THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

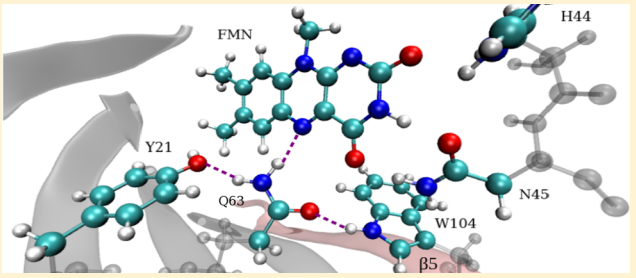
The Central Role of Gln63 for the Hydrogen Bonding Network and UV−Visible Spectrum of the AppA BLUF Domain

Ya-Wen Hsiao,[‡] Jan P. Götze, and Walter Thiel*

Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470, Mülheim an der Ruhr, Germany

Supporting Information

ABSTRACT: In blue-light sensing using flavin (BLUF) domains, the side-chain orientation of key residues close to the flavin chromophore is still under debate. We report quantum refinements of the wild-type AppA BLUF protein from *Rhodobacter sphaeroides* starting from two published X-ray structures (1YRX and 2IYG) with different arrangements of the residues around the chromophore. Quantum refinement uses the same experimental X-ray raw data as conventional refinement, but includes data from quantum mechanics/molecular mechanics (QM/MM) calculations as restraints, which is expected to be more reliable than the normally employed MM data. In addition to quantum refinement, pure QM/MM geometry optimizations are performed for the 1YRX and 2IYG structures and for five models derived therefrom. Vertical excitation energies are computed at the QM(DFT/MRCI)/MM level to assess the resulting structures. The experimental absorption maximum of the dark state of wild-type AppA is well reproduced for structures that contain the Gln63 residue in 1YRX-type orientation. The computed excitation energies are red-shifted for structures with a flipped Gln63 residue in 2IYG-type orientation. The calculated 1YRX- and 2IYG-type hydrogen-bonding networks are discussed in detail, particularly with regard to the orientation of the chromophore and the Gln63, Trp104, and Met106 residues.



INTRODUCTION

Blue-light sensing using flavin (BLUF) proteins belong to the family of flavin-dependent blue-light photoreceptors.[1] Such receptors play an important role in the light regulatory mechanisms of bacteria, plants, and mammals. Other members in this photoreceptor family are, e.g., the cryptochromes and light−oxygen−voltage (LOV) domain-containing proteins. Upon blue-light illumination, the BLUF-containing proteins exhibit a red-shift of about 15 nm in the UV−visible (UV−vis) absorption spectrum,[2,3] from 445 to 460 nm for the lowest absorption of the chromophore.[3] This red-shift is believed to be related to the formation of a signaling state.[2] Additionally, there is a downshift of around 20 cm^{-1} of the C4═O4 stretching wavenumber in FTIR spectra associated with the signaling state.[4,5]

Both the BLUF and LOV domains contain a flavin chromophore. In contrast to the well-characterized LOV domains, the nature of the conformational changes in the BLUF proteins upon illumination is not yet fully understood. There is consensus that the isoalloxazine ring of the flavin chromophore is not covalently bound to the protein and remains structurally unchanged after photon absorption.[2−4,6] The neighboring hydrogen-bonding network is believed to undergo rearrangement after illumination, which then gives rise to the observed red-shift in the UV−vis absorption spectrum. However, even in the dark state (i.e., the initial nonsignaling conformation that is present before photoexcitation), the structure of the hydrogen-bonding network is still under debate. There are two types of dark-state X-ray crystal structures available, i.e., one obtained by Anderson et al.[7] (PDB code 1YRX) for wild type AppA, the most studied BLUF protein from the purple photosynthetic bacterium *Rhodobacter sphaeroides*, and the other one determined by Jung et al.[8] (PDB code 2IYG) for the AppA1-124C20S mutant. These published crystal structures are similar in their overall appearance, but differ in the arrangement of the active-site residues, in particular with regard to the side-chain orientation of Trp104, Met106, and Gln63 (see the schematic representation in Figure 1). The side chain of Trp104 is oriented inside the active site in 1YRX, forming a hydrogen bond with the Gln63 side-chain carbonyl oxygen atom, while Trp104 has an outward orientation in 2IYG, with the side chain of Met106 being inside instead. The Gln63 residue is flipped in the two structures: it has a hydrogen bonding interaction with Tyr21 through its amide nitrogen atom in 1YRX, and through its amide oxygen atom in 2IYG. It has been pointed out[8] that the 2IYG structural model of the dark state corresponds well with those for the BLUF domains of other proteins such as BlrB.[9] On the other hand, a solution NMR study of wild type AppA in the dark state supports the 1YRX structural model: more than 20 NOE signals show that Trp104 is buried in the core of wild type AppA and is able to form a hydrogen bond with the Gln63 carbonyl oxygen atom.[10]

Received: March 26, 2012
Revised: June 12, 2012
Published: June 13, 2012

dx.doi.org/10.1021/jp3028758 | J. Phys. Chem. B 2012, 116, 8064−8073

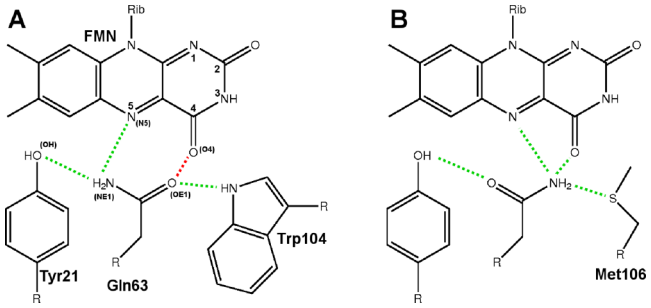


Figure 1. Schematic representation of the central BLUF hydrogen bonding network in 1YRX (A) and 2IYG (B). Green: hydrogen bonds; red: oxygen−oxygen repulsion.

This inside conformation of Trp104 in the dark state of wild-type AppA is also supported by steady-state fluorescence[11] and ultraviolet resonance Raman[12] spectroscopy studies. A recent spectroscopic investigation of the hydrogen-bond switch reaction in the BlrB protein has also addressed possible alternative hydrogen-bond patterns of the active-site residues in BlrB and AppA.[13] The possible Trp/Met arrangements are sketched in Figure 2.

Another controversial topic is how the orientation of the active-site residues changes in AppA BLUF domains upon illumination. Fluorescence and biochemical studies suggest that the orientation of Trp104 switches from "inside" to "outside" when going from the dark state to the signaling state.[14,15] This view is supported by other spectroscopic experiments.[16,17] On the other hand, steady-state fluorescence spectroscopy[11] indicates that Trp104 does not become exposed to solvent during the BLUF photocycle and thus remains in an inside conformation, whereas a recent ultraviolet resonance Raman study concludes that Trp104 remains buried but undergoes some conformational changes upon formation of the signaling state.[12] The AppA1-124C20S mutant, in which Trp104 adopts the outside conformation already in the dark-state structure (2IYG), also exhibits a characteristic red shift upon illumination (5 cm^{-1}, i.e., smaller than usual), and it has been shown that the light-induced conformational alterations involve significant movement of the Met106 side chain.[8] Finally, according to a solution NMR study of wild-type AppA, the Gln63 residue flips upon illumination such that the signaling state has strong hydrogen bonds between the Gln63 carbonyl group and Tyr21 as well as between the Gln63 amide nitrogen atom and the flavin C4═O group.[10]

On the theoretical side, classical molecular dynamics (MD) simulations[18] favor structural model 2IYG over 1YRX, because MD runs starting from the 1YRX geometry of wild-type AppA soon led to a flip of the Gln63 side chain, resulting in 2IYG-type arrangement; by contrast, in 10 ns MD simulations of the AppA1-124C20S mutant and two other BLUF protein (BlrB and T110078), the 2IYG-type conformation of Gln63 remained stable (as well as the outside conformation of Trp104). It was thus concluded that Gln63 adopts a 2IYG-type orientation both in the dark and the signaling state (albeit with dynamically changing hydrogen bond strengths), and the instability of the 1YRX-orientation was ascribed to unfavorable electrostatic repulsions between the flavin O4 and Gln63 OE1 atoms.[18] A more recent systematic MD study[19] using different classical force fields (GROMOS 45A4 and GROMOS 53A6 as opposed to AMBERff03 in the previous work[18]) as well as quantum mechanics/molecular mechanics (QM/MM) potentials arrived at different results: the 1YRX and 2IYG geometries remained stable in classical 20 ns MD simulations started from the corresponding two crystal structures,[7,8] and the overall comparison of all the results obtained led to a slight preference for the 1YRX structure of the dark state. The quality of the different X-ray and NMR structures was assessed in a recent DFT/MM investigation of flavin IR spectra in BLUF domains, which assigned a 1YRX-type inside conformation of Trp104 to the dark state of AppA BLUF and favored for the signaling state (without being definitive) an inside conformation of Met106.[20] A comprehensive QM/MM study[21] addressed the photo-induced reaction cascade in BLUF photoreceptors and proposed a detailed mechanism for the formation of the signaling state: these QM/MM calculations were based on the crystal structure of BlrB[9] with a 2IYG-type orientation and covered different conformations of the crucial Trp and Met residues. The computed mechanism involved a tautomerization rather than a rotation of the active-site Gln residue, and the signaling state was thus predicted to contain the imidic acid isomer of the Gln residue.[21] Another QM/MM study on wild-type AppA[22] reported geometry optimizations that were started from the published 1YRX and 2IYG crystal structures: while the latter converged to a geometry close to the initial 2IYG structure (with an outside orientation of Trp104), the former resulted in significant structural relaxation, and the tautomeric imidic acid form of Gln63 was suggested to be present in 1YRX-type structures (with an inside orientation of Trp104). Further QM model calculations supported the notion that

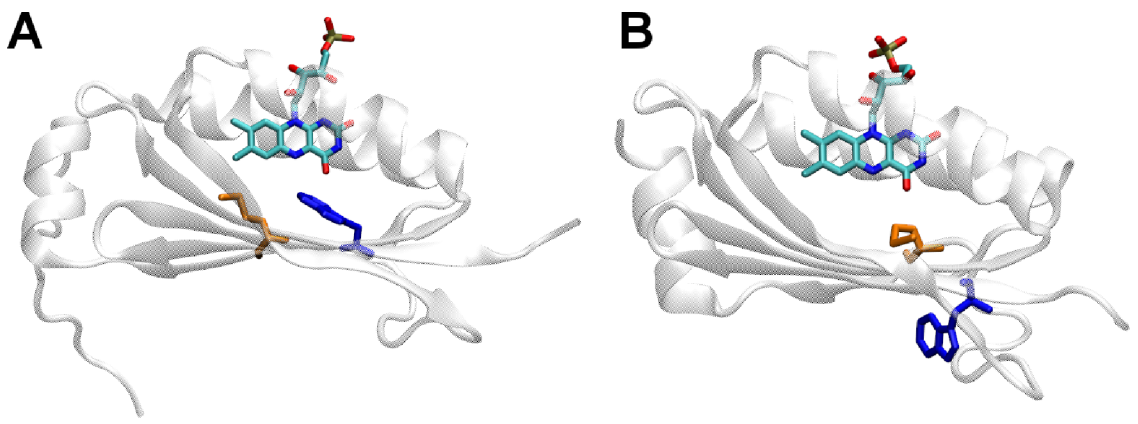


Figure 2. Orientation of Trp104 and Met106 in X-ray structures of the BLUF domain, 1YRX (A) and 2IYG (B). Blue: Trp104; orange: Met106.

dx.doi.org/10.1021/jp3028758 | J. Phys. Chem. B 2012, 116, 8064−8073

photon absorption by the flavin in the 2IYG-type dark state leads to tautomerization and rotation of Gln63, with formation of the signaling state.[22] Finally, in a QM study[23] of an AppA model system (165 atoms) derived from a solution NMR structure (2BUN), it was concluded from the computed UV−vis spectra (TD-B3LYP/6-31G*) that the light-induced switching of Gln63 and Ser41 is essential for explaining the spectral 15 nm red shift, whereas the geometrical changes for Trp104 and Met106 upon illumination have little effect on the spectra.

In view of all these conflicting results and structural proposals, we decided to reinvestigate the two published crystal structures (1YRX and 2IYG) using quantum refinement[24] and QM/MM[25] methods. To assess the resulting geometries, we computed vertical excitation energies using a combination of density functional theory and multireference configuration interaction (DFT/MRCI).[26] Furthermore, we constructed additional structural models computationally, by deliberately modifying the side-chain orientation of key residues, in order to explore the relationship between structures and spectroscopic properties in more detail. Quantum refinement is considered to give more reliable heteromolecular structures than conventional X-ray crystallographic refinement: while both approaches use the same experimental X-ray raw data, quantum refinement includes QM/MM energy and gradient information, which is generally more reliable than the MM force field information employed in the conventional refinement process. Alternatively, quantum refinement can be regarded as QM/MM geometry optimization restrained by experimental X-ray raw data, which will help to avoid artifacts of purely theoretical QM/MM calculations.[27,28] Quantum refinement has been applied successfully in a number of recent studies.[24,27−38] The comparison between DFT/MRCI excitation energies at the resulting geometries and the corresponding experimental data then provides another check on the quality of the derived structures, as demonstrated in previous work.[28,38]

COMPUTATIONAL DETAILS

Quantum Refinement and QM/MM. The original protein coordinates and structure factors were obtained from the Protein Data Bank[39] (PDB codes 1YRX and 2IYG at resolution 2.3 Å). Missing hydrogen atoms (needed for QM/MM) were added by using the HBUILD module in CHARMM[40] using standard protonation state conventions, except for His. All His residues were assigned as HSE (only NE2 protonated), with the following exceptions: His44 was treated as HSP (both ND1 and NE2 protonated) in both structures. His78 was also assigned as HSP in 1YRX due to the presence of oxygen atoms at hydrogen-bond distance, but not in 2IYG (HSE) where the His78 side chain was flipped in the original assignment compared to that of 1YRX. His85 was defined with ND1 being unprotonated, because it is located near Ser23 and points with ND1 toward the hydroxyl oxygen of Ser23 that is less than 3 Å apart. The phosphate group of flavin was assumed to carry a charge of −2, which leads to a total charge of −1 for each chain of 1YRX; the surface residue Glu94 was protonated in each chain to neutralize the 1YRX structure. The 2IYG structure was made neutral by protonating the surface residues Asp15, Glu74, and Glu93 in the A chain as well as Asp15, Glu39, and Glu74 in the B chain. The positions of the added hydrogen atoms were then relaxed at the MM level with 10 000 steps of steepest descent, followed by an augmented basis Newton−Raphson (ABNR) minimization.

The QM region in the QM/MM calculations consisted of the chromophore flavin mononucleotide (FMN) with the phosphate group being removed and the side chains of the active-site residues Tyr21, His44, Asn45, Gln63, and Trp104/Met106 (1YRX/2IYG), which were included because of their geometrical closeness to the chromophore. The covalent bonds across the QM-MM boundary were capped with hydrogen link atoms. There were six such junctions: between CA and CB for Tyr21, His44, and Asn45; between CB and CG for Gln63 and Trp104/Met106; and between C1′ and C2′ for FMN.

Turbomole[41] and DL_POLY[42] were used to calculate the energies and gradients of the QM and MM parts, respectively. ChemShell[43,44] was employed to handle the coordinate, energy, and gradient data communication between the QM, MM, and crystallographic refinement programs, and to perform the geometry optimizations through the HDLC optimizer.[45] All QM calculations were done using DFT with the BP86 functional[46,47] and the 6-31G* basis set.[48,49] This combination has proven to be efficient and reliable as long as geometry is the main concern: it normally reproduces bond lengths of organic molecules to within 0.02 Å, which is better than medium-resolution protein crystal structures where the average error in bond lengths is about 0.1 Å. The BP86/6-31G* combination has been successfully applied in several earlier studies.[24,28−33,38] Empirical dispersion corrections were added within the QM region after the SCF iterations.[50,51] The MM force field parameters for proteins were taken from CHARMM22.[52]

Quantum refinement in the form of an X-ray-data-restrained QM/MM minimization has recently been implemented in the ChemShell package.[28] In this procedure, CNS[53] is applied to assist in the refinement; it calculates the X-ray structure factor amplitudes F for each new protein structure and then evaluates maximum likelihood with amplitudes (MLF) to obtain a hypothetical energy E_{xref} and the corresponding gradient.[54] E_{xref} is proportional to $(|F_o(hkl)| - \langle|F_o(hkl)|\rangle)^2$, where the expected value of $\langle|F_o(hkl)|\rangle$ is derived from the observed F_o and the calculated F_c values (for details, see ref 49). The energy function of quantum refinement takes the following form:

$$E_{total} = E_{QM/MM} + w_{xref}E_{xref} \quad (1)$$

w_{xref} values of 0.5 and 1.0 were used in this study. The quality of a refined model is judged by its R_{free} factor, which is defined in analogy to the R factor:

$$R = \frac{\sum_{hkl} ||F_o(hkl)| - |F_c(hkl)||}{\sum_{hkl} |F_o(hkl)|} \quad (2)$$

The refinements include all reflections from the original PDB data sets (27 504 for 1YRX and 16 135 for 2IYG). The R_{free} factor is obtained from a set of randomly chosen reflections (5.3% for 1YRX and 5% for 2IYG), which are set aside from the beginning and not included in the refinement. It provides a cross-validation in order to avoid overfitting of the diffraction data.[55]

According to our previous experience, quantum refinement with electrostatics included seems to perform equally well or better in protein structure determination than without.[28,38] Therefore we chose an electrostatic quantum refinement in the present study. We used our standard setup for QM/MM calculations with electrostatic embedding. Briefly, the protein was solvated in a sphere of TIP3P water molecules[56] with a radius of 30 Å, centered at the N5 atom of FMN in the A chain. The water sphere covered the whole A chain that was the focus

dx.doi.org/10.1021/jp3028758 | J. Phys. Chem. B 2012, 116, 8064−8073

Table 1. Quantum Refinement: Hydrogen Bonding Network around the FMN Chromophore (Interatomic Distances in Å) in the 1YRX and 2IYG Structures, and the Corresponding Calculated Lowest Excitation Energies ΔE and Oscillator Strengths f

model	1YRX			2IYG		
w_{xref}	orig	0.5	1.0	orig	0.5	1.0
R_{free}	0.273	0.346	0.345	0.233	0.283	0.283
OH(Y21)−NE2(Q63)	2.52	2.95	2.89			
OH(Y21)−OE1(Q63)				2.57	2.62	2.62
NE2(Q63)−N5(FMN)	3.05	2.96	2.96			
OE1(Q63)−N5(FMN)				3.36	3.18	3.15
OE1(Q63)−O4(FMN)	2.67	2.94	2.91			
NE2(Q63)−O4(FMN)				2.78	2.86	2.81
ND2(N45)−O4(FMN)	2.89	2.87	2.85	3.21	3.09	3.17
OD1(N45)−N3(FMN)	2.74	2.75	2.76	2.82	2.80	2.84
ND1(H44)−O2(FMN)	3.14	2.94	2.94	3.33	3.19	3.31
ND1(H44)−OD1(N45)	3.64	3.48	3.55	3.43	3.33	3.36
OE1(Q63)−NE1(W104)	3.29	2.89	3.01			
NE2(Q63)−CE(M106)				3.17	3.21	3.20
ΔE(eV)	2.37	2.77	2.81	2.16	2.68	2.69
ΔE(nm)	523	447	441	575	462	460
f	0.23	0.31	0.30	0.20	0.27	0.27

of our investigations, while the B chain (and the C chain in 1YRX) were kept fixed. To retain the shape of the surrounding water droplet and to avoid evaporation of surface water molecules, a spherical quartic boundary potential was applied acting on the corresponding oxygen atoms. The water molecules were equilibrated at the MM level using CHARMM (program version 31b1), and the system underwent cycles of rehydration and constant-temperature MD as described previously:[27,38] after initial placement of the water sphere, overlapping water molecules were deleted, and the remaining ones were energy minimized for 1000 cycles followed by a 15 ps constant-temperature (300 K) classical MD simulation. Subsequently, a new water sphere was superimposed on the system with overlapping atoms deleted, and was again subjected to a 15 ps MD run. This procedure was repeated 12 times, with the equilibration time being increased to 30 ps in the last three cycles. During these classical setup calculations, the heavy-atom coordinates of the protein were not allowed to change. In the subsequent quantum refinement and the QM/MM calculations with electrostatic embedding, no cutoff was applied for the electrostatic QM/MM interactions. The evaluation of the target function in the refinement did not include the added solvating water molecules. Further technical details on the ChemShell implementation of quantum refinement are available in the original report.[28]

DFT/MRCI. DFT/MRCI[26] was employed to calculate the excitation energies of the chromophore. In this approach, dynamic electron correlation is taken into account by DFT, while static correlation effects are captured by a limited configuration interaction (CI) expansion. The configuration state functions in the MRCI expansion are built from Kohn−Sham orbitals, and the diagonal elements of the Hamiltonian are constructed from the corresponding Hartree−Fock-based expression and a DFT-specific correction term. Using the standard values for the five empirical DFT/MRCI parameters, the typical error in vertical excitation energies is around 0.2 eV compared to experimental values[26] and to high-level ab initio reference data.[57−59]

Molecular orbitals (MOs) were generated with Turbomole using the BH-LYP functional[46,60,61] and the TZVP basis set.[62,63] The active space for the MRCI treatment was built from double excitations of 10 electrons in 10 frontier MOs. The CI space was kept moderate by selecting only configurations with an estimated energy below a cutoff δE_{sel} above the highest reference-space energy. We adopted the DFT/MRCI parameters recently developed in the Grimme group[64] in combination with a tight selection criterion δE_{sel} for enhanced computational efficiency. An initial cutoff of 0.6 hartree was applied in the selection, which was increased to 0.8 hartree from the second cycle onward. The reference configurations were determined iteratively: configurations with a squared coefficient greater than 0.003 were included in the reference space for the next CI calculation, and the iterations were stopped when the lowest 10 roots with singlet multiplicity were converged to 10^{-6} hartree.

DFT/MRCI has been established as a reliable tool for predicting UV−vis spectra.[27,58,59] Therefore it is particularly suitable for the purpose of this study: to find structures corresponding to the dark state based on the closeness of the calculated excitation energies to the experimental values.

RESULTS AND DISCUSSION

Quantum Refinement. Quantum refinement was performed on the published crystal structures 1YRX and 2IYG using two weight factors (w_{xref} = 0.5 and 1.0). The resulting refined geometries were quite similar in both cases, and the final R_{free} values for these two choices turned out to be alike. They were higher than those from the published refinements (see Table 1) which might be due to the treatment of only one of the multiple chains and the use of a different X-ray crystallographic refinement program in the present work. Interatomic distances between key residues around the flavin chromophore are shown in Table 1 for the published structures (orig) and the current quantum refinements, along with the calculated spectroscopic properties. The calculated vertical excitation energies are rather low when computed at the published geometries (1YRX, 2.37 eV; 2IYG, 2.16 eV), but much higher and rather uniform when using the currently refined geometries (1YRX-based, 2.77−2.81 eV; 2IYG-based, 2.68−2.69 eV).

The overall agreement of the presently determined structures with the published crystal structures is measured by the

dx.doi.org/10.1021/jp3028758 | J. Phys. Chem. B 2012, 116, 8064−8073

corresponding root-mean-square deviations (rmsd). In the case of 1YRX, a purely theoretical QM/MM geometry optimization yields rmsd values of 0.561 Å for the A chain and 0.355 Å for the QM region. Quantum refinements with w_{xref} = 0.5 (1.0) reduce these values significantly to 0.173 (0.167) Å for the A chain and 0.181 (0.186) Å for the QM region, in spite of the fact that the contribution of the X-ray reference data to the total energy function in eq 1 amounts to only 4% (8%) of the QM/MM contribution (in absolute value). Similar results are found in the case of 2IYG: for example, the rmsd values for the QM region decrease from 0.327 Å in the QM/MM structure to 0.120 (0.109) Å in the quantum refinement structures obtained with w_{xref} = 0.5 (1.0). Rather weak restraints by the X-ray reference data are thus sufficient to overcome the major overall differences between the QM/MM optimized and the PDB structures. Since the remaining deviations are quite similar for the two chosen weight factors, we did not consider it necessary to run further quantum refinements with larger w_{xref} values.

For the sake of brevity, we address in the following discussion of the refined structures only the results obtained with w_{xref} = 1.0 (see Tables S1−S2 of the Supporting Information for the Cartesian coordinates obtained for the QM atoms). In the case of 1YRX, quantum refinement leads to large changes (>0.2 Å) from the published structure for the distances R(OH(Y21)−NE2(Q63)), lengthened by 0.37 Å; R(OE1(Q63)−O4(FMN)), lengthened by 0.24 Å; and R-(OE1(Q63)−NE1(W104)), shortened by 0.28 Å. These three changes all involve Gln63 and are correlated with each other, since they reflect Gln63 being placed further away from Tyr21 and FMN, and closer to Trp104 (see Figure 3). Gln63 is thus able to form a stronger hydrogen bond with the side chain of Trp104 in the refined structure, while decreasing its interaction with the flavin C4═O4 carbonyl group. An overlay of the refined and the original 1YRX structures (not shown) confirms that Gln63 moves during quantum refinement, while the position of Tyr21 does not change much.

In the case of 2IYG, there is only one major structural change upon quantum refinement, namely, a shortening of the distance R(OE1(Q63)−N5(FMN)) by 0.21 Å; even so, it is then still 0.19 Å longer than its counterpart R(NE2(Q63)−N5(FMN)) in 1YRX. R(NE2(Q63)−O4(FMN)) does not change much during quantum refinement and is 0.10 Å shorter than R(OE1(Q63)−O4(FMN)) in 1YRX. R(OH(Y21)−OE1-(Q63)) changes only slightly and is much shorter than the corresponding distance R(OH(Y21)−NE2(Q63)) in 1YRX, by 0.27 Å. R(NE2(Q63)−CE(M106)) remains long after quantum refinement, arguing against a strong hydrogen-bonding interaction between Gln63 and Met106 in 2IYG. According to the computed distances in 2IYG after quantum refinement, Gln63 forms strong hydrogen bonds with Tyr21 (via OE1, 2.62 Å) and the FMN C4═O4 carbonyl group (via NE2, 2.81 Å), but only weak such interactions with the flavin N5 atom (via OE1, 3.15 Å) and with Met106 (via NE2, 3.20 Å). Experimentally, formation of the signaling state upon illumination is associated with the formation of strong hydrogen bonds between OH(Tyr21) and OE1(Gln63) as well as between the flavin C4═O4 group and NE2(Gln63),[10] and with a decrease of the C4═O4 stretching frequency by about 20 cm^{-1} compared with the dark state (indicating stronger hydrogen bonding interactions involving C4═O4 in the signaling state). According to our quantum refinement results, all these features are consistent with a 2IYG-type structure for the signaling state.

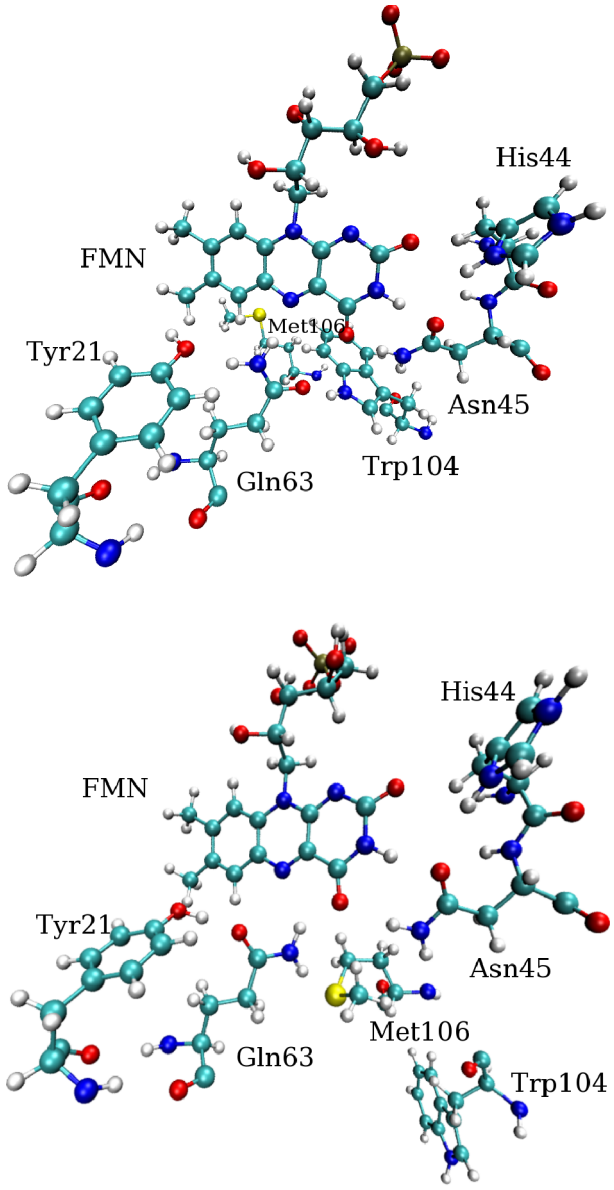


Figure 3. The chromophore and its neighboring residues: 1YRX structure (upper panel) and 2IYG structure (lower panel).

The CN and CO bond lengths in the Gln63 side chain are determined to be 1.34−1.35 and 1.24−1.26 Å, respectively, both in the 1YRX and 2IYG quantum refinement structures. These values are typical for amides. In the alternative imidic tautomer of Gln63, the CN (CO) bond would be expected to be significantly shorter (longer). The quantum refinement thus retains the amide character of the Gln63 residue found in the crystal structures.

The vertical excitation energy of wild-type AppA is 2.79 eV (445 nm).[2,3] DFT/MRCI calculations of this quantity have an error bar of typically 0.2 eV.[58,59] The DFT/MRCI results at the published X-ray geometries show much larger deviations (1YRX too low by 0.42 eV, 2IYG too low by 0.63 eV), indicating that these geometries are not realistic. At the currently refined structures, the DFT/MRCI excitation energies are 2.81 eV (1YRX-based) and 2.69 eV (2IYG-based). The former is in almost perfect agreement with experiment, while the latter is somewhat too low (but still within the usual error

dx.doi.org/10.1021/jp3028758 | J. Phys. Chem. B 2012, 116, 8064−8073

range); hence, these results favor the 1YRX-based structure, however without being conclusive.

QM/MM Geometry Optimizations. As already mentioned, much of the recent debate on the AppA active-site structure has focused on the side-chain orientations of Gln63, Trp104, and Met106. To explore the influence of these orientations on the spectral properties, we carried out QM/MM geometry optimizations on the two structures obtained from quantum refinement and on the following five models derived therefrom (depicted in Figure S1 of the Supporting Information).

1 Model 1-Q63$_{flip}$ is derived from the 1YRX structure by flipping the Gln63 side chain so that the amide oxygen points at Tyr21. This generates a wild-type AppA structure with Gln63 positioned as in the 2IYG structure of the AppA1-124C20S mutant.
2 Model 2-Q63$_{flip}$ is obtained from the 2IYG structure by flipping the Gln63 chain side so that the amide nitrogen points at Tyr21. This generates an AppA1-124C20S mutant structure with Gln63 oriented as in the 1YRX structure of wild-type AppA. It differs from the first model, especially in the arrangement of the Trp104 side chain (either outside or inside the active site).
3 Model W104F is a mutant derived from the 1YRX structure by replacing Trp104 with Phe. The side chain of Phe does not have a strong hydrogen bond donor as does Trp104, and therefore this model may offer some insight into the influence of the hydrogen bond between Gln63 and Trp104.
4 Model W104$_{out}$ is generated from the 1YRX structure by moving the Trp104 side chain out of the active site. This should help us to assess the effects of this residue.
5 Model W104$_{out}$M106$_{in}$ is obtained from the 1YRX structure by moving the Trp104 side chain out of the active site and having the Met106 side chain point into the active site. This creates a wild-type AppA structure with Trp104 and Met106 arranged as in the 2IYG structure of the AppA1-124C20S mutant.

The starting geometries for constructing these models were taken from the quantum refinement (w_{xref} =1.0) for 1YRX and 2IYG (see preceding section) and were then manually modified. Thereafter the first two models were directly subjected to QM/MM geometry optimization, whereas a regular setup and solvation procedure was carried out for the other three models before performing the QM/MM optimization (see Tables S3−S9 of the Supporting Information for the optimized Cartesian coordinates obtained for the QM atoms).

Table 2 lists the QM/MM optimized distances in the hydrogen-bond network around the flavin chromophore for the three structures with the highest calculated absorption energies (2.73−2.76 eV), namely, 1YRX, 2-Q63$_{flip}$, and W104F. Table 3 contains the corresponding data for the four structures with the lowest computed excitation energies (2.63−2.68 eV), i.e., 1-Q63$_{flip1}$, W104$_{out}$, W104$_{out}$M106$_{in}$, and 2IYG. Experimentally, the maximum of the first absorption band is at 445 nm (2.79 eV) for the dark state, and at 460 nm (2.70 eV) for the red-shifted state of wild-type AppA formed upon illumination.[2] As already noted, errors in DFT/MRCI excitation energies are typically on the order of 0.2 eV.

We first discuss the results in Table 2. The QM/MM optimized structure for the 1YRX model is generally compatible with the corresponding structure from quantum refinement (1), but there are also some notable differences. Upon QM/MM optimization, Gln63 moves closer to Trp104 and further away from O4(FMN), as indicated by the significant increase of R(OE1(Q63)−O4(FMN)) and decrease of R(OE1(Q63)−NE1(W104)); in addition, the hydrogen bond between His44 and Asn45 becomes considerably shorter [see R(ND1(H44)−OD1(N45))].

The QM/MM optimized structure of the 2-Q63$_{flip}$ model is by construction similar to the 2IYG mutant structure, except for the flipped Gln63 residue with a local 1YRX-type environment. Consequently, the hydrogen-bond distances involving Gln63 closely resemble those in the 1YRX quantum refinement structure (Table 1) [see R(OH(Y21)−NE2(Q63)), R(NE2-

Table 2. QM/MM Optimization: Hydrogen Bonding Network around the FMN Chromophore (Interatomic Distances in Å) of Structures with a Calculated Lowest Absorption Energy above 2.70 eV, and the Corresponding Calculated Lowest Excitation Energies Δ*E* and Oscillator Strengths *f*

	1YRX	2-Q63$_{flip}$	W104F
OH(Y21)−NE2(Q63)	3.03	2.93	3.05
NE2(Q63)−N5(FMN)	3.11	3.05	3.13
OE1(Q63)−O4(FMN)	3.56	3.07	3.32
ND2(N45)−O4(FMN)	2.94	2.91	2.93
OD1(N45)−N3(FMN)	2.84	2.75	2.80
ND1(H44)−O2(FMN)	3.05	2.94	3.07
ND1(H44)−OD1(N45)	2.98	3.28	3.06
OE1(Q63)−NE1(W104)	2.74		
OE1(Q63)−CE(M106)		2.79	
OE1(Q63)−CE2(F104)			2.97
Δ*E*(eV)	2.74	2.73	2.76
Δ*E*(nm)	452	455	450
f	0.32	0.30	0.43

Table 3. QM/MM Optimization: Hydrogen Bonding Network around the FMN Chromophore (Interatomic Distances in Å) of Structures with a Calculated Lowest Absorption Energy below 2.70 eV, and the Corresponding Calculated Lowest Excitation Energies Δ*E* and Oscillator Strengths *f*

	1-Q63$_{flip}$	W104$_{out}$	W104$_{out}$M106$_{in}$	2IYG
OH(Y21)−NE2(Q63)		2.98	2.97	
OH(Y21)−OE1(Q63)	2.69			2.64
NE2(Q63)−N5(FMN)		3.01	2.95	
OE1(Q63)−N5(FMN)	3.17			3.34
OE1(Q63)−O4(FMN)		3.12	2.97	
NE2(Q63)−O4(FMN)	2.82			3.37
ND2(N45)−O4(FMN)	2.90	2.90	2.92	2.94
OD1(N45)−N3(FMN)	2.83	2.79	2.78	2.76
ND1(H44)−O2(FMN)	2.96	3.04	2.93	2.83
ND1(H44)−OD1(N45)	2.94	3.00	3.04	3.61
NE2(Q63)−NE1(W104)	3.21			
OE1(Q63)−CE(M106)			8.47	
NE2(Q63)−CE(M106)				3.32
OE1(Q63)−O(H_2O_{2689})		2.74	2.75	
OE1(Q63)−O(H_2O_{2996})		2.64	2.70	
Δ*E*(eV)	2.63	2.68	2.66	2.63
Δ*E*(nm)	471	463	466	471
f	0.27	0.30	0.28	0.26

dx.doi.org/10.1021/jp3028758 | J. Phys. Chem. B 2012, 116, 8064−8073

Table 4. DFT/MRCI Results for the Structures Obtained from the PDB Database (orig), from Quantum Refinement (w_{xref} = 1), and from QM/MM Geometry Optimization: Vertical Excitation Energies (eV)/Wavelengths (nm), Oscillator Strengths (in parentheses), and Characterization of the Two Lowest Intense Transition to Singlet Excited States[a]

system	transition		ΔE
1YRX (orig)	LE(FMN)	HOMO→LUMO	2.37/523 (0.23)
	LE-CT(Y21+Q63+FMN→FMN)	HOMO-3→LUMO	3.27/380 (0.13)
1YRX (w_{xref} = 1)	LE(FMN)	HOMO→LUMO	2.81/441 (0.30)
	LE-CT(Q63+FMN→FMN)	HOMO-3→LUMO	3.49/355 (0.18)
1YRX(QM/MM)	LE(FMN)	HOMO→LUMO	2.74/452 (0.32)
	LE(FMN)	HOMO-2→LUMO	3.48/357 (0.23)
2-Q63$_{flip}$	LE(FMN)	HOMO→LUMO	2.73/455 (0.30)
	LE(FMN)	HOMO-3→LUMO	3.36/369 (0.24)
W104F	LE(FMN)	HOMO→LUMO	2.76/450 (0.32)
	LE(FMN)	HOMO-3→LUMO	3.43/361 (0.24)
2IYG (orig)	LE-CT(Y21+FMN→FMN)	HOMO-1→LUMO	2.16/575 (0.20)
	LE(FMN)	HOMO-2→LUMO	3.06/405 (0.23)
2IYG (w_{xref} = 1)	LE(FMN)	HOMO-1→LUMO	2.69/460 (0.27)
	LE(FMN)	HOMO-2→LUMO	3.35/370 (0.28)
2IYG(QM/MM)	LE(FMN)	HOMO-1→LUMO	2.63/471 (0.26)
	LE(FMN)	HOMO-2→LUMO	3.22/385 (0.27)
1-Q63$_{flip}$	LE(FMN)	HOMO→LUMO	2.63/471 (0.27)
	LE(FMN)	HOMO-2→LUMO	3.34/371 (0.26)
W104$_{out}$	LE(FMN)	HOMO→LUMO	2.68/463 (0.30)
	LE(FMN)	HOMO-2→LUMO	3.38/367 (0.24)
W104$_{out}$M106$_{in}$	LE(FMN)	HOMO-1→LUMO	2.66/466 (0.28)
	LE(FMN)	HOMO-3→LUMO	3.37/368 (0.25)

[a]See text for notation.

(Q63)−N5(FMN)), and R(OE1(Q63)−O4(FMN))], which are all slightly lengthened. Upon QM/MM optimization, the side chain of Met106 remains in the active site and moves closer to Gln63 (see R(OE1(Q63)−CE(M106)), 2.79 Å). Apparently, a bidentate CH···O hydrogen-bond-like interaction is established between OE1(Q63) and two hydrogen atoms at CE(M106). Such interactions have been discussed in the literature:[65−68] although much weaker than OH···O or NH···O interactions, they may be crucial for bioreversibility, e.g., in protein−ligand binding. In the 2-Q63$_{flip}$ model, the side chain of Met106 thus does not by itself associate with the chromophore in the dark state (as has been suggested for 2IYG[8,18]), but, like Trp104, it should be able to tune the spectrum presumably by interacting with OE1(Q63).

The QM/MM optimized structure of the mutant W104F (with Phe replacing Trp104) is quite similar to that of the 1YRX wild type (Table 2). The closest distances between the OE1 atom of Gln63 and carbon atoms on the Phe side chain involve CZ (3.10 Å) and CE2 (2.97 Å), indicating again a bidentate CH···O interaction. Experimentally, it has been shown that the W104F mutant of AppA undergoes a similar spectral change upon illumination as the wild type, but the lifetime of its signaling state is shorter.[15] Moreover, there is biochemical and genetic evidence that the hydrogen bond between Gln63 and Trp104 in AppA plays a critical role in blue light sensing.[15] In our QM/MM optimized structures, there is a hydrogen bond between the OE1 atom of Gln63 and Trp104 in the case of 1YRX, which is replaced by hydrogen-bond-like interactions with Met106 and Phe104 in the 2-Q63$_{flip}$ and W104F models, respectively; in all three cases, however, the computed excitation energies remain almost the same (2.73−2.76 eV; see Table 2). This is consistent with conclusions from earlier work that the side chain orientation of Trp104 has only a minor influence on the absorption energy.[12,23]

We now turn to the four models with lower calculated absorption energies. (Table 3). The 1-Q63$_{flip}$ model was constructed by merely exchanging the positions of OE1 and NE2 of Gln63 in the 1YRX model. At the QM/MM optimized geometries, the excitation energy of the 1-Q63$_{flip}$ model is lower than that of 1YRX (2.63 vs 2.74 eV), which corresponds to a computed red shift of 19 nm. Upon QM/MM optimization, the side chain of Trp104 stays in the active site of the 1-Q63$_{flip}$ model, but the hydrogen-bonding interactions with the flipped Gln63 residue are weaker than those with unflipped Gln63 in the 1YRX model, simply because the NH_2 amide moiety of flipped Gln63 is a weaker hydrogen bond acceptor than the C═O group in unflipped Gln63; this leads to a much larger hydrogen bond distance (3.21 vs 2.74 Å). The presence of the flipped Gln63 residue is thus sufficient to cause a substantial red shift, irrespective of the fact that the other active-state residues retain their basic orientation.

The QM/MM optimized geometries for models W104$_{out}$ and W104$_{out}$M106$_{in}$ are very similar to each other, and the same applies to the calculated excitation energies (2.68 vs 2.66 eV). We note that the C4═O4 group of the chromophore is engaged in a weak hydrogen-bonding interaction with Gln63 in both models, see R(OE1(Q63)−O4(FMN)), which is absent in the QM/MM optimized 1YRX and W104F structures (see Table 2). In models W104$_{out}$ and W104$_{out}$M106$_{in}$, there is no hydrogen bond between OE1(Q63) and the protein β5 strand (residues 101−109). During our standard setup procedure, we find in both cases that solvent water molecules move in and form hydrogen bonds with OE1(Q63). Since water molecules are mobile, this kind of interaction is less directional, compared with the making and breaking of distinct hydrogen bonds to the protein, which can be mechanistically relevant. To further explore these models, we selected two extra starting geometries from MD snapshots in each case and optimized them at the

dx.doi.org/10.1021/jp3028758 | J. Phys. Chem. B 2012, 116, 8064−8073

QM/MM level (see Supporting Information for the protocol to generate these snapshots). The resulting structures show some variation, but retain most of the features of the parent models, and the calculated excitation energies are again on the low side, between 2.63 and 2.65 eV (see Table S10 of Supporting Information for detailed results).

The QM/MM optimized structure for the 2IYG model (Table 3) shares most of the general features with the corresponding structure from quantum refinement (Table 1), but there are also some distinct differences. In the QM/MM structure, the C4═O4 group of the chromophore is mainly involved in the hydrogen bond with Asn45, with a substantially shorter ND2−O4 distance than in the quantum refinement structure (2.94 vs 3.17 Å), while the hydrogen bond with Gln63 is lost as judged from the NE2−O4 distance (3.37 vs 2.81 Å). Moreover, the C2═O2 moiety of the chromophore interacts more strongly with His44 in the QM/MM structure (ND1−O2 distance of 2.83 vs 3.31 Å), albeit via a rather bent ND1−H···O2 angle of 134°.

The properties of the models included in Table 3 are generally in line with the experimental data associated with the signaling state, in particular with regard to the absorption energy, the downshift of the C4═O4 vibrational frequency, and the presence of a strong hydrogen bond between Tyr21 and Gln63.

In summary, we suggest that the Gln63 residue adopts in the dark state of the wild-type AppA protein the orientation proposed in the published 1YRX structure (and not the flipped form found in the AppA1-124C20S mutant). In the preferred dark-state structure, Gln63 is involved in four hydrogen bonds: the amide nitrogen atom (NE2) interacts with the chromophore (N5) and Tyr21 (OH), while the carbonyl oxygen atom (OE1) also interacts with the chromophore (C4═O4) and with Trp104 (NE1). The orientation of Gln63 and its hydrogen bonds with Tyr21 and with the β5 strand of the protein (Trp104 in wild-type AppA) play crucial roles in spectral tuning. Structures with a flipped Gln63 residue have red-shifted absorption energies and a different hydrogen bonding pattern: here, the amide nitrogen atom (NE2) interacts well with the chromophore (C4═O4) and only weakly with Trp104 (NE1), while the carbonyl oxygen atom (OE1) interacts strongly with Tyr21 (OH) and less so with the chromophore (N5). In early experimental work,[7,10] such structures have been associated with the signaling state, whereas recent theoretical work favors the presence of a Gln63 imidic acid isomer in the signaling state.[21,22] The distances between the flavin chromophore and the Asn45 and His44 residues vary considerably in the structures currently obtained by quantum refinement and QM/MM geometry optimization, without any obvious correlation to the calculated excitation energies, indicating that the influence of Asn45 and His44 on the spectra is quite limited.

DFT/MRCI. The results from single-point DFT/MRCI calculations at the structures obtained from the PDB database, from quantum refinement, and from QM/MM geometry optimization are collected in Table 4. They include the vertical excitation energies, the oscillator strengths, and the characterization of the two lowest intense singlet transitions. For most of the structures considered, DFT/MRCI predicts a few (one to three) low-intensity transitions (with oscillator strengths typically below 0.01) between the first and second intense band, which are of mixed character and involve substantial charge transfer (CT) from active-site residues to FMN; these weak absorptions are disregarded in the following.

Generally speaking, the two lowest intense singlet transitions correspond to local excitations (LEs) of $\pi\pi^*$ type on the FMN chromophore. There is some minor admixture of CT excitations from active-site residues to FMN in three cases, i.e., for both PDB structures (1YRX, second transition, CT from Tyr21 and Gln63; 2IYG, first transition, CT from Tyr21) and for one quantum refinement structure (1YRX, second transition, CT from Gln63), but in all these cases, the $\pi\pi^*$ LE character is dominant. The lowest-energy transition usually corresponds to a LE within FMN from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO), but in a few cases, the HOMO is mainly located on Tyr21 so that the LE then starts from HOMO-1 (see Table 4). The second intense transition is generally also of $\pi\pi^*$ LE type and occurs within the FMN chromophore; it is dominated by a single excitation, normally from HOMO-3 to LUMO (see Table 4), which is accompanied by some electron density transfer mainly from the phenyl ring toward the other half of the FMN ring system. The three orbitals that are mainly involved in these two transitions (HOMO-3, HOMO, and LUMO) are plotted in the Supporting Information (Figure S2) for the 2IYG-based quantum refinement structure.

Experimentally, the two lowest intense bands of the dark state of the AppA BLUF protein are observed at 445 nm (2.79 eV) and 365 nm (3.40 eV), with the former band being more intense than the latter one.[3] The computed vertical excitation energies for the original PDB structures (1YRX, 2.37 and 3.27 eV; 2IYG, 2.16 and 3.06 eV) are obviously much too low, with deviations from experiment that are beyond the usual DFT/MRCI error margin. By contrast, the DFT/MRCI results for the 1YRX-based quantum refinement structure agree well with experiment, both for the energies (2.81 and 3.49 eV) and for the relative intensities (0.30 vs 0.18). In the case of the 2IYG-based quantum refinement structure, the overall agreement with experiment is almost as good for the energies (2.69 and 3.35 eV), but less so for the relative intensities (0.27 vs 0.28). Comparison with other theoretical work[21−23] is difficult because of the use of different models and geometries. Therefore, we only quote the published vertical excitation energies for the local $\pi\pi^*$ excitations within the FMN chromophore that are considered the "best" by the respective authors, i.e., the CC2 value of 2.99 eV for the first transition in the BlrB protein,[21] the MCQDPT2-CASSCF values of 2.80 and 3.43 eV for an AppA BLUF model cluster (Trp_{out}),[22] and the TD-B3LYP values of 2.75 and 3.46 eV for another AppA BLUF model cluster (average over 114 MD snapshots).[23]

The question remains whether we can attribute the computed spectral differences to specific structural changes. Large shifts in the calculated lowest excitation energies are found when going from the original PDB structures to the quantum refinement structures. We first address the blue shift of 0.44 eV observed in the case of 1YRX (2.37 vs 2.81 eV; see above). For both geometries, the lowest excited singlet state is dominated by a HOMO−LUMO $\pi\pi^*$ single excitation, with a weight of more than 80% in the DFT/MRCI wave function. It is thus not surprising that the major part of the calculated blue shift arises from a corresponding increase in the HOMO−LUMO orbital energy gap: this raise amounts to 0.41 (0.35) eV in single-point DFT calculations for the QM region without (with) the surrounding MM point charges. Further single-point

dx.doi.org/10.1021/jp3028758 | J. Phys. Chem. B 2012, 116, 8064−8073

calculations for selected parts of the QM region show that almost the entire increase in the HOMO−LUMO gap comes from the FMN chromophore itself (0.40 eV) and that the active-site residues (Tyr21, Gln63, Asn45) modulate this increase only slightly in the range between 0.38−0.45 eV (see Table S11 of Supporting Information). It is thus the geometry of the FMN chromophore that makes the difference: judging from the agreement with the experimental spectra (see above), the quantum refinement yields a much more realistic structure of the FMN chromophore than the original X-ray refinement. Inspection of the FMN geometries (see Figure S3 of the Supporting Information) shows that the bond lengths in the flavin rings are indeed quite different in the structures obtained from crystallographic and quantum refinement: these bond lengths are rather uniform in the former case (CC 1.39−1.42 Å, CN 1.37−1.40 Å) and span a much larger range in the latter case (CC 1.38−1.48 Å, CN 1.31−1.43 Å). The more pronounced bond length alternation in the quantum refinement structure causes a larger HOMO−LUMO gap, and hence a larger vertical excitation energy that is much closer to the experimental value. The situation is entirely analogous for the blue shift encountered in the 2IYG-based structures (see Supporting Information). Our present findings thus provide further support for the recommendation that the calculation of spectral properties of proteins should be based on reoptimized chromophore structures (rather than X-ray PDB structures) to avoid artifacts arising from inaccurate geometries.[69]

Compared with the changes caused by the bond length alternation in the FMN chromophore, the effects of the neighboring residues on the HOMO−LUMO gap are much smaller, generally less than 0.1 eV in either direction (see Table S11 of Supporting Information). The effects of Tyr21, Gln63, and Asn45 appear to be additive, yielding a reduction of the HOMO−LUMO gap of 0.09−0.10 eV for the 1YRX-based structures and of 0.15−0.19 eV for the 2IYG-based structures (Table S11) and thus contributing to the generally somewhat lower excitation energies computed in the latter case. However, since these changes are very small, we refrain from an attempt to link them to specific structural features or specific chromophore-residue interactions.

CONCLUSION

DFT/MRCI calculations of vertical excitation energies of photoreceptors can be used to assess the quality of the underlying protein structure. In the case of the wild-type AppA protein, they strongly underestimate the experimental absorption energies when evaluated at the published X-ray geometries (1YRX and 2IYG). Therefore we performed quantum refinements starting from the published 1YRX and 2IYG structures, employing the X-ray raw data (PDB) and restraints from QM/MM rather than MM energies. Single-point DFT/MRCI calculations at the resulting refined structures gave excitation energies close to the experimental value, within the typical error bar of DFT/MRCI; the agreement is slightly better for the 1YRX-based structure.

For further analysis, we carried out pure QM/MM geometry optimizations starting from the 1YRX and 2IYG structures obtained by quantum refinement and, in addition, from five manually modified model structures with judicious changes in the orientation of crucial active-site residues. DFT/MRCI calculations at these QM/MM optimized structures again afforded excitation energies reasonably close to the experimental value, with one group of three structures reproducing the experiment almost exactly, and another group of four structures having somewhat lower excitation energies (red-shifted by about 0.1 eV). Experimentally, the absorption of the signaling state is red-shifted with respect to the dark state, by about 15 nm, and it is thus tempting to associate the common structural features of the first (second) group with the dark (signaling) state of AppA BLUF proteins.

In the first group, the Gln63 residue is oriented as in the published 1YRX crystal structure: it is hydrogen bonded via its carbonyl oxygen atom OE1 to the β5 strand of the protein (to Trp104 in wild-type AppA) and via its amide nitrogen atom NE2 to Tyr21 (less strongly); in addition, there are hydrogen-bonding interactions of Gln63 with the chromophore (N5 and partly also C4═O4). The models of the first group have the side chain of either Trp104 or Met106 pointing inside toward the active side (1YRX and 2-Q63$_{flip}$); since both arrangements lead to high absorption energies, we are not able to distinguish on this basis whether the side chain of Trp104 or Met106 is preferred in the active site of the AppA dark state. Another common feature in the first group is the hydrogen bond between Gln63 and the β5 strand of the protein, but the corresponding partner (Trp104 in wild-type AppA) seems to have little influence on the computed absorption energy (regardless of whether it points inside the active site or not).

The second group of structures with lower excitation energies contains Gln63 in a flipped 2IYG-type orientation: as a consequence, Gln63 interacts only weakly with the β5 strand (via NE2), but forms a strong hydrogen bond with Tyr21 (via OE1); in addition, there are two hydrogen-bonding interactions with the chromophore (C4═O4 and N5). Judging from the computed short distances, the carbonyl oxygen atom (O4) of the chromophore interacts rather strongly with Asn45 (ND2) and Gln63 (NE2 and OE1) in the 2IYG-type models, and one may thus expect that this will cause a reduction of the C4═O4 stretching frequency, as has indeed been observed experimentally in the signaling state.

The structures obtained from quantum refinement and from QM/MM geometry optimization share the same general features. The changes relative to the structures determined by conventional X-ray crystallographic refinement (PDB: 1YRX and 2IYG) are somewhat more pronounced in the case of the purely theoretical QM/MM optimizations, which lack the experimental restraints incorporated during quantum refinement and which may thus overestimate the required corrections. Overall, we thus regard the structures from quantum refinement as the most reliable ones. For our present purposes, the differences between them and the QM/MM optimized structures are rather minor, however, since they both yield consistent sets of DFT/MRCI excitation energies.

ASSOCIATED CONTENT

Supporting Information

Cartesian coordinates of the QM atoms in refined and optimized structures, technical details of the protocol for generating additional MD snapshots and associated geometrical data, comments on the applied dispersion corrections, HOMO−LUMO gaps, graphical representations of the models used for analysis, plots of the MOs involved in the two lowest intense electronic transitions for the 1YRX(w_{xref} = 1) structure, and comparisons of chromophore structures. This material is available free of charge via the Internet at http://pubs.acs.org.

dx.doi.org/10.1021/jp3028758 | J. Phys. Chem. B 2012, 116, 8064−8073

AUTHOR INFORMATION

Corresponding Author

*E-mail: thiel@mpi-muelheim.mpg.de.

Present Address

‡AstraZeneca R&D Södertälje, SE-151 85 Södertälje, Sweden.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Max Planck Society and the Volkswagenstiftung. We are grateful to Wilfred van Gunsteren and Katharina Meier (ETH Zürich) for helpful discussions and comments.

REFERENCES

(1) van der Horst, M. A.; Hellingwerf, K. J. *Acc. Chem. Res.* **2004**, *37*, 13.
(2) Masuda, S.; Bauer, C. *Cell* **2002**, *110*, 613.
(3) Kraft, B. J.; Masuda, S.; Kikuchi, J.; Dragnea, V.; Tollin, G.; Zaleski, J. M.; Bauer, C. E. *Biochemistry* **2003**, *42*, 6726.
(4) Masuda, S.; Hasegawa; Ono, T.-A. *Biochemistry* **2005**, *44*, 1215.
(5) Unno, M.; Sano, R.; Masuda, S.; Ono, T.-A.; Yamauchi, S. *J. Phys. Chem. B* **2005**, *109*, 12620.
(6) Gomelsky, M.; Kaplan, S. *J. Biol. Chem.* **1998**, *273*, 35319.
(7) Anderson, S.; Dragnea, V.; Masuda, S.; Ybe, J.; Moffat, K.; Bauer, C. *Biochemistry* **2005**, *44*, 7998.
(8) Jung, A.; Reinstein, J.; Domratcheva, T.; Shoeman, R. L.; Schlichting, I. *J. Mol. Biol.* **2006**, *362*, 717.
(9) Jung, A.; Domratcheva, T.; Tarutina, M.; Wu, Q.; Ko, W. H.; Shoeman, R. L.; Gomelsky, M.; Gardner, K. H.; Schlichting, I. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12350.
(10) Grinstead, J. S.; Avila-Perez, M.; Hellingwerf, K. J.; Boelens, R.; Kaptein, R. *J. Am. Chem. Soc.* **2006**, *128*, 15066.
(11) Toh, K. C.; van Stokkum, I. H. M.; Hendriks, J.; Alexandre, M. T. A.; Arents, J. C.; Avila-Perez, M.; van Grondelle, R.; Hellingwerf, K. J.; Kennis, J. T. M. *Biophys. J.* **2008**, *95*, 312.
(12) Unno, M.; Kikuchi, S.; Masuda, S. *Biophys. J.* **2010**, *98*, 1949.
(13) Mathes, T.; van Stokkum, I. H. M.; Bonetti, C.; Hegemann, P.; Kennis, J. T. M. *J. Phys. Chem. B* **2011**, *115*, 7963.
(14) Yuan, H.; Anderson, S.; Masuda, S.; Dragnea, V.; Moffat, K.; Bauer, C. *Biochemistry* **2006**, *45*, 12687.
(15) Masuda, S.; Tomida, Y.; Ohta, H.; Takamiya, K. *J. Mol. Biol.* **2007**, *368*, 1223.
(16) Dragnea, V.; Waegele, M.; Balascuta, S.; Bauer, C.; Dragnea, B. *Biochemistry* **2005**, *44*, 15978.
(17) Hasegawa, K.; Masuda, S.; Ono, T. *Biochemistry* **2004**, *43*, 14979.
(18) Obanayama, K.; Kobayashi, H.; Fukushima, K.; Sakurai, M. *Photochem. Photobiol.* **2008**, *84*, 1003.
(19) Meier, K.; Thiel, W.; van Gunsteren, W. *J. Comput. Chem.* **2012**, *33*, 363.
(20) Rieff, B.; Bauer, S.; Gerald, M.; Tavan, P. *J. Phys. Chem. B* **2011**, *115*, 11239.
(21) Sadeghian, K.; Bocola, M.; Schütz, M. *J. Am. Chem. Soc.* **2008**, *130*, 12501.
(22) Domratcheva, T.; Grigorenko, B. L.; Schlichting, I.; Nemukhin, A. V. *Biophys. J.* **2008**, *94*, 3872.
(23) Götze, J.; Saalfrank, P. *J. Photochem. Photobiol. B* **2009**, *94*, 87.
(24) Ryde, U.; Olsen, L.; Nilsson, K. *J. Comput. Chem.* **2002**, *23*, 1058.
(25) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198.
(26) Grimme, S.; Waletzke, M. *J. Chem. Phys.* **1999**, *111*, 5645.
(27) Sanchez-Garcia, E.; Doerr, M.; Hsiao, Y.-W.; Thiel, W. *J. Phys. Chem. B* **2009**, *113*, 16622.
(28) Hsiao, Y.-W.; Sanchez-Garcia, E.; Doerr, M.; Thiel, W. *J. Phys. Chem. B* **2010**, *114*, 15413.
(29) Hsiao, Y.-W.; Drakenberg, T.; Ryde, U. *J. Biomol. NMR* **2005**, *31*, 97.
(30) Hsiao, Y.-W.; Tao, Y.; Shokes, J. E.; Scott, R. A.; Ryde, U. *Phys. Rev. B* **2006**, *74*, 214101.
(31) Ryde, U.; Hsiao, Y.-W.; Rulíšek, L.; Solomon, E. I. *J. Am. Chem. Soc.* **2007**, *129*, 726.
(32) Hsiao, Y.-W.; Ryde, U. *Inorg. Chim. Acta* **2005**, *359*, 1081.
(33) Hersleth, H.-P.; Hsiao, Y.-W.; Ryde, U.; Görbitz, C. H.; Andersson, K. K. *Biochem. J.* **2008**, *412*, 257.
(34) Yu, N.; Yennawar, H. P.; Merz, K. M. *Acta Crystallogr. D* **2005**, *61*, 322.
(35) Yu, N.; Hayik, S. A.; Wang, B.; Liao, N.; Reynolds, C. H.; Merz, K. M. *J. Chem. Theory Comput.* **2006**, *2*, 1057−1069.
(36) Li, X.; He, X.; Wang, B.; Merz, K. *J. Am. Chem. Soc.* **2009**, *131*, 7742−7754.
(37) Li, X.; Hayik, S. A.; Merz, K. M. *J. Inorg. Biochem.* **2010**, *104*, 512−522.
(38) Hsiao, Y.-W.; Thiel, W. *J. Phys. Chem. B* **2011**, *115*, 2097.
(39) The Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank; http://www.pdb.org (accessed June 11, 2012).
(40) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(41) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
(42) Smith, W.; Forester, T. R. *J. Mol. Graphics* **1996**, *14*, 136.
(43) Sherwood, P.; et al. *THEOCHEM* **2003**, *632*, 1.
(44) ChemShell, A Computational Chemistry Shell; http://www.chemshell.org (accessed June 11, 2012).
(45) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177.
(46) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(47) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.
(48) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.
(49) Hariharan, P. C.; Pople, J. A. *Theor. Chem. Acc.* **1973**, *28*, 213.
(50) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.
(51) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.
(52) MacKerell, A. D. J.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586.
(53) Brunger, A. T.; Adams, P. D.; Clore, G. M.; Delano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, I.; Niges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., *Crystallography & NMR System CNS*, version 1.2; Yale University: New Haven, CT, 2000.
(54) Brunger, A. T.; Adams, P. D. *Acc. Chem. Res.* **2002**, *35*, 404.
(55) Brunger, A. T. *Nature* **1992**, *355*, 472.
(56) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(57) Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P. A.; Thiel, W. *J. Chem. Phys.* **2008**, *128*, 134110.
(58) Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. *J. Chem. Phys.* **2008**, *129*, 104103.
(59) Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. *J. Chem. Phys.* **2010**, *133*, 174318.
(60) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(61) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
(62) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.
(63) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.
(64) Gerenkamp, M. Entwicklung und Anwendung quantenchemischer Methoden zur Berechnung komplexer chemischer Systeme. Ph.D. Thesis, Universität Münster, Münster, Germany, 2005.
(65) Wahl, M.; Sundaralingam, M. *Trends Biochem. Sci.* **1997**, *22*, 97.
(66) Pierce, A.; Sandretto, K.; Bemis, G. *Proteins* **2002**, *49*, 567.
(67) Klaholz, B.; Moras, D. *Structure* **2002**, *10*, 1197.
(68) Sarkhel, S.; Desiraju, G. R. *Proteins* **2004**, *54*, 247.
(69) Dreuw, A.; Harbach, P. H. P.; Mewes, J. M.; Wormit, M. *Theor. Chem. Acc.* **2010**, *125*, 419.