Supporting Information

Role of Active Site Conformational Changes in Photocycle Activation of the AppA BLUF Photoreceptor

Puja Goyal[†] and Sharon Hammes-Schiffer[†]*

[†]*Department of Chemistry, 600 South Mathews Avenue,*

University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

*corresponding author
E-mail: shs3@illinois.edu
Phone: (217) 300-0335

Table of Contents

A. Computational methodology

A.1. Molecular mechanical (MM) studies of Trp104, Met106 and Gln63 movement

A.1.1. System preparation and equilibration. Starting from the BLUF domain crystal structure (PDB ID 1YRX, chain A) (1), hydrogen atoms were added using HBUILD in CHARMM (2) assuming standard protonation states for all titratable residues at pH 7. The protein was solvated in a periodically replicated cubic box of water molecules with edge length 85.4 Å. To neutralize the net charge of the protein, Na^+ ions were added, followed by addition of equal numbers of Na^+ and Cl^- ions to achieve an ionic concentration of 0.150 M NaCl. The CHARMM36 force field (3) for amino acids and ions and the modified TIP3P model (4) for water were used. Force field parameters for the flavin mononucleotide (FMN) cofactor were obtained from a previous study by Schulten and coworkers (5). Electrostatic interactions were treated using the Particle Mesh Ewald (PME) method (6) with a grid size of ~0.9 Å, and van der Waals interactions were truncated at atom-atom distances of 14.5 Å using the SHIFT scheme. For computational efficiency of the molecular dynamics simulations, the domain decomposition (DOMDEC) module (7) in CHARMM was used. All bonds involving hydrogen were constrained using the SHAKE algorithm (8). A molecular dynamics time step of 1 fs was used.

The system was equilibrated on the ground state, as well as the locally excited (LE) and charge transfer (CT) excited states, according to the following procedure. Keeping the protein frozen, 500 steps of steepest descent minimization followed by adopted basis Newton Raphson (ABNR) minimization with a gradient tolerance of 0.01 kcal mol^{-1} $Å^{-1}$ was performed. Subsequently, 250 ps heating to 300 K using velocity scaling was conducted, followed by 250 ps equilibration at 300 K using velocity scaling and a 750 ps NVE simulation for the solvent and ions. Subsequently, keeping the solvent and ions frozen, geometry optimization for the protein was performed using the same protocol described above. Restraining the protein heavy atoms to their optimized positions with a mass-weighted harmonic restraint with force constant 1.0 kcal mol^{-1} $Å^{-2}$ first and then 0.5 kcal mol^{-1} $Å^{-2}$, two 250 ps equilibration trajectories at 300 K were propagated. Subsequently, an unrestrained 500 ps equilibration NVT trajectory at 300 K was propagated. Finally, a 20 ns unrestrained trajectory in the NPT ensemble was propagated on the ground state, using the extended system constant pressure and temperature algorithm implemented in CHARMM based on the work of Andersen (9), Nosé, Klein (10), and Hoover (11). The mass of the pressure piston, Langevin piston collision frequency, reference pressure,

mass of the thermal piston, and reference temperature were set to 500 a.m.u., 10 ps^{-1}, 1 atm., 2000 kcal ps^2 mol^{-1} and 300 K, respectively. Within ~5 ps, the box size stabilized at ~83 Å.

Starting from the structure at the end of the 20 ns NPT trajectory on the ground state, geometry optimization for the protein and FMN chromophore was carried out with the FMN and Tyr21 partial charges corresponding to the LE and CT states. The partial charges of FMN and Tyr21 for the excited electronic states were obtained from time-dependent density functional theory (TDDFT) (12) calculations using Gaussian 09 (13), as described in Section C. This geometry optimization was followed by equilibration as described above, and then by a 20 ns NVT production simulation on each excited state.

A.1.2. Trp104 and Met106 movement. To calculate the minimum free energy path and the associated free energy profile for interconversion between the Trp_{in}/Met_{out} and Trp_{out}/Met_{in} conformations (Figs. 3 and 4 in the main paper), the adaptively biased path optimization (ABPO) method (14) implemented in CHARMM was used. ABPO combines the path optimization algorithm of the finite temperature string (FTS) method (15) with the sampling power of an adaptive biasing potential method. Unlike the FTS method, ABPO allows trajectories to diffuse along the path and accelerates this diffusion using an adaptive biasing potential. Sampling is further aided by running independent trajectories for several images located at different points along the path.

The initial linear path between the Trp_{in}/Met_{out} and Trp_{out}/Met_{in} conformations was constructed according to the following procedure. Starting from the Trp_{in}/Met_{out} structure at the end of the 20 ns NPT trajectory on the ground state, Trp104 and Met106 were pulled away from and brought closer to FMN, respectively, using a series of 1 ns restrained NVT trajectories on the ground state. The restraints were harmonic, with a force constant of 600 kcal mol^{-1} $Å^{-2}$ on both the O4(FMN)-NE1(Trp104) and O4(FMN)-SD(Met106) distances (Fig. 3*B* in the main paper), where NE1 and SD are the side-chain N and S atoms of Trp104 and Met106, respectively. The ranges of O4(FMN)-NE1(Trp104) and O4(FMN)-SD(Met106) distances at which the restraints were centered were 3.4–15.4 Å with an interval of 0.6 Å and 3.66–10.33 Å with an interval of ~0.33 Å, respectively.

Prior to path optimization using the ABPO method, the pseudo-diffusion tensor D was evaluated using a 200 ps trajectory for each of the 21 images, with the instantaneous value of D calculated every 50 fs. This procedure was followed by 16 cycles of path optimization with a 15

ps trajectory per image and 10 cycles of path optimization with a 50 ps trajectory per image. The adaptive bias strength of 0.9 corresponded to well-tempered metadynamics and allowed efficient sampling along the path. The path was discretized into 200 points. The radius of the path tube was 1.0 Å $g^{0.5}$, and the force constant of the tube wall potential was 3.0 kcal g^{-1} $Å^{-2}$. The ratio of the standard deviation of the Gaussian mollifiers used to smooth the histogram and the path to the total path length was 0.05. In these calculations, 500 time steps were pre-deposited into the histogram, and the path was updated whenever the histogram reached the sampling threshold of 50 time steps everywhere along the path.

A.1.3. Gln63 rotation. From the set of structures used to construct the initial linear path for ABPO calculations, the structure with the O4(FMN)-NE1(Trp104) and O4(FMN)-SD(Met106) distances close to 13.0 Å and 5.0 Å, respectively, was used to initiate an unrestrained 20 ns NVT simulation on the ground state and obtain an equilibrated Trp_{out}/Met_{in} structure on the ground state. From this structure, equilibrated Trp_{out}/Met_{in} structures on the LE and CT states were obtained following the same procedure as described in Section A.1.1.

Starting from equilibrated Trp_{in}/Met_{out} and Trp_{out}/Met_{in} structures on each of the three electronic states of interest, umbrella sampling (16) simulations were carried out along the $r_{CN}-r_{CO}$ coordinate, with r_{CN} and r_{CO} denoting the C6(FMN)-NE2(Gln63) and C6(FMN)-OE1(Gln63) distances, respectively, where NE2(Gln63) and OE1(Gln63) represent the side-chain N and O atoms of Gln63. For these simulations, 10−11 windows with a spacing of 0.4 Å were employed. Initial structures for the windows were obtained from 100 ps trajectories starting from a structure from an adjacent window using a restraint force constant of 500 kcal mol^{-1} $Å^{-2}$. Subsequently, a 100 ps equilibration and 500 ps production trajectory with a restraint force constant of 50 kcal mol^{-1} $Å^{-2}$ was propagated for each window. The weighted histogram analysis method (WHAM) (17) was used to obtain the final potential of mean force (PMF) along the chosen reaction coordinate.

A.2. Quantum mechanical/molecular mechanical (QM/MM) studies of Gln63 tautomerization.

A.2.1. System preparation and equilibration. To calculate the free energy of Gln63 tautomerization, we carried out QM/MM (18) umbrella sampling simulations in which Gln63 was treated quantum mechanically using the density functional tight-binding (DFTB) method (19). For computational efficiency, the generalized solvent boundary potential (GSBP) method

(20) was used in place of periodic boundary conditions for the treatment of electrostatics. Starting from equilibrated Trp_{in}/Met_{out} and Trp_{out}/Met_{in} structures on each of the three electronic states of interest, all water molecules except those with their O atom within 26 Å of the N5 atom of FMN were deleted. All ions except 3 Na^+ ions were also deleted, maintaining the net neutrality of the system. The system was divided into an inner region, which contains all atoms within 28 Å of the N5 atom of FMN, and an outer region, which includes all other atoms. Solvent in the outer region was represented by a dielectric continuum with $\varepsilon = 80$. The protein atoms in the outer region and those in the inner region close to the inner-outer boundary (i.e., CHARMM groups with any inner region atom directly bonded to an outer region atom, bonded to an inner region atom directly bonded to an outer region atom, or between a distance of 26 Å and 28 Å from the N5 atom of FMN) were frozen during the simulations.

The solvent-shielded static field due to the outer region atoms and the reaction field contribution to interactions within the inner region were determined using Poisson-Boltzmann (PB) calculations in the framework of the GSBP method. In the PB calculations, a focusing scheme that places a 64.8 Å cube of fine grid (0.4 Å spacing) into a larger 194.4 Å cube of coarse grid (1.2 Å spacing) was used. The inner region charge density was expressed using the first 30 spherical harmonics with a total of 900 basis functions. The optimized radii of Roux and Nina were adopted to define the solvent-solute dielectric boundary (21). The salt concentration was set to be 150 mM.

In accordance with the stochastic boundary MD protocol (22), Newton's equations of motion were solved for the MD region (within 24 Å from the N5 atom of FMN), and Langevin equations of motion with a temperature bath of 300 K were solved for the buffer region (24-28 Å from the N5 atom of FMN). All protein atoms in the buffer region were harmonically restrained with force constants determined directly from the B-factors in the PDB file. Langevin atoms were updated every 20 steps during the simulation to correctly treat protein groups and water molecules that switch regions during the simulation. All water molecules, ions, and fully mobile protein atoms in the MD region were subjected to a weak GEO (geometrical) type restraining potential to keep them inside the inner sphere with the MMFP module of CHARMM. The GEO restraining potential is in the form of a quartic polynomial on each heavy atom: $k\Delta^2(\Delta^2 - V_p)$, where $\Delta = r - r_{off}$, k is the restraining quartic force constant (0.5 kcal mol^{-1} $Å^{-4}$), r is the distance of the heavy atom from the N5 atom of FMN, r_{off} is the cutoff distance (24 Å) below which the

GEO restraint is set to zero, and V_P is an offset value taken to be 2.25 Å^2. These parameters lead to a restraining potential on the heavy atoms that smoothly turns on at 24 Å, reaches a well at 25 Å with a depth of ~0.6 kcal/mol, and then quickly rises to be repulsive beyond 25.5 Å. Nonbonded interactions within the inner sphere are treated with an extended electrostatics model, in which groups beyond 12 Å from each other interact as multipoles and up to quadrupoles.

Equilibration of the Trp_{in}/Met_{out} and Trp_{out}/Met_{in} structures on each of the three relevant electronic states was carried out as follows. For the GSBP setups with the Trp_{in}/Met_{out} and Trp_{out}/Met_{in} structures equilibrated with classical MD as described in Section A.1 for each of the three electronic states of interest, ABNR geometry optimization with a gradient tolerance of 0.01 kcal mol^{-1} Å^{-2} was performed with all bonds involving hydrogen constrained using SHAKE. This procedure was followed by 50 ps heating to 300 K and 50 ps equilibration at 300 K using the MM force field with a 1 fs time step. A hydrogen link atom between the Cα and Cβ atoms of Gln63 was added to the final structure in each case. The DIV link atom scheme (23), which distributes the charge on the Cα atom to the other atoms in the same CHARMM group, was used to treat the QM/MM boundary. The QM region was described by third-order DFTB (DFTB3) with the 3OB parameter set and calculated Hubbard derivatives (19). Finally, 50 ps QM/MM equilibration was performed with a mass-weighted harmonic restraint force constant of 0.5 kcal mol^{-1} Å^{-2} on the FMN heavy atoms about their initial positions, no SHAKE constraints on the QM region atoms, and a 0.5 fs time step.

A.2.2. Gln63 tautomerization. To obtain the free energy profile for glutamine tautomerization, umbrella sampling simulations with $r_{HN}-r_{HO}$ as the reaction coordinate were conducted, where H, N, and O represent the HE21, NE2, and OE1 side-chain atoms of Gln63. The reaction coordinate was sampled between −1.5 Å and 1.5 Å. Initial structures for the windows were obtained from 15 ps trajectories starting from a structure from an adjacent window using a restraint force constant of 400 kcal mol^{-1} Å^{-2}. Subsequently, with a restraint force constant of 200 kcal mol^{-1} Å^{-2} along the reaction coordinate, a 50 ps equilibration trajectory with a harmonic restraint force constant of 0.5 kcal mol^{-1} Å^{-2} on the FMN heavy atoms followed by a 50 ps equilibration trajectory and a 50 ps production trajectory with a harmonic restraint force constant of 0.15 kcal mol^{-1} Å^{-2} on the FMN heavy atoms was conducted for each window. WHAM was used to obtain the final PMF along the chosen reaction coordinate.

Next we calculated the free energy profile for conversion of the "E-form" of the glutamine tautomer obtained in the calculations above to its "Z-form". The final structure from the window centered at the reaction coordinate value of 1.5 Å was used to initiate a 50 ps QM/MM equilibration trajectory with no restraint along the $r_{HN}-r_{HO}$ coordinate. Subsequently, umbrella sampling simulations along the dihedral angle $\phi_{OCNH'}$ were performed. Here O, C, N, and H′ denote the OE1, CD, NE2, and HE22 side-chain atoms of Gln63. Initial structures for the windows were obtained from 15 ps trajectories starting from a structure from an adjacent window using a restraint force constant of 50 kcal mol^{-1} rad^{-2}. Subsequently, a 50 ps equilibration trajectory and 100 ps production trajectory with a restraint force constant of 50 kcal mol^{-1} rad^{-2} were propagated for each window. All of these simulations were performed with a weak mass-weighted harmonic restraint force constant of 0.15 kcal mol^{-1} $Å^{-2}$ on the FMN heavy atoms about their initial positions.

B. Convergence of ABPO calculations

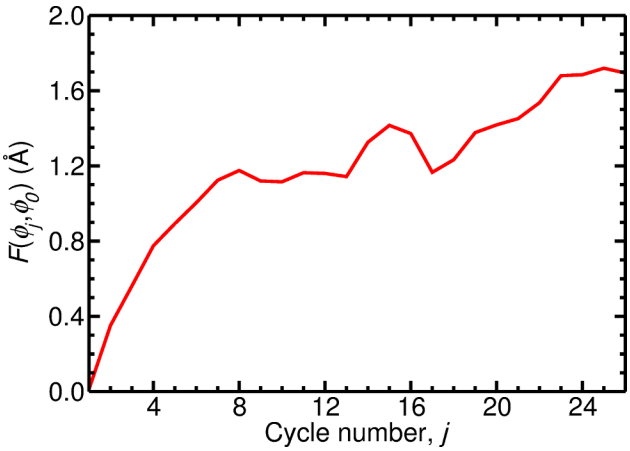

Fig. S1. Evolution of the Fréchet distance between the minimum free energy path (MFEP) from the ABPO calculations and the initial linear path between the Trp_{in}/Met_{out} and Trp_{out}/Met_{in} conformations with the cycle number in the ABPO calculations.

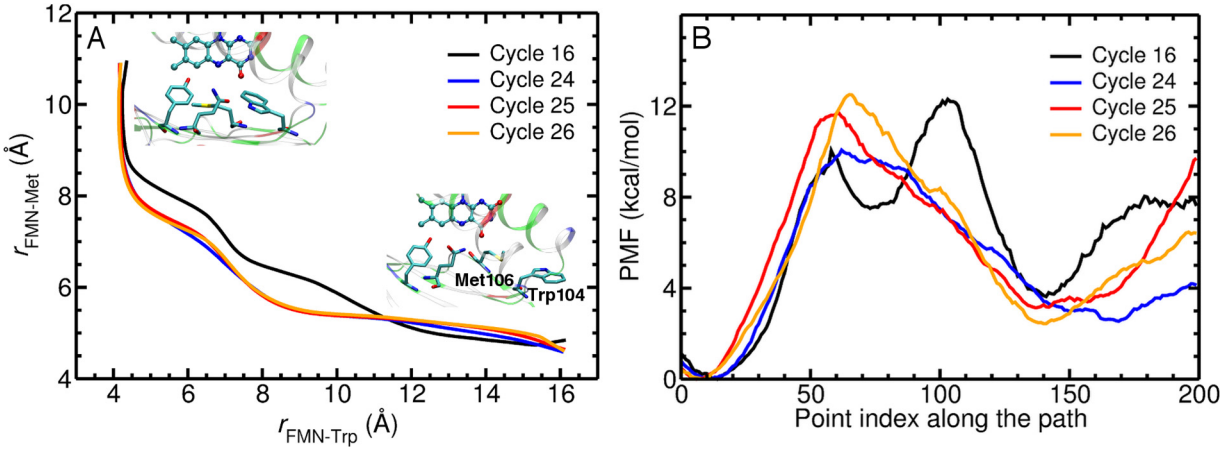

Fig. S2. Depiction of (*A*) the MFEP and (*B*) the PMF associated with the interconversion between the Trp_{in}/Met_{out} and Trp_{out}/Met_{in} conformations in cycles 16, 24, 25, and 26 of the ABPO calculations. The MFEP is calculated as a function of the O4(FMN)-NE1(Trp104) and O4(FMN)-SD(Met106) distances, denoted as $r_{FMN-Trp}$ and $r_{FMN-Met}$, respectively. Each PMF in (*B*) is calculated along the corresponding MFEP shown in (*A*). These ABPO simulations appear to be converged in terms of the MFEP and are converged to within ~1 kcal/mol for the reaction free energy and to within ~3 kcal/mol for the free energy barrier. As for all simulations of this type, however, full convergence cannot be proven.

C. Charge Parameters for the LE and CT States

The coordinates of FMN and Tyr21 were extracted from chain A of two crystal structures of the AppA BLUF domain (PDB ID 1YRX and 2IYG) (1, 24). Hydrogen atoms were added such that the Tyr OH group pointed away from or toward the C=O groups of FMN for the model based on 1YRX or 2IYG, respectively. FMN and Tyr were modified to lumiflavin, representing the isoalloxazine part of the FMN chromophore, and *p*-cresol, respectively. After gas phase geometry optimization in the presence of constraints to avoid large structural changes (Fig. S3 *A* and *B*), gas phase TDDFT calculations with the CAM-B3LYP functional (25) and the 6-31+G** basis set were performed. Atomic charges fit to the electrostatic potential at points selected according to the Merz-Singh-Kollman scheme and constrained to reproduce the dipole moment (26, 27) were obtained on the ground state (GS), locally excited (LE) state, and charge transfer (CT) state (Fig. S3 *C* and *D*). The changes in atomic charges upon photoexcitation, averaged over the two models, were used to modify the ground state force field partial charges of FMN and Tyr21 from Ref. (5) and the CHARMM force field, respectively, to obtain partial charges corresponding to the LE and CT states (Table S1). A similar approach was adopted by previous studies in the literature to obtain partial charges for atoms on excited electronic states (28-30). The force field parameters other than the partial charges on the flavin and Tyr21 remained the same for the excited electronic states. The changes in bonding parameters, as well as the effects on the partial charges of nearby amino acids, were neglected because the direct electrostatic effects are expected to dominate the active site conformations in this study.

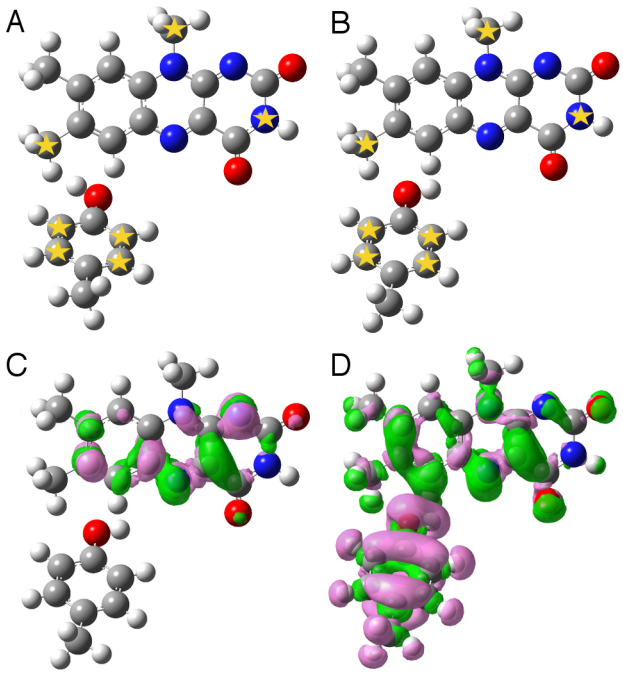

Fig. S3. DFT/CAM-B3LYP/6-31+G**-optimized structures of gas phase models comprised of lumiflavin and *p*-cresol with initial coordinates obtained from FMN and Tyr21 in the AppA BLUF domain crystal structures (*A*) 1YRX and (*B*) 2IYG. During the optimization, the atoms labeled with stars were kept fixed. For the model shown in (*B*), the isosurfaces of electronic density differences between (*C*) the LE and ground states, and (*D*) the CT and ground states from TDDFT/CAM-B3LYP/6-31+G** calculations are depicted. Pink and green regions represent decrease and increase of electronic density upon vertical photoexcitation and correspond to isovalues of –0.0015 and 0.0015 $Bohr^{-3}$, respectively.

Table S1. Partial charges of the FMN and Tyr21 atoms on the GS, LE state and CT state. Only the atoms with charges on the LE or CT state different from those on the GS are listed. The GS charges for FMN and Tyr21 are from Ref. (5) and the CHARMM force field, respectively, and the LE and CT state charges were obtained from atomic charge differences between these states and the GS as obtained from TDDFT calculations.

	Partial charges		
	GS	LE state	CT state
FMN			
C9	-0.102850	-0.123917	-0.105019
C9A	-0.014855	0.060435	-0.094487
C5A	0.298220	0.378787	0.404998
C6	-0.102850	-0.202444	-0.221345
C7	-0.070741	0.053219	-0.001243
C8	-0.070741	-0.132667	-0.205117
N10	0.216248	0.173885	0.275876
C10	0.294628	0.366747	0.224758
C4A	0.293794	0.086130	0.175887
N5	-0.543286	-0.591230	-0.710738
N1	-0.601187	-0.532146	-0.647610
C2	0.543059	0.579735	0.582087
N3	-0.296627	-0.344982	-0.328360
C4	0.446255	0.488996	0.475085
H9	0.087969	0.096621	0.094073
H6	0.087969	0.095660	0.078850
C7M	-0.170000	-0.201520	-0.219457
H71	0.090000	0.101805	0.070723
H72	0.090000	0.103722	0.061139
H73	0.090000	0.103032	0.109013
C8M	-0.170000	-0.146699	-0.080039
H81	0.090000	0.089131	0.063209
H82	0.090000	0.078543	0.036458
H83	0.090000	0.078460	0.057016
C1'	-0.180000	-0.115882	-0.143689
H1'	0.090000	0.065626	0.038877
H1"	0.090000	0.065626	0.038877
O2	-0.474192	-0.469239	-0.559939
H3	0.271484	0.280120	0.255413
O4	-0.462297	-0.485553	-0.558296
Tyr21			
CZ	0.110	0.110	0.300
CE1	-0.115	-0.115	-0.061
CD1	-0.115	-0.115	-0.079
CG	0.000	0.000	0.116
CD2	-0.115	-0.115	-0.044
CE2	-0.115	-0.115	-0.109
OH	-0.540	-0.540	-0.419
HH	0.430	0.430	0.442
HE1	0.115	0.115	0.162
HD1	0.115	0.115	0.141
HD2	0.115	0.115	0.134
HE2	0.115	0.115	0.164
CB	-0.180	-0.180	-0.266
HB1	0.090	0.090	0.176
HB2	0.090	0.090	0.176

D. Gln63 rotation in unrestrained classical MD

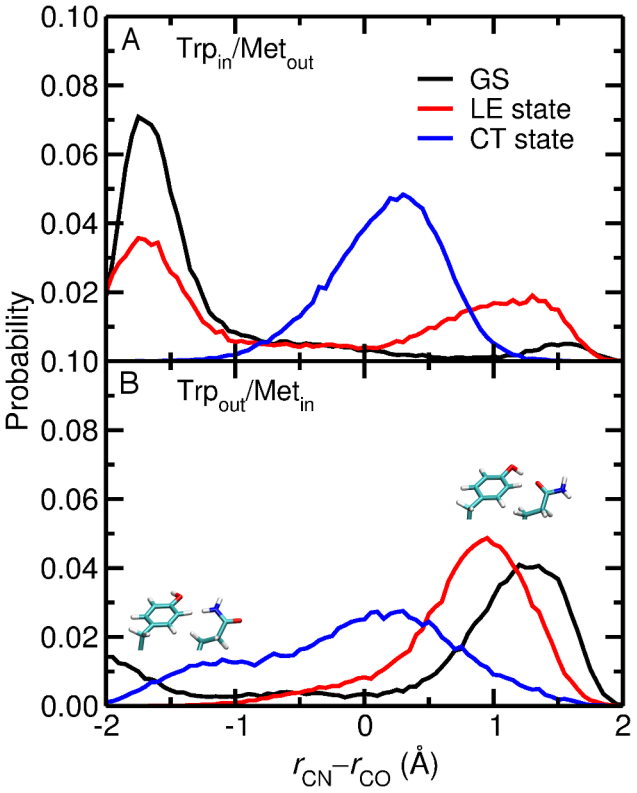

Fig. S4. Probability distribution of r_{CN}–r_{CO}, where C, N, and O denote the C6 atom of FMN, the side-chain NE2 atom of Gln63, and the side-chain OE1 atom of Gln63, respectively, in 20 ns unrestrained classical MD trajectories of the AppA BLUF domain on different electronic states in the (*A*) Trp$_{in}$/Met$_{out}$ and (*B*) Trp$_{out}$/Met$_{in}$ conformations. These distributions reflect the conformation of Gln63 in the active site (compare to Fig. 6 in the main paper). Note that these probabilities are qualitatively consistent with the free energy profiles calculated with umbrella sampling (Fig. 6 in main paper), with slight differences observed for the Trp$_{in}$/Met$_{out}$ conformation on the LE state, presumably due to sampling limitations.

E. Free energy of Gln63 tautomerization

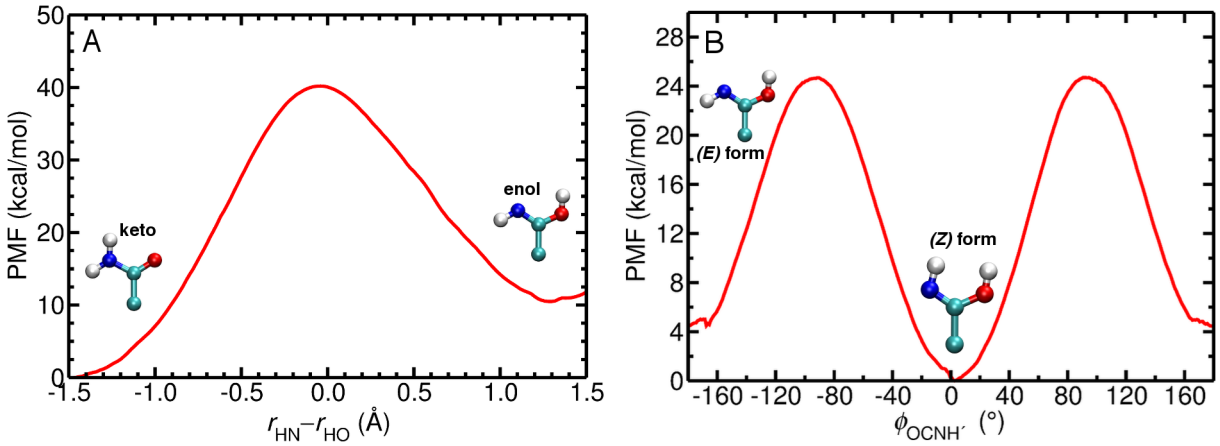

Fig. S5. Free energy profiles for (*A*) keto-enol tautomerization of Gln63 and (*B*) interconversion between the (*E*) and (*Z*) forms of the enol tautomer of Gln63 on the LE state in the Trp_{in}/Met_{out} conformation.

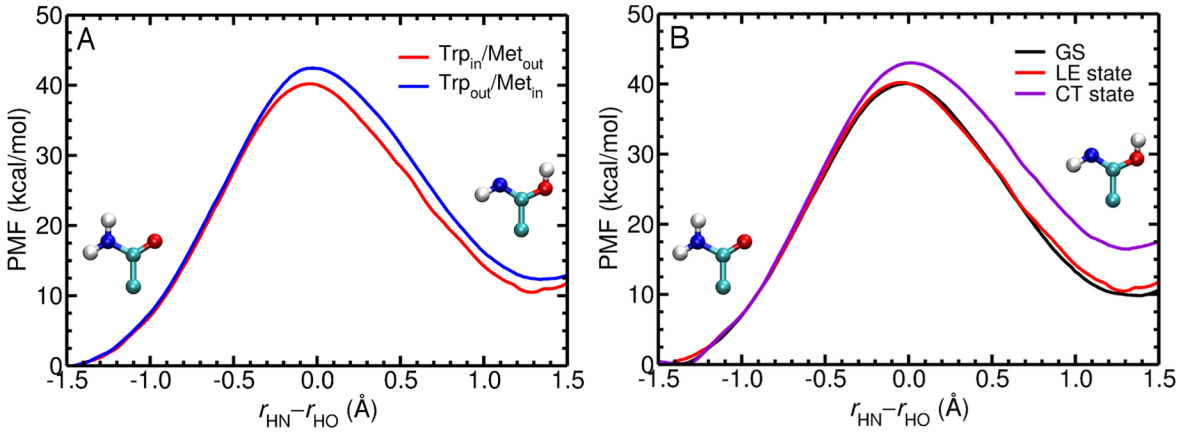

Fig. S6. Free energy profile for keto-enol tautomerization of Gln63 (*A*) on the LE state in the Trp_{in}/Met_{out} and Trp_{out}/Met_{in} conformations and (*B*) on the GS, LE state, and CT state in the Trp_{in}/Met_{out} conformation.

F. Energy gap between the LE and CT states

Starting from the AppA BLUF domain crystal structure (PDB ID 1YRX, chain A), hydrogen atoms were added using HBUILD in CHARMM assuming standard protonation states for all titratable residues at pH 7. The coordinates of residues Tyr21, His44, Asn45, Phe61, Gln63, His78, His85, Trp104, and the chromophore FMN were extracted, and all amino acid backbone atoms were deleted. The C_β atoms of the Tyr, His, Phe, and Trp residues were used for building methyl group caps. Asn45 and Gln63 were modified to ethanamide, and FMN was modified to lumiflavin. In the gas phase model thus constructed, the positions of the N and O atoms of ethanamide representing Gln63 were interchanged to create a second model. In this model, the H atom of the hydroxyl group of *p*-cresol representing Tyr21 was also repositioned so that *p*-cresol donates a hydrogen bond to ethanamide representing Gln63. The H atom positions in these two models were optimized in the gas phase at the DFT/CAM-B3LYP/6-31+G** level of theory. We refer to the partially optimized models with *p*-cresol as a hydrogen bond acceptor and a hydrogen bond donor as A and A_{inv}, respectively (Fig. S7 *A* and *B*).

Smaller models were built from models A and A_{inv} by retaining the residues representing (1) FMN, Tyr21, Gln63, and Trp104 (models B and B_{inv}), (2) FMN, Tyr21, and Gln63 (models C and C_{inv}), and (3) FMN and Tyr21 (models D and D_{inv}). These models are depicted in Fig. S7 *C*-*H*. TDDFT/CAM-B3LYP/6-31+G** calculations were performed on these gas phase models to obtain the vertical excitation energies for the LE and CT states. The CT state involving ET from Tyr21 to FMN is shown in Fig. 2*B* in the main paper. Two different low-lying CT states involving ET from Trp104 to FMN are shown in Fig. S8. The vertical excitation energy to the CT state involving Tyr21 is higher than that to the CT states involving Trp104 for the models in which the Tyr21 and Gln63 orientations are not conducive to PT from Tyr21 to FMN (Tables S2, S3).

The effect of the protein and aqueous environment on the excited state properties was examined as follows. Geometry optimization of the protein in the equilibrated Trp_{in}/Met_{out} structure with Tyr21 and Gln63 not oriented for PT to FMN was carried out using the MM force field with periodic boundary conditions (PBC). The protein geometry optimization was also performed after manually reorienting the Tyr21 and Gln63 hydroxyl and amide groups, respectively, so that they are well-oriented for PT to FMN. In addition, another geometry optimization of the protein was performed for an equilibrated Trp_{in}/Met_{out} structure with Tyr21

and Gln63 oriented for PT to FMN to determine the impact of the specific protein configuration on the results. Hydrogen link atoms were added between the C_α and C_β atoms of Tyr21, the C_α and C_β atoms of Gln63, and the C1' and C2' atoms of FMN, and the charges on the MM host atoms (C_α of Tyr21, C_α of Gln63, and C2' of FMN) were zeroed out. Other link atom schemes were also tested and found to have very small effects on the results. The Tyr21 and Gln63 side chains and flavin, along with the hydrogen link atoms, were treated quantum mechanically at the TDDFT/CAM-B3LYP/6-31+G** level of theory using Gaussian 09. The water molecules and ions in the primary PBC box of edge length 83.29 Å centered at the N5 atom of FMN and the rest of the protein were allowed to polarize the QM region wavefunction via their MM partial charges. TDDFT calculations were performed to obtain the excited electronic state energies of the QM region in the presence of the aqueous and protein environment as well as in the gas phase.

Table S4 illustrates that inclusion of the protein and aqueous environment does not qualitatively impact the difference between the CT-LE gaps for the two different active site hydrogen-bonding patterns. The most significant effect observed is for the vertical excitation to the CT state in the conformation oriented for PT. The CT-LE energy gap for the conformation with Tyr21/Gln63 oriented for PT is still lower by ~0.6 eV compared to that for the conformation with Tyr21/Gln63 not oriented for PT. Moreover, the vertical excitation energies of the LE and CT states and the CT-LE energy gap for the two conformations in the gas phase are similar to those for the models C and C_{inv} (Tables S2, S4), indicating that the trends in the results are not sensitive to the exact geometry of the QM region, but rather are dominated by the overall hydrogen-bonding pattern. Furthermore, for the active site conformation with Tyr21 and Gln63 oriented for PT, the use of an equilibrated structure with this conformation leads to very similar results as a structure generated from an equilibrated structure with Tyr21 and Gln63 not oriented for PT (Table S4), indicating that the results are not sensitive to the specific protein configuration.

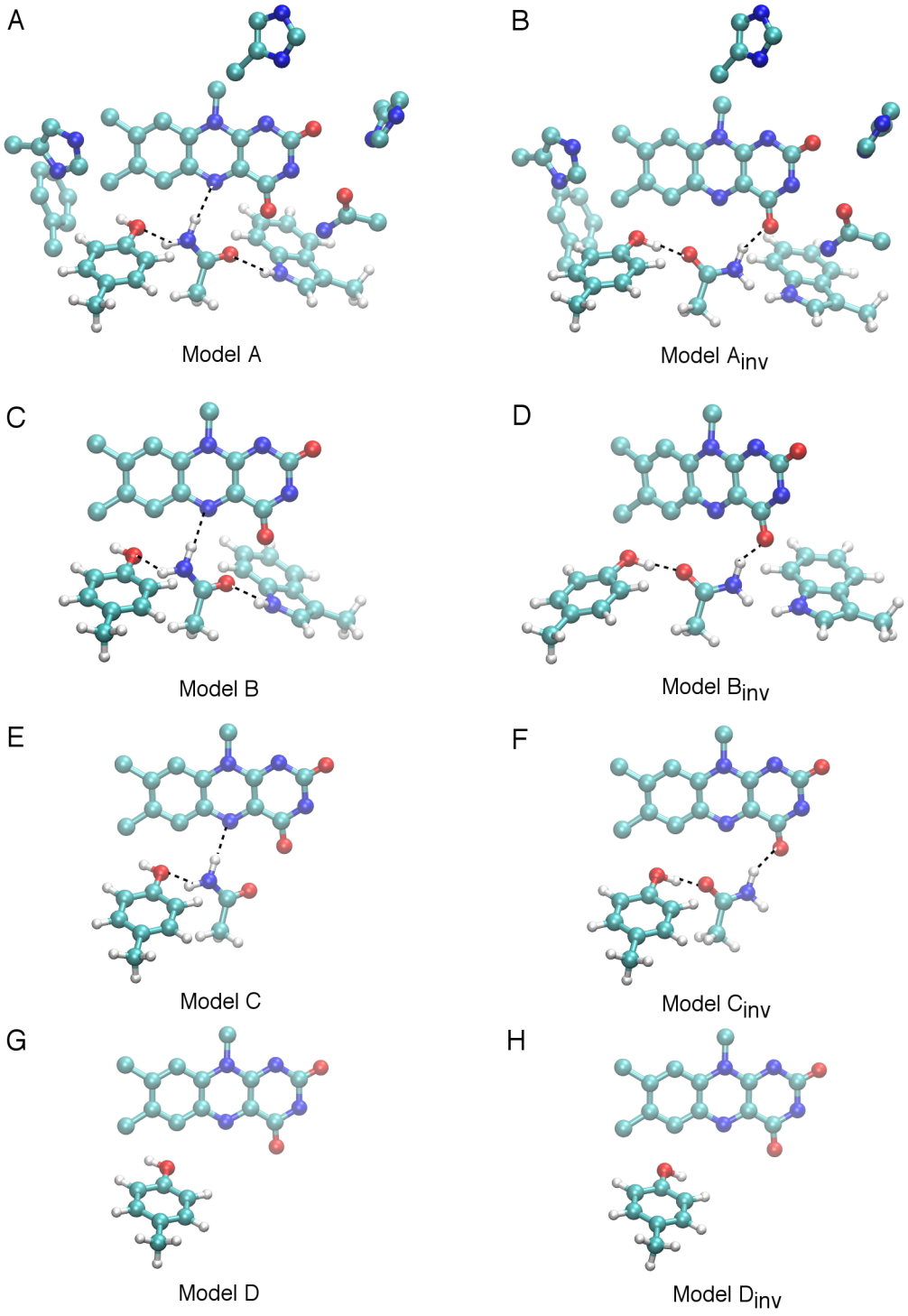

Fig. S7. Geometries of the different gas phase models used to estimate the CT–LE energy gap using TDDFT/CAM-B3LYP/6-31+G**. For simplicity, only the hydrogen atoms of the residues representing Tyr21, Gln63, and Trp104 are shown here.

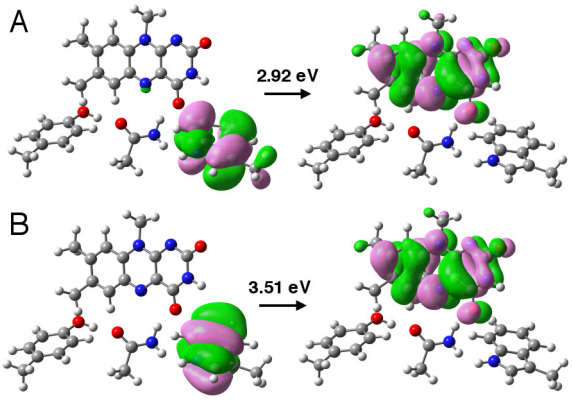

Fig. S8. For model B_{inv}, natural transition orbitals from TDDFT/CAM-B3LYP/6-31+G** calculations depicting the two lowest-energy excitations involving ET from the Trp104 analog to the FMN analog.

Table S2. Vertical excitation energies and oscillator strengths for excitation to the LE state and the CT state, which involves ET from Tyr21 to FMN, in different gas phase models of the active site.

Model	GS→LE[a]		GS→CT[a]		CT−LE gap
	E_{exc} (eV)	f	E_{exc} (eV)	f	ΔE(eV)
A	2.7792	0.2357	4.3917	0.0006	1.6125
A_{inv}	2.7317	0.2082	3.2410	0.0137	0.5093
B	2.8553	0.2230	4.3577	0.0007	1.5024
B_{inv}	2.7996	0.1841	3.1459	0.0066	0.3463
C	2.8798	0.2301	4.3179	0.0010	1.4381
C_{inv}	2.8107	0.1820	3.0491	0.0040	0.2384
D	2.9167	0.2372	3.8556	0.0729	0.9389
D_{inv}	2.8658	0.2296	3.9969	0.0113	1.1311

[a]Fig. 2 in the main paper depicts the electronic density difference plots corresponding to the LE and CT states, respectively, for the C_{inv} model. The "inv" subscript indicates that the conformation is conducive to the proton relay, while the absence of this subscript indicates the alternative conformation.

Table S3. Vertical excitation energies and oscillator strengths for excitation to the LE state and the CT_{Trp} state, which involves ET from Trp104 to FMN, in different gas phase models of the active site.

Model	GS→LE		GS→CT_{Trp}[a]		CT_{Trp}−LE gap
	E_{exc} (eV)	f	E_{exc} (eV)	f	ΔE(eV)
A	2.7792	0.2357	2.8278	0.0007	0.0486
			3.3606	0.0001	0.5814
A_{inv}	2.7317	0.2082	3.5526	0.0000	0.8209
			4.0587	0.0002	1.3270
B	2.8553	0.2230	2.1771	0.0000	-0.6782
			2.7752	0.0016	-0.0801
B_{inv}	2.7996	0.1841	2.9226	0.0000	0.1230
			3.5091	0.0002	0.7095

[a]The two entries for each model are for two excited electronic states involving ET from Trp104 to FMN, the upper and lower numbers corresponding to the transitions depicted in Figs. S8 *A* and *B*, respectively, for the B_{inv} model. The "inv" subscript indicates that the conformation is conducive to the proton relay, while the absence of this subscript indicates the alternative conformation.

Table S4. Vertical excitation energies and oscillator strengths for excitation to the LE state and the CT state, which involves ET from Tyr21 to FMN, in different active site conformations in the gas phase and in the presence of the full protein and aqueous environment.[a]

		GS→LE		GS→CT		CT-LE gap
		E_{exc} (eV)	f	E_{exc} (eV)	f	ΔE(eV)
Tyr21/Gln63 not oriented for PT	Gas phase	3.0307	0.2523	4.2089	0.0002	1.1782
	Protein and aqueous environment	3.0542	0.2398	4.2844	0.0004	1.2302
Tyr21/Gln63 oriented for PT[b]	Gas phase I	2.9923	0.2387	3.3775	0.0000	0.3852
	Protein and aqueous environment I	3.0129	0.2200	3.6675	0.0014	0.6546
	Protein and aqueous environment II	3.0089	0.2052	3.6496	0.0696	0.6407

[a]The full protein and aqueous environment is described by an electrostatic embedding scheme.

[b]Structure I was generated by protein geometry optimization after manual rotation of the Tyr21 hydroxyl and Gln63 amide groups in an equilibrated structure with Tyr21 and Gln63 not oriented for PT. Structure II was generated by protein geometry optimization of an equilibrated structure with Tyr21 and Gln63 oriented for PT.

References

1. Anderson S, et al. (2005) Structure of a novel photoreceptor, the BLUF domain of AppA from Rhodobacter sphaeroides. *Biochemistry* 44(22):7998-8005.
2. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. *J Comput Chem* 30(10):1545-1614.
3. Best RB, et al. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput* 8(9):3257-3273.
4. Jorgensen WL (1981) Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J Am Chem Soc* 103(2):335-340.
5. Freddolino PL, Gardner KH, Schulten K (2013) Signaling mechanisms of LOV domains: New insights from molecular dynamics studies. *Photochem Photobiol Sci* 12(7):1158-1170.
6. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems. *J Chem Phys* 98(12):10089-10092.
7. Hynninen A-P, Crowley MF (2014) New faster CHARMM molecular dynamics engine. *J Comput Chem* 35(5):406-413.
8. Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23(3):327-341.
9. Andersen HC (1980) Molecular dynamics simulations at constant pressure and/or temperature. *J Chem Phys* 72(4):2384-2393.
10. Nose S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. *Mol Phys* 50(5):1055-1076.
11. Hoover WG (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys Rev A* 31(3):1695-1697.
12. Casida ME, Huix-Rotllant M (2012) Progress in time-dependent density-functional theory. *Annu Rev Phys Chem* 63:287-323.
13. Frisch MJ, et al. (2009) Gaussian 09, Revision E.01. *Gaussian, Inc., Wallingford CT*.
14. Dickson BM, Huang H, Post CB (2012) Unrestrained computation of free energy along a path. *J Phys Chem B* 116(36):11046-11055.
15. E W, Ren W, Vanden-Eijnden E (2005) Finite temperature string method for the study of rare events. *J Phys Chem B* 109(14):6688-6693.
16. Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J Comput Phys* 23(2):187-199.
17. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules .I. The method. *J Comput Chem* 13(8):1011-1021.
18. Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. *Angew Chem Int Ed* 48(7):1198-1229.
19. Gaus M, Cui Q, Elstner M (2011) DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J Chem Theory Comput* 7(4):931-948.
20. Im W, Berneche S, Roux B (2001) Generalized solvent boundary potential for computer simulations. *J Chem Phys* 114(7):2924-2937.

21. Nina M, Im W, Roux B (1999) Optimized atomic radii for protein continuum electrostatics solvation forces. *Biophys Chem* 78(1-2):89-96.
22. Brooks III CL, Karplus M (1983) Deformable stochastic boundaries in molecular dynamics. *J Chem Phys* 79(12):6312-6325.
23. Konig PH, Hoffmann M, Frauenheim T, Cui Q (2005) A critical evaluation of different QM/MM frontier treatments with SCC-DFTB as the QM method. *J Phys Chem B* 109(18):9082-9095.
24. Jung A, Reinstein J, Domratcheva T, Shoeman RL, Schlichting I (2006) Crystal structures of the AppA BLUF domain photoreceptor provide insights into blue light-mediated signal transduction. *J Mol Biol* 362(4):717-732.
25. Yanai T, Tew DP, Handy NC (2004) A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem Phys Lett* 393(1-3):51-57.
26. Singh UC, Kollman PA (1984) An approach to computing electrostatic charges for molecules. *J Comput Chem* 5(2):129-145.
27. Besler BH, Merz Jr. KM, Kollman PA (1990) Atomic charges derived from semiempirical methods. *J Comput Chem* 11(4):431-439.
28. Nilsson L, Halle B (2005) Molecular origin of time-dependent fluorescence shifts in proteins. *Proc Natl Acad Sci USA* 102(39):13867-13872.
29. Lin B, Gao Y, Li Y, Zhang JZH, Mei Y (2014) Implementing electrostatic polarization cannot fill the gap between experimental and theoretical measurements for the ultrafast fluorescence decay of myoglobin. *J Mol Model* 20(4):2189.
30. Sobieraj M, et al. (2015) A QM-MD simulation approach to the analysis of FRET processes in (bio)molecular systems. A case study: complexes of E. coli purine nucleoside phosphorylase and its mutants with formycin A. *J Mol Model* 21(4):75.