here we have performed microsecond-scale simulations of the C426A and iLOV variants originally engineered by Chapman et al.[11] The goal is 2-fold. First, we aim at characterizing the conformational dynamics of the two systems in solution. Second, we want to verify that the "crowding and flipping" is not an artificial effect due to the crystal structure but it occurs in solution. This goal is achieved through an "in silico" experiment which compares the dynamic interactions between the chromophore and the neighboring residues in the two variants and the resulting absorption spectra and Stokes shifts. The spectroscopic properties are here simulated through a quantum mechanics/molecular mechanics approach which accounts for mutual polarization effects between the QM chromophore and the classical protein residues and solvent molecules.[27]

From this investigation, it comes out that the FMN-protein interactions in the binding pockets are characterized by a large flexibility in the microsecond time scale in both systems. However, this flexibility does not seem to significantly affect the excitation and emission energies. On the other hand, the additional mutations present in iLOV are responsible for a reduced mobility of the isoalloxazine ring of FMN within the pocket. This effect also slightly reduces the extent of intermolecular reorganization experienced by the chromophore after excitation.

2. COMPUTATIONAL DETAILS

2.1. Molecular Dynamics. MD simulations were performed on the two LOV variants described by Christie et al. through crystallography analysis:[12] the photochemically inactivated LOV2 domain from *A. thaliana*, named C426A mutant (PDB: 4EER), and the iLOV protein containing five single-point mutations (S394T, S409G, C426A, I452T, and F470L) (PDB: 4EES). In Figure S1 of the Supporting Information we report the multiple sequence alignment of wild-type LOV2 domain, C426A, and iLOV.

By using the *tleap* module of AmberTools,[28] we carried out the preparation of the two investigated systems, their solvation within a truncated octahedron TIP3P water box (with ~23 000 water molecules), as well as the addition of Na$^+$ and Cl$^-$ ions at 0.15 M. Extra Na$^+$ ions were also add so as to achieve system charge neutrality. The two simulated systems have, in total, approximately 70 000 particles.

The system minimization was done by first minimizing hydrogen atoms, next the solvent components and the protein-chromophore complex, and finally the whole system. For the MD simulations, we first performed the system heating which was divided into two steps: the first one from 0 to 100 K (5 ps in the NVT ensemble) constraining all the system with a harmonic potential (4.0 kcal mol^{-1} Å^{-1}) and the second one from 100 to 300 K (100 ps in the NPT ensemble) constraining just the protein backbone. Next, a 5 ns NPT equilibration step at 300 K was done initially applying the same constraint on the protein backbone but releasing the harmonic force constant by 1 kcal mol^{-1} Å^{-1} at each 1 ns. Then, we carried out 5 μs of production at 300 K in the NPT ensemble for both C526A and iLOV systems. Two replica MDs were performed on each system, for a total sampling time of 20 μs.

Both minimization and MD simulations were performed using the Amber16 program employing the *ff14SB*[29] force field for protein. The parameters for the FMN chromophore were obtained from the literature.[30] In all MD simulations, the time step was set to 2 fs. For system temperature and pressure

control we employed a Langevin thermostat and an anisotropic barostat, both implemented in the Amber16. The particle-mesh Ewald algorithm[31] was used to describe the long-range electrostatic interactions. The MD analysis was performed by using both the *cpptraj*[32] module of AmberTools and locally developed tools.

The network of hydrogen bonds around the FMN chromophore was investigated by computing distances between the FMN and H-bonding residues, as well as among these residues, along both replicas of each system. A principal component analysis (PCA) on these distances, followed by a clustering with the HDBSCAN algorithm,[33,34] was used to find different conformations of the FMN pocket (clusters). Ten structures were randomly extracted from each cluster to be employed in the following calculations. The interaction of nonpolar residues with the FMN ring was quantified by computing the overlap integral between the pseudoelectronic densities of the FMN ring and the residue side chain:

$$\text{Ovlp} = \int d\mathbf{r}\, \tilde{\rho}_A(\mathbf{r}) \tilde{\rho}_B(\mathbf{r}) \tag{1}$$

where the pseudoelectronic densities $\tilde{\rho}_X(\mathbf{r})$ $(X = A,B)$ were computed as a sum of Gaussian distributions centered on the heavy atoms of each fragment:

$$\tilde{\rho}_X(\mathbf{r}) = \sum_{i \in X} \frac{1}{\sigma_i \sqrt{2\pi}} \, e^{-(\mathbf{r}-\mathbf{r}_i)^2/(2\sigma_i^2)} \tag{2}$$

The standard deviations of the Gaussian distributions, σ_i, were taken equal to the van der Waals radii of the elements.

2.2. QM/MM(Pol) Calculations. The multiscale calculations, for both crystal structures and MD snapshots, were divided into four steps: (i) ground state (S_0) geometry optimization; (ii) vertical excitation energy calculation; (iii) excited state (S_1) geometry optimization; and (iv) vertical emission energy calculation. The initial coordinates for step i were obtained from the crystal structures and from configurations extracted from the MD trajectories, and the optimized S_0 structures were used as input for step iii. In all steps, only the isoalloxazine group of FMN was included in the QM region while the ribityl tail was treated at classical level (see Figure 2).

All geometry optimizations (steps i and iii) were performed at the ONIOM(QM:MM) level,[35,36] with the QM subsystem allowed to move and the rest of the system (the ribityl tail of FMN and the protein) kept frozen. For the configurations
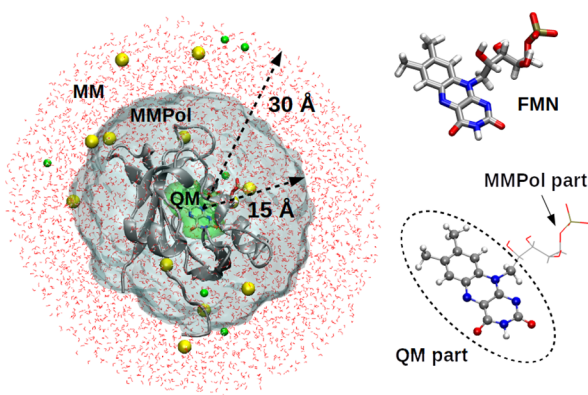

Figure 2. System partitioning employed in the QM/MMPol/MM calculations.

obtained from MD, a solvation sphere of about 20 Å of radius was defined around the chromophore. The Na$^+$ and Cl$^-$ ions were stripped from the solvent. The QM part was described at the B3LYP/6-31G(d) level for ground-state optimizations (step i) and TD-ωB97X-D/6-31+G(d) level for the excited state optimizations (step iii). The MM part was described by the same force field used for the MD simulation.

All excitation and emission calculations (steps ii and iv) were performed using a three-layer model (see Figure 2): (1) the QM subsystem described at TD-ωB97X-D/6-31+G(d) level; (2) the protein, the water molecules and ions within 15 Å from FMN described using a polarizable MM model (from now on MMpol), (3) the water molecules and the ions in a shell of 15−30 Å from the chromophore treated at the MM level. The polarizable first shell was described with the *pol12 AL* Amber force field.[37,38]

We note that the latter functional gives significantly blue-shifted excitation energies with respect to experiments (about +0.4 eV). A better agreement with experiments was found for B3LYP but, with that functional, the characterization of the excited state was unclear due to artificial mixing. As here the goal was to obtain consistent excitation and emission energies, to predict Stokes shifts, we preferred to use ωB97X-D and obtain well characterized excited states with excitation and emission energies of comparable accuracy, even if both are blue-shifted.

All the calculations have been performed with a locally modified version of Gaussian 16[39] where the QM/MMpol approach has been implemented.[40,41]

2.3. Lineshape Calculations. The homogeneous line shape $\sigma(\omega - \omega_{01})$ of the flavin chromophores was computed using the second-order cumulant expansion in the displaced harmonic oscillator (DHO) formalism.[42]

$$\sigma(\omega - \omega_{01}) = \Re \int_0^\infty dt \exp\left[i(\omega - \omega_{01})t + \int_0^\infty d\omega \right.$$
$$\frac{J(\omega)}{\omega^2}\left(\coth\left(\frac{\beta\hbar\omega}{2}\right)(\cos(\omega t) - 1) \right.$$
$$\left. - i(\sin(\omega t) - \omega t) \right) \Bigg] \tag{3}$$

where the spectral density (SD) $J(\omega)$ is defined in order to encode the vibronic coupling with all normal modes:

$$J(\omega) = \pi \sum_k S_k \omega_k^2 (\delta(\omega - \omega_k) - \delta(\omega + \omega_k)) \tag{4}$$

where S_k is the Huang−Rhys factor along mode k. The Huang−Rhys factors were calculated on the crystal structures of both iLOV and C426A, by first computing the ground-state normal modes, in the same ONIOM scheme used for the optimizations in step (i). The excited-state gradient computed at the B3LYP/6-31G(d) level of theory was then projected onto each normal mode to get the Huang−Rhys factor.

The final spectra were computed, for each cluster, by summing the contribution of N structures with excitation energies $\omega_{01}^{(j)}$ and transition dipoles $\boldsymbol{\mu}_{01}^{(j)}$:

$$A(\omega) = \omega \frac{1}{N} \sum_{j=1}^N \left| \boldsymbol{\mu}_{01}^{(j)} \right|^2 \sigma(\omega - \omega_{01}^{(j)}) \tag{5}$$

where $\sigma(\omega)$ is the absorption line shape computed as described above.
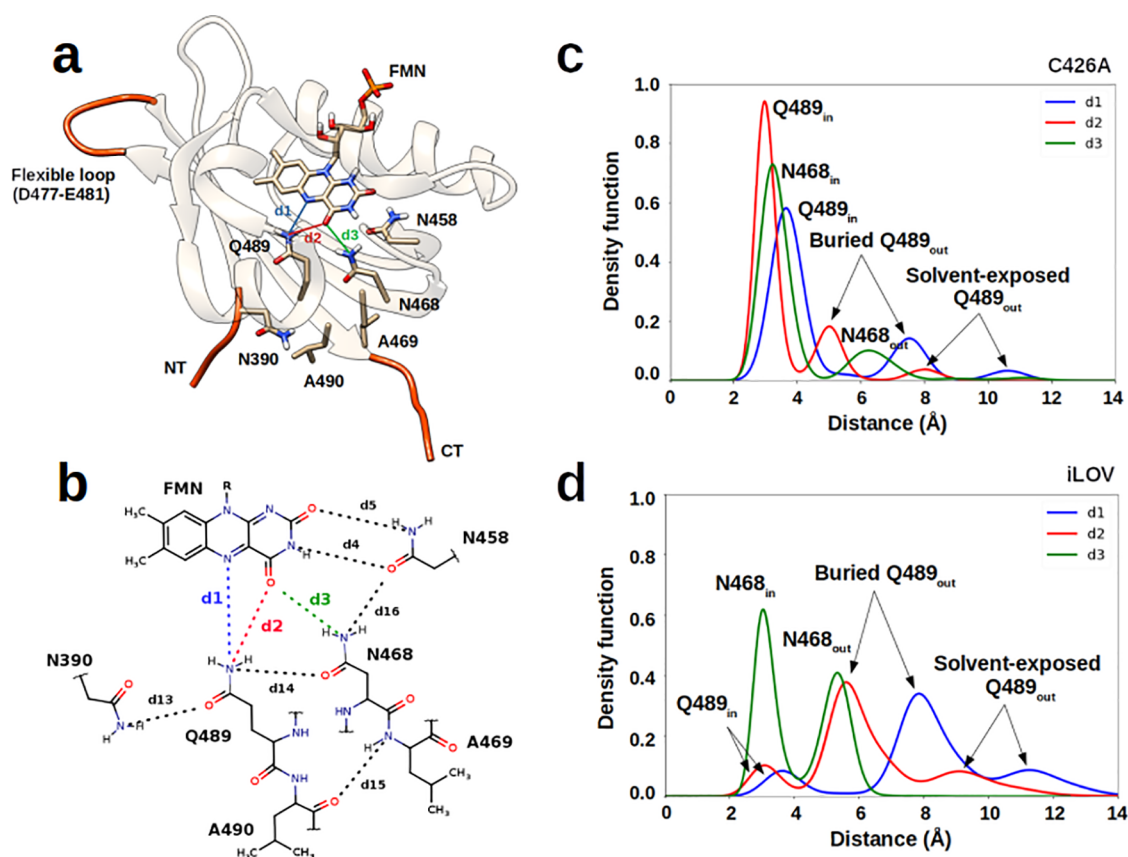
Figure 3. (a) Crystal structure of iLOV showing the conserved flexible residues in the FMN-binding site. In red color are also indicated the regions of the protein that showed great flexibility during our MD simulations (either in C426A or iLOV). NT and CT refer to C- and N-terminal ends. (b) Map of interactions indicating the distances employed in the H-bond analysis (d1−5, d13−16). (c,d) Distributions obtained for FMN-Q489 and FMN-N468 H-bond distances (d1−3) for C426A and iLOV.

3. RESULTS AND DISCUSSION

3.1. Structural Analysis. To achieve a clear description of the structural and dynamic specificities of the two systems which can be related to their different fluorescence behavior, we split the analysis in two parts: one focused on the H-bonding network which characterizes the FMN-binding pocket and the other investigating the specific residues involved in the "crowding and flipping" mechanism, illustrated in Figure 1b,c.

As a preliminary and general comment, however, we note that in our microsecond-long MD simulations C- and N-termini showed to be highly flexible in both C426A and iLOV domains, with up to 4 Å RMSD from the crystal structure (Figure S2). Excluding those termini, however, both C426A and iLOV showed a rather rigid protein backbone, with the average RMSD value around 2 Å with respect to the crystallographic structures (Figures S3 and S4). No significant changes in the secondary structure were observed. In one replica, C426A showed increasing RMSD, up to 3 Å after ~4000 ns. This deviation arises from a conformational change in a flexible loop between residues D477 and E481 (Figure S3). As this loop is external to the protein and far from the binding pocket (Figure 3a), we can exclude an influence of its conformation on the properties of the chromophore.

Moving now to the analysis of the binding pocket, we start by focusing on the extended H-bonding network which is characterized by a total of 22 distances: 12 protein-FMN (d1−12), 8 protein−protein interactions (d13−20) nearby the FMN-binding pocket, and 2 intra-FMN interactions (d21,

d22). The full map of the H-bonding distances are reported in Figure S5.

In general, most of the H-bonding interactions here analyzed showed to be stable in both C426A and iLOV during the whole MD simulations. In particular, all the interactions between the ribityl tail of FMN and side chains N425, R427, Q430, and R443 (d7−12) present in the crystal structures were conserved along the MD trajectories. The same occurred for the protein−protein interactions involving the same four residues (d17−20), for the two intra-FMN H-bonds (d21, d22), as well as for the H-bonds between the isoalloxazine ring of FMN and the N458 side chain (d4, d5). On the other hand, differences with respect to the crystal structures have been found for the H-bonding interactions involving residues N390, N458, N468, A469, Q489, and A490 (d1−5, d13−16) (see Figure 3a,b). To give an example of these differences, in Figure 3c,d we compare the distributions of distances between FMN and either glutamine (Q489) or asparagine (N468) residues, in the two systems.

These findings can be explained by noticing that N468 and Q489 chains adopt multiple conformations along the MD trajectories, which correspond to different distances from FMN (Figure 3c,d). As it can be seen from the figures, some conformations are oriented toward the chromophore, with distances less than 4 Å; such conformations are referred to as *in*. Other conformations are instead either solvent-oriented or "buried", and present distances longer than 5 Å. Such conformations are referred to as *out*. Importantly, we have
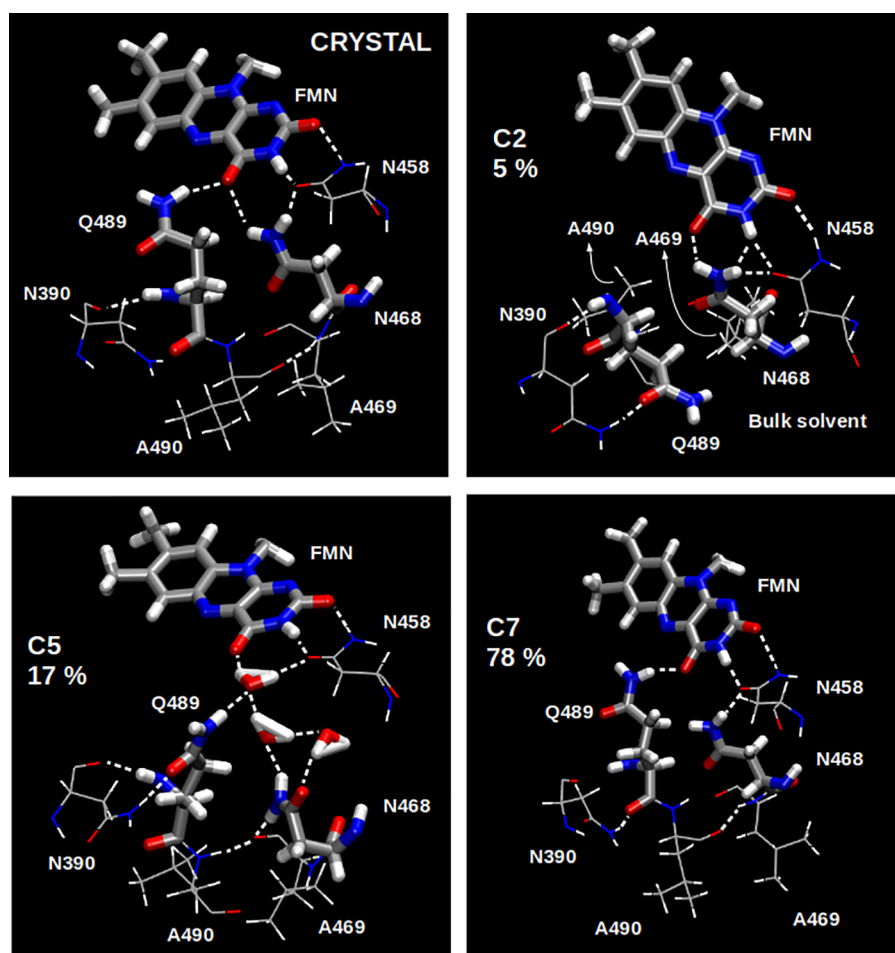
Figure 4. Comparison of the binding pocket in the crystal structure and in representative configurations for the three clusters obtained for C426A (C2, C5, and C7). The percentage values indicate the population of each cluster in the simulated trajectories.

observed the occurrence of *out* conformations of Q489 after at least 300 ns for iLOV, and after more than 2 μs for C426A (Figures S6, S7). This suggests that the crystal conformation is a local minimum in the free-energy landscape of these systems, which might have been stabilized by crystal packing. The long time scales needed to equilibrate the protein, however, suggest caution in interpreting the relative population of *in* and *out* conformations.

All these different conformations of the FMN pocket were clusterized on the basis of distances d1−5 and d13−16. The clustering algorithm was able to distinguish Q489$_{in}$ and several types of Q489$_{out}$ conformations (Figure S6, S7). After excluding clusters with population <1%, we obtained three clusters for C426A (named C2, C5, and C7) and six clusters for iLOV (named C0, C1, C2, C3, C6, and C7). The larger number of iLOV clusters with respect to C426A arises from a greater population of the Q489$_{out}$ and N468$_{out}$ side chain conformations. A comparison of the network of interactions of FMN with the residues of the binding pocket in the crystal structures and representative configurations for the different clusters are reported in Figure 4 and Figure 5, for C426A and iLOV, respectively. For the latter, only the three most populated clusters are shown while the others are reported in Figure S8.

By analyzing the different clusters we can notice that the Q489$_{out}$ and N468$_{out}$ orientations favor the entry of water molecules into the active site in both C426A and iLOV. These

water molecules can replace Q489 and/or N468 in the H-bonding network with FMN and form a bridge between the FMN and the residues. In such a case, the Q489 side chain adopts a buried conformation Q489$_{out}$ relatively closer to the FMN with respect to the solvent-exposed Q489$_{out}$ orientation. The two FMN-N458 H-bonds (d4,5), in turn, showed to be quite persistent in both C426A and iLOV (see Figure 4, 5). No *out* orientations were observed for N458.

Here we will use "IN" to refer to the structures in which both Q489$_{in}$ and N468$_{in}$ occur simultaneously, similarly to what found in the crystal structures. And we will use "OUT" to refer to structures in which either Q489$_{out}$ or N468$_{out}$ are present. Using this classification, we observe that one cluster of C426A (C7) and one of iLOV (C1) are of type IN. Concerning the OUT structures, our simulations indicate that the Q489$_{out}$ side chain can form H-bonds with N390 and N468 (d13 and d14 in Figure S9). Intriguingly, substituting N390 with a serine was shown to improve photostability of the iLOV *in vivo*, probably by restricting the movement of Q489.[12] In addition, we observed interactions between Q489$_{out}$ and water molecules in the active site or from the bulk solvent when Q489 is solvent-exposed (see Figures 4 and 5 and Figure S8). Unlike Q489, N468$_{out}$ is never solvent-exposed, but it adopts a buried conformation in both systems. In most of the OUT structures analyzed, the N468$_{out}$ conformation is characterized by water molecules bound to the FMN oxygen. Sometimes, the hydrogen bond between FMN and N468
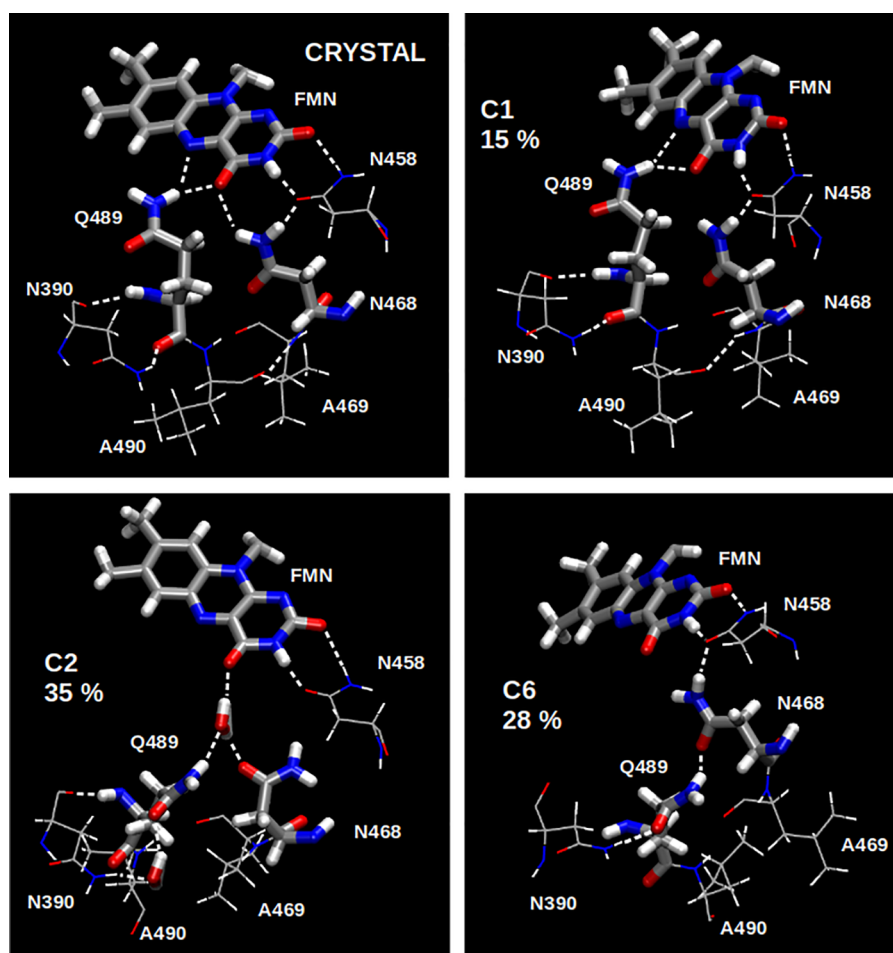
Figure 5. Comparison of the binding pocket in the crystal structure and in representative configurations for the three most populated clusters obtained for iLOV (C1, C2 and C6). The percentage values indicate the population of each cluster in the simulated trajectories.

occurs through a bridging water molecule while an interaction between N458 and N468 is established (see C3 in Figure S8).

Previous MD studies of iLOV or natural LOV domains have also reported the change of orientation undergone by the glutamine in the active site.[17,23−25,43] Experiments and MD simulations showed that K489 is as flexible as glutamine, and prefers an *out* conformation.[17] MD studies on the LOV1 domain from *Chlamydomonas reinhardtii* also pointed out that the H-bonding interaction between FMN and Q120 side chain (equivalent to Q489 in iLOV) is also unstable, persisting for only 33% of the simulation time (60 ns).[44] Similarly, Lokhandwala et al.[45] have observed buried and solvent exposed conformations for the Q204 side chains in MD studies involving a short LOV domain from *Trichoderma reesei*. The $N468_{out}$ conformer (or its equivalent in protein sequence) was not reported in previous MD studies, probably due to the short time windows investigated by those simulations (30 to 60 ns). In our simulations, no significant conformational changes of the N468 side chain were observed before 300 ns, suggesting that long simulation times or enhanced sampling methods are needed to assess the conformation of the flavin-binding side chains.

Although largely reported in MD simulation studies, the flexibility of conserved residues in the FMN biding pocket has not been observed in the crystal structures, which consistently present a $Q489_{in}$ conformation.[12] As suggested before,[23] the more compact $Q489_{in}$ conformation might be stabilized by crystal packing. The functional relevance of the dynamics of such residues in natural LOV proteins also remains unclear, but they may have implications for both decay pathways and signal transduction.[44−46] Regarding the artificial LOV domains, it has been suggested that the Q489 side chain may be involved in both spectral and fluorescence-efficiency tuning.[23]

Instead, Christie et al.[12] used the evidence given by the crystal structure to propose that the fluorescence increase in iLOV is mainly related to the optimized protein-chromophore van der Waals interactions induced by the "crowding and flipping" mechanism illustrated in Figure 1b,c. To assess if the "flipping" of the side chain of L472, observed in the crystal as a consequence of the F470L mutation, is conserved also in solution, we monitored the dihedral angles of the 470 and 472 apolar side chains present in the active-site of iLOV and C426A (Figure 6). Leucine L472 remains essentially in the crystal conformation, with a rotation of about 120° in iLOV when compared to the single mutant (see Figure 6, top right panel). Also the side chain of residue 470 (Leu in iLOV, Phe in C426A) reproduces the crystal structure geometry for both systems (see Figure S10); however, L470 is somewhat more flexible than F470.

In summary, our MD simulations confirm that the F470L mutation stabilizes the flipping of the L472 residue in iLOV, because L470 has a smaller van der Waals volume compared to F470. This allows the L472 side chain to fit better in the
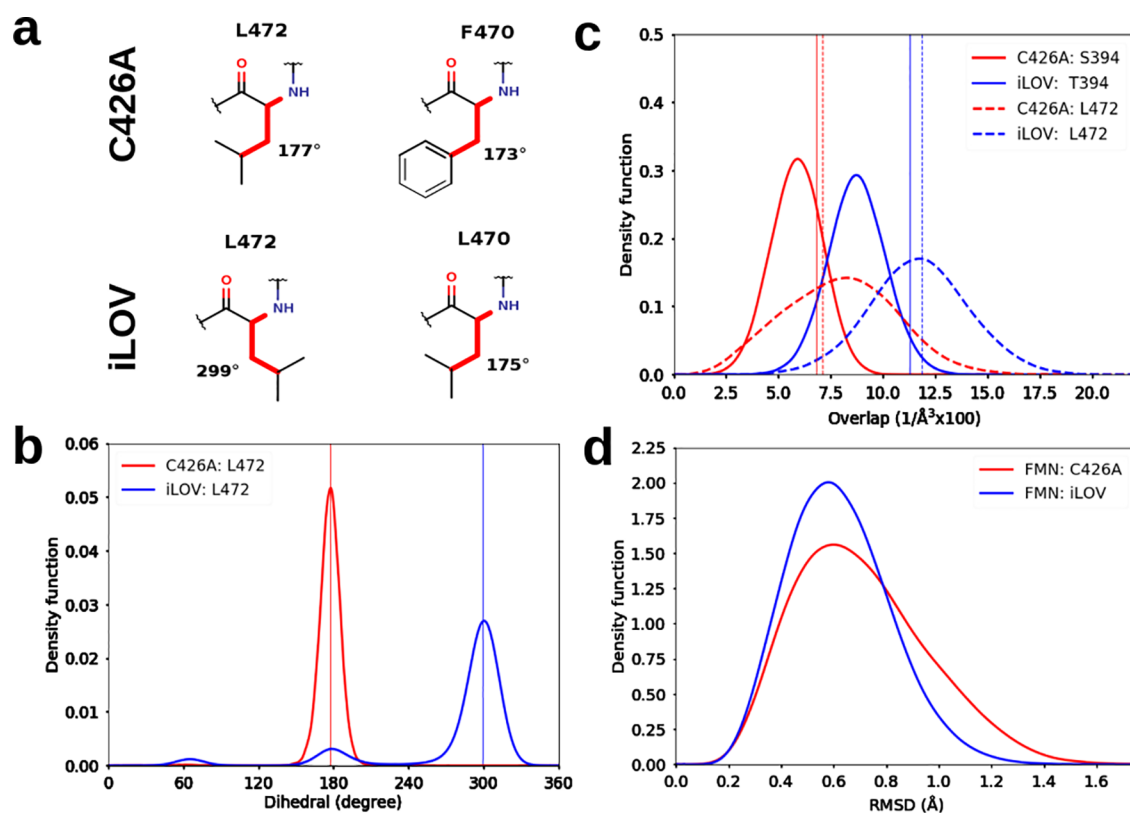
Figure 6. (a) Dihedral angles (red color) analyzed for C426A (top) and iLOV (bottom). The values in degree refer to the crystal structures. (b) L472 dihedral distributions for C426A (red) and iLOV (blue). The vertical lines indicate the values found in the crystal structures. The dihedral distributions for F470 (in C426A) and L470 (in iLOV) are available in Figure S10. (c) Distributions of the overlap between the FMN and the indicated residues for both systems. The vertical lines indicate the overlap values obtained from the crystal structures. (d) RMSD distributions for FMN isoalloxazine ring along DM simulations of C426A (red) and iLOV (blue). The backbone of the protein crystal structures was used as reference for the alignment (the N- and C-terminal residues were excluded).

Table 1. Calculated Vertical Emission (EMI) and Excitation (EXC) Energies and Stokes Shifts (SS) for the Crystal Structure and the Most Populated Clusters of C426A and iLOV[a]

	C426A				
	CRY	$C2_{OUT}$	$C5_{OUT}$	$C7_{IN}$	avg
pop. (%)		5	17	78	
EXC	3.17	3.27	3.23	3.22	3.22
EMI	2.78	2.88	2.84	2.82	2.83
SS	3154	3203	3145	3192	3185

	iLOV							
	CRY	$C0_{OUT}$	$C1_{IN}$	$C2_{OUT}$	$C3_{OUT}$	$C6_{OUT}$	$C7_{OUT}$	avg
pop. (%)		11	15	35	7	28	3	
EXC	3.15	3.25	3.21	3.25	3.25	3.25	3.26	3.24
EMI	2.77	2.87	2.82	2.87	2.87	2.86	2.87	2.86
SS	3062	3059	3098	3090	3079	3089	3090	3087

[a]The weighted average (avg) values are also indicated. For each cluster we also indicate its IN or OUT conformation and the corresponding population (%). EMI and EXC energies are expressed in eV and the Stokes shifts (SS) in cm^{-1}.

flipped position, closer to the isoalloxazine ring, making this orientation stable also in the presence of thermal fluctuations.

To investigate the crowding effect, instead, we calculated the van der Waals overlap (eq 1) between the FMN and the apolar residues in the positions 394 and 472 of the two systems (namely S394 and L472 for C426A and T494 and L472 for iLOV). The resulting distributions along the MD trajectories are reported in Figure 6c. From these graphs it clearly appears that the FMN in iLOV is significantly more overlapped with both residues than in the other variant. The crowding of the

active site in iLOV is induced cooperatively by S394T and F470L mutations, since both side chains of T394 and flipped L472 are closer to the chromophore, thus contributing to a better residue packing around the FMN. Such optimized packing could increase the stiffness of the chromophore in iLOV, as also indicated by the lower RMSD for the isoalloxazine ring obtained from the corresponding trajectories (Figure 6d). All these indications suggest that the presence of T349 and L472 contributes to an improved protein−chromophore interaction which reduces the mobility of the

chromophore in the active site of iLOV, exactly as proposed in the crowding mechanism.

3.2. Spectroscopic Analysis. The clusters obtained from the analysis of the H-bonding network were finally used in combination with multiscale QM/MMPol calculations, to evaluate the impact of H-bonding dynamics on absorption spectra and Stokes shifts. From each cluster, 10 configurations were extracted and used to compute the chromophore vertical excitation and emission energies (see the Computational Details section and Figure 2).

The calculated vertical energies are reported in Table 1 for each cluster of both systems. In addition, for each system, we also report the population-weighted average values. All clusters give similar results, but some configurations have a more blue-shifted excitation/emission energy. The most blue-shifted cluster presents $Q489_{out}$ and $N468_{in}$ side chain conformations, with $Q489_{out}$ in its solvent-exposed conformation and no water molecules bound to the chromophore (cluster C2 in C426A).

By comparing the different clusters, we can estimate that the loss of Q489-FMN H-bond causes a blue-shift of about 400 cm^{-1} in the excitation energy. A parallel quantification for the N468-FMN H-bond was not possible because, as mentioned before, in the frames in which we observe the occurrence of $N468_{out}$, the N468 side chain is always replaced by water molecules.

The absorption lineshapes were computed separately for each cluster as detailed in section 2.3. The resulting spectra are reported in Figure 7, along with the population-weighted average and with the experiments. Owing to the large population of in conformations, the calculated average C426A spectrum is essentially the same as the one of cluster C7. On the contrary, in iLOV the average spectrum is determined by the *out* clusters, which are all very similar, and account for ~85% of the trajectory.

The calculated absorption lineshapes compare well with experiments, reproducing the two main vibronic bands at ~21 000 and ~23 000 cm^{-1}, as well as the shoulder at higher energies. We can thus conclude that the broadening of the absorption band is mainly vibronic in nature, and that the modulation of pigment−protein interactions has a relatively small effect on the spectrum. While the experimental absorption spectra are very similar for the two proteins, our calculated spectra differ slightly by a small blue-shift for iLOV. We ascribe this difference to the different cluster populations observed in the two sets of MDs. Nonetheless, the spectra of different clusters are shifted by ~300 cm^{-1}, that is, less than the vibronic broadening, suggesting that the IN and OUT conformations could be populated in both C426A and iLOV, without influencing the absorption line shape. Our results suggest that the absorption spectrum of C426A and iLOV is likely the result of several pocket conformations, comprising both *in* and *out* conformations for the glutamine and asparagine side chains.

Regarding the Stokes shift (SS) values (Table 1), no significant differences were observed among the clusters of each system. However, the SS obtained in the C426A clusters is systematically 80−100 cm^{-1} greater than that obtained for iLOV clusters. The Stokes shifts computed on the crystal structures present the same trend for the two systems, with values comparable to the MD ones. From these data, it is clear that the Stokes shift does not depend on the dynamic H-bonding network, but it is apparently controlled by the mutations S349T and F470L discussed above. By hampering
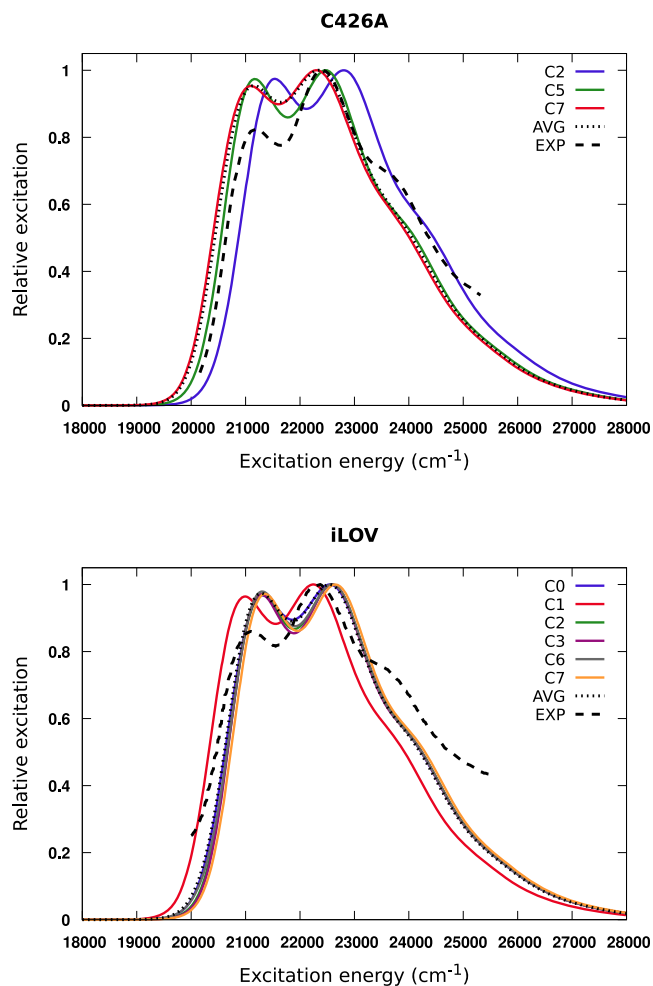

Figure 7. Absorption spectra for C426A (top) and iLOV (bottom) computed from MD using 10 frames for each cluster. The figures show the average spectra for each cluster (solid lines), the weighted average spectra (AVG) for all clusters (dotted line), as well as the experimental absorption spectra (EXP) obtained by Chapman and co-workers[11] (dashed line). All spectra intensities were normalized so that the maximum is 1. All computed spectra were shifted by −3500 cm^{-1} to match the maximum absorption wavenumber of C426A.

the mobility of the isoalloxazine ring, the "crowding" effect slightly reduces the extent of intermolecular reorganization experienced by the chromophore after excitation to the S_1 state. We speculate that the same mechanism also contributes to slowing down the internal conversion of FMN toward the ground state. Intriguingly, the measured fluorescence spectra of C426A and iLOV, while very similar,[11] show some non-negligible differences (Figure S11). Indeed, the fluorescence spectrum of iLOV is narrower and slightly blue-shifted, in agreement with the Stokes shifts calculated in this work.

4. CONCLUSIONS

We analyzed the microsecond time-scale dynamics of two LOV-based fluorescent protein variants, namely C426A and iLOV, by means of MD simulations. In particular, we characterized the dynamics of the H-bonding network of conserved residues in the FMN binding pocket of the two systems and found a significant flexibility especially for the glutamine (Q489) and asparagine (N486) side chains which undergo conformational changes in time scales in the order of

200−300 ns. These findings stress the importance of a proper sampling for any study aimed at optimizing LOV-derived fluorescent proteins.[26]

Our simulations confirmed the "flipping and crowding" effect induced in iLOV by the additional mutations and revealed its mechanism of action. In particular, the crowding is cooperatively triggered by S394T and F470L mutations. The former leads to a direct packing effect, while the latter allows the conserved L472 side chain to change conformation and fit closer to the chromophore. On the other hand, we showed that these mutations, and the resulting differences in the composition and flexibility of the binding pockets, are not reflected in significant shifts of the excitation and emission energies, in agreement with the similarity of the spectra measured for the two systems. However, a small but consistent reduction was found in the Stokes shift, moving from C426A to iLOV. This suggests that the reduction induced by F470 and S394T mutations in the mobility of the isoalloxazine ring of FMN in iLOV also reduces the intermolecular reorganization experienced by the chromophore after excitation. Here, we hypothesize that the same mechanism could also contribute to slowing down the internal conversion of FMN toward the ground state and to improve the fluorescence of iLOV.

■ ASSOCIATED CONTENT

ⓈI Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c10834.

> RMSD plots for the different replicas of the systems. Additional analyses of the H-bonding distances in the binding pockets. Representative structures of the less populated clusters of iLOV (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Felipe Cardoso Ramos − *Dipartimento di Chimica e Chimica Industriale, University of Pisa, Pisa I-56124, Italy;*
Email: cfelip2@gmail.com

Benedetta Mennucci − *Dipartimento di Chimica e Chimica Industriale, University of Pisa, Pisa I-56124, Italy;*
ⓘ orcid.org/0000-0002-4394-0129;
Email: benedetta.mennucci@unipi.it

Author

Lorenzo Cupellini − *Dipartimento di Chimica e Chimica Industriale, University of Pisa, Pisa I-56124, Italy;*
ⓘ orcid.org/0000-0003-0848-2908

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpcb.0c10834

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **1994**, *263*, 802−805.

(2) Chia, H. E.; Marsh, E. N. G.; Biteen, J. S. Extending fluorescence microscopy into anaerobic environments. *Curr. Opin. Chem. Biol.* **2019**, *51*, 98−104.

(3) Ozbakir, H. F.; Anderson, N. T.; Fan, K.-C.; Mukherjee, A. Beyond the Green Fluorescent Protein: Biomolecular Reporters for Anaerobic and Deep-Tissue Imaging. *Bioconjugate Chem.* **2020**, *31*, 293−302.

(4) Drepper, T.; Eggert, T.; Circolone, F.; Heck, A.; Krauß, U.; Guterl, J.-K.; Wendorff, M.; Losi, A.; Gärtner, W.; Jaeger, K.-E. Reporter proteins for in vivo fluorescence without oxygen. *Nat. Biotechnol.* **2007**, *25*, 443−445.

(5) Mukherjee, A.; Schroeder, C. M. Flavin-based fluorescent proteins: Emerging paradigms in biological imaging. *Curr. Opin. Biotechnol.* **2015**, *31*, 16−23.

(6) Losi, A.; Viappiani, C. In *Encyclopedia of Biophysics*; Roberts, G., Watts, A., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2020; pp 1−9.

(7) Anderson, N. T.; Weyant, K. B.; Mukherjee, A. Characterization of flavin binding in oxygen-independent fluorescent reporters. *AIChE J.* **2020**, *66*, 4784−9.

(8) Sakai, T.; Kagawa, T.; Kasahara, M.; Swartz, T. E.; Christie, J. M.; Briggs, W. R.; Wada, M.; Okada, K. Arabidopsis nph1 and npl1: blue light receptors that mediate both phototropism and chloroplast relocation. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 6969−6974.

(9) Herrou, J.; Crosson, S. Function, structure and mechanism of bacterial photosensory LOV proteins. *Nat. Rev. Microbiol.* **2011**, *9*, 713−723.

(10) Kopka, B.; Magerl, K.; Savitsky, A.; Davari, M. D.; Röllen, K.; Bocola, M.; Dick, B.; Schwaneberg, U.; Jaeger, K. E.; Krauss, U. Electron transfer pathways in a light, oxygen, voltage (LOV) protein devoid of the photoactive cysteine. *Sci. Rep.* **2017**, *7*, 1−17.

(11) Chapman, S.; Faulkner, C.; Kaiserli, E.; Garcia-Mata, C.; Savenkov, E. I.; Roberts, A. G.; Oparka, K. J.; Christie, J. M. The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 20038−20043.

(12) Christie, J. M.; Hitomi, K.; Arvai, A. S.; Hartfield, K. A.; Mettlen, M.; Pratt, A. J.; Tainer, J. A.; Getzoff, E. D. Structural tuning of the fluorescent protein iLOV for improved photostability. *J. Biol. Chem.* **2012**, *287*, 22295−22304.

(13) Shu, X.; Lev-Ram, V.; Deerinck, T. J.; Qi, Y.; Ramko, E. B.; Davidson, M. W.; Jin, Y.; Ellisman, M. H.; Tsien, R. Y. A genetically encoded tag for correlated light and electron microscopy of intact cells, tissues, and organisms. *PLoS Biol.* **2011**, *9*, e1001041.

(14) Mukherjee, A.; Weyant, K. B.; Walker, J.; Schroeder, C. M. Directed evolution of bright mutants of an oxygen-independent flavin-binding fluorescent protein from Pseudomonas putida. *J. Biol. Eng.* **2012**, *6*, 20.

(15) Wingen, M.; Potzkei, J.; Endres, S.; Casini, G.; Rupprecht, C.; Fahlke, C.; Krauss, U.; Jaeger, K. E.; Drepper, T.; Gensch, T. The photophysics of LOV-based fluorescent proteins-new tools for cell biology. *Photochem. Photobio. Sc.* **2014**, *13*, 875−883.

(16) Mukherjee, A.; Walker, J.; Weyant, K. B.; Schroeder, C. M. Characterization of flavin-based fluorescent proteins: an emerging class of fluorescent reporters. *PloS one* **2013**, *8*, No. 64753, DOI: 10.1371/journal.pone.0064753.

(17) Davari, M. D.; Kopka, B.; Wingen, M.; Bocola, M.; Drepper, T.; Jaeger, K. E.; Schwaneberg, U.; Krauss, U. Photophysics of the LOV-Based Fluorescent Protein Variant iLOV-Q489K Determined by Simulation and Experiment. *J. Phys. Chem. B* **2016**, *120*, 3344−3352.

(18) Losi, A.; Gardner, K. H.; Möglich, A. Blue-light receptors for optogenetics. *Chem. Rev.* **2018**, *118*, 10659−10709.

(19) Shcherbakova, D. M.; Shemetov, A. A.; Kaberniuk, A. A.; Verkhusha, V. V. Natural photoreceptors as a source of fluorescent proteins, biosensors, and optogenetic tools. *Annu. Rev. Biochem.* **2015**, *84*, 519−550.

(20) Kobayashi, H.; Ogawa, M.; Alford, R.; Choyke, P. L.; Urano, Y. New strategies for fluorescent probe design in medical diagnostic imaging. *Chem. Rev.* **2010**, *110*, 2620−2640.

(21) Dittrich, M.; Freddolino, P. L.; Schulten, K. When Light Falls in LOV: A Quantum Mechanical/Molecular Mechanical Study of Photoexcitation in Phot-LOV1 of Chlamydomonas reinhardtii. *J. Phys. Chem. B* **2005**, *109*, 13006−13013.

(22) Salzmann, S.; Silva-Junior, M. R.; Thiel, W.; Marian, C. M. Influence of the LOV Domain on Low-Lying Excited States of Flavin: A Combined Quantum-Mechanics/Molecular-Mechanics Investigation. *J. Phys. Chem. B* **2009**, *113*, 15610−15618.

(23) Khrenova, M. G.; Nemukhin, A. V.; Domratcheva, T. Theoretical characterization of the flavin-based fluorescent protein iLOV and its Q489K mutant. *J. Phys. Chem. B* **2015**, *119*, 5176−5183.

(24) Khrenova, M. G.; Meteleshko, Y. I.; Nemukhin, A. V. Mutants of the Flavoprotein iLOV as Prospective Red-Shifted Fluorescent Markers. *J. Phys. Chem. B* **2017**, *121*, 10018−10025.

(25) Meteleshko, Y. I.; Nemukhin, A. V.; Khrenova, M. G. Novel flavin-based fluorescent proteins with red-shifted emission bands: a computational study. *Photochem. Photobio. Sc.* **2019**, *18*, 177−189.

(26) Nemukhin, A. V.; Grigorenko, B. L.; Khrenova, M. G.; Krylov, A. I. Computational Challenges in Modeling of Representative Bioimaging Proteins: GFP-Like Proteins, Flavoproteins, and Phytochromes. *J. Phys. Chem. B* **2019**, *123*, 6133−6149.

(27) Bondanza, M.; Nottoli, M.; Cupellini, L.; Lipparini, F.; Mennucci, B. Polarizable embedding QM/MM: the future gold standard for complex (bio)systems? *Phys. Chem. Chem. Phys.* **2020**, *22*, 14433−14448.

(28) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R.; Ghoreishi, D. et al.*AMBER 2018*; University of California: San Francisco, CA, 2018 .

(29) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.

(30) Schneider, C.; Sühnel, J. A molecular dynamics simulation of the flavin mononucleotide−RNA aptamer complex. *Biopolymers* **1999**, *50*, 287−302.

(31) Darden, T.; Perera, L.; Li, L.; Pedersen, L. New tricks for modelers from the crystallography toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* **1999**, *7*, 55−60.

(32) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−95.

(33) Campello, R. J.; Moulavi, D.; Sander, J. Density-based clustering based on hierarchical density estimates. *Lecture Notes in Computer Science* **2013**, *7819 LNAI*, 160−172.

(34) McInnes, L.; Healy, J. Accelerated Hierarchical Density Based Clustering. *IEEE International Conference on Data Mining Workshops, ICDMW 2017,* **2017** (Nov), 33−42.

(35) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. The ONIOM method: its foundation and applications to metalloenzymes and photobiology. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 327−350.

(36) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; et al. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678−5796.

(37) Wang, J.; Cieplak, P.; Li, J.; Hou, T.; Luo, R.; Duan, Y. Development of polarizable models for molecular mechanical calculations I: parameterization of atomic polarizability. *J. Phys. Chem. B* **2011**, *115*, 3091−9.

(38) Wang, J.; Cieplak, P.; Li, J.; Wang, J.; Cai, Q.; Hsieh, M.; Lei, H.; Luo, R.; Duan, Y. Development of Polarizable Models for Molecular Mechanical Calculations II: Induced Dipole Models Significantly Improve Accuracy of Intermolecular Interaction Energies. *J. Phys. Chem. B* **2011**, *115*, 3100−3111.

(39) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, revision A.03; Gaussian Inc.: Wallingford CT, 2016 .

(40) Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. Electronic energy transfer in condensed phase studied by a polarizable QM/MM model. *J. Chem. Theory Comput.* **2009**, *5*, 1838−1848.

(41) Lipparini, F. General Linear Scaling Implementation of Polarizable Embedding Schemes. *J. Chem. Theory Comput.* **2019**, *15*, 4312−4317.

(42) Mukamel, S. *Principles of Nonlinear Optical Spectroscopy*; Oxford University Press: New York, 1995.

(43) Nazarenko, V. V.; Remeeva, A.; Yudenko, A.; Kovalev, K.; Dubenko, A.; Goncharov, I. M.; Kuzmichev, P.; Rogachev, A. V.; Buslaev, P.; Borshchevskiy, V.; et al. A thermostable flavin-based fluorescent protein from Chloroflexus aggregans: a framework for ultra-high resolution structural studies. *Photochem. Photobio Sc.* **2019**, *18*, 1793−1805.

(44) Freddolino, P. L.; Dittrich, M.; Schulten, K. Dynamic switching mechanisms in LOV1 and LOV2 domains of plant phototropins. *Biophys. J.* **2006**, *91*, 3630−3639.

(45) Lokhandwala, J.; Silverman y de la Vega, R. I.; Hopkins, H. C.; Britton, C. W.; Rodriguez-Iglesias, A.; Bogomolni, R.; Schmoll, M.; Zoltowski, B. D. A native threonine coordinates ordered water to tune Light-Oxygen-Voltage (LOV) domain photocycle kinetics and osmotic stress signaling in Trichoderma reesei ENVOY. *J. Biol. Chem.* **2016**, *291*, 14839−14850.

(46) Yee, E. F.; Diensthuber, R. P.; Vaidya, A. T.; Borbat, P. P.; Engelhard, C.; Freed, J. H.; Bittl, R.; Möglich, A.; Crane, B. R. Signal transduction in light−oxygen−voltage receptors lacking the adduct-forming cysteine residue. *Nat. Commun.* **2015**, *6*, 1−10.