JCTC Journal of Chemical Theory and Computation

# Quantum Chemistry Calculations Provide Support to the Mechanism of the Light-Induced Structural Changes in the Flavin-Binding Photoreceptor Proteins

M. G. Khrenova,*[,†] A. V. Nemukhin,[†,‡] B. L. Grigorenko,[†] A. I. Krylov,[§] and T. M. Domratcheva[||]

*Chemistry Department, M.V. Lomonosov Moscow State University, Leninskie Gory 1/3, Moscow, 119991, Russian Federation, N.M. Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Kosygina 4, Moscow, 119334, Russian Federation, Department of Chemistry, University of Southern California, Los Angeles, California 90089, and Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, 69120 Heidelberg, Germany*

Received April 5, 2010

**Abstract:** The proposed mechanisms of photoinduced reactions in the blue light using flavin chromophore photoreceptor proteins are primarily based on the results of X-ray crystallography and spectroscopy studies. Of particular value are the observed band shifts in optical and vibrational spectra upon formation of the signaling (light-induced) state. However, the same set of experimental data has given rise to contradictory interpretations suggesting different structures of the dark and signaling states. To verify the specific mechanism of light-induced changes involving the rotation/tautomerization transformations with the conserved Gln residue near the flavin chromophore, we performed accurate quantum chemical calculations of the equilibrium structures, vibrational and absorption bands of the model systems mimicking the BLUF domain of flavoprotein AppA. Geometry optimization and calculations of vibrational frequencies were carried out with the QM(B3LYP/cc-pVDZ)/MM(AMBER) approach starting from the representative molecular dynamics (MD) snapshots. The MD simulations were initiated from the available crystal structures of the AppA protein. Calculations of the vertical excitation energies were performed with the scaled opposite spin configuration interaction with single substitutions SOS-CIS(D) method that enables efficient treatment of excited states in large molecular systems. The computed molecular structures as well as the spectral shifts (the red shift by 12÷16 nm in absorption and the downshift by 25 $cm^{-1}$ for the C4═O flavin vibrational mode) are in excellent agreement with the experimental results, lending a strong support to the mechanism proposed by Domratcheva et al. (*Biophys. J.* **2008,** *94*, 3872).

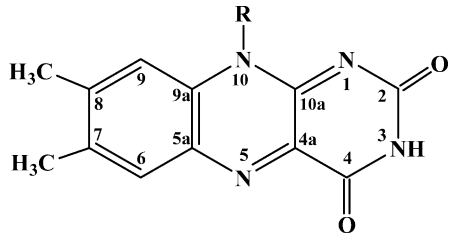
## Introduction

Accurate characterization of electronically excited states and vibrational spectra in large molecular systems is a key step in establishing mechanisms of light-induced biophysical processes. Studies of the blue light using flavin adenine dinucleotide (FAD) sensor proteins called BLUF[1] provide prominent examples. The BLUF domains belong to the family of photoreceptor proteins whose function is linked to the subtle changes in the chromophore environment.[2] The primary photochemistry of their activation results in changes in the hydrogen-bond network near the chromophore and presumably in conformational changes involving the nearby

* Corresponding author. Telephone: +74959391096. Fax: +74959390283. E-mail: wasabiko13@gmail.com.
† M.V. Lomonosov Moscow State University.
‡ Russian Academy of Sciences.
§ University of Southern California.
|| Max Planck Institute for Medical Research.

10.1021/ct100179p 
Published on Web 07/01/2010

***Scheme 1.*** The Isoalloxazine Ring of FAD and Numbering of Its Atoms




amino acid residues. Upon light excitation the signaling intermediate state of the BLUF photocycle is formed that is characterized by a slight red shift (10÷15 nm) of the intense absorption bands in UV−vis spectra. The light-induced red shifts decay within minutes, and the dark state is recovered. The detailed mechanism of these events at the atomic resolution is still actively debated. All interpretations agree that there is a certain reconstruction of the hydrogen-bond network in the vicinity of the isoalloxazine ring of FAD (Scheme 1), however, a considerable disagreement on the light-induced specific changes in the peptide groups, especially near the N5, C4═O, N3H, and C2═O moieties of the chromophore, persists in the literature.[3−15]

The main focus of this work is on the BLUF domain containing protein AppA whose function is to control photosynthesis gene expression in the purple bacterium *Rhodobacter sphaeroides*.[3] Among other BLUF proteins, AppA is distinguished by the longest recovery time to the dark state. The absorption spectrum of the protein in the dark state consists of the broad peak at 443 nm attributed to the $S_0-S_1$ electronic transition of flavin buried in the chromophore-containing pocket, while in the light-induced transient form, this band is shifted to 456 nm, i.e., by 13 nm (0.08 eV).[4] The definite change in the vibrational spectrum upon formation of the signaling state is a red shift of about 20−23 $cm^{-1}$ in the flavin C4═O stretching vibration.[4,16]

Such small shifts in the optical and vibrational bands are attributed to rearrangements in the immediate vicinity of the chromophore, but this is the only consensus of opinions among various research groups. The three-dimensional (3D) structures of the protein derived from the X-ray studies[6,17] have not resolved the disagreement but actually initiated the dispute. Figure 1 illustrates the main point of the debate. The molecular groups in the chromophore-containing pocket are from the side chains of Tyr21, Gln63, His44, and Asn45, as shown in Figure 1. The contradictory opinions concern the side chain of Gln63 in the dark state since X-ray crystallography cannot resolve the location of hydrogen atoms and unambiguously distinguish oxygen and nitrogen atoms. Consequently, different forms of Gln63 (in particular, amidic and imidic) and different orientations of its side chain toward Tyr21 have been proposed for the dark and light states of AppA.[6,17] One more amino acid side chain in the vicinity of the isoalloxazine ring of FAD (either Met106 or Trp104) is shown schematically in Figure 1. According to the results of Jung et al.,[17] the side chain of Met106 is located near FAD in the dark state, while Trp104 is located outside this region (the so-called, $Trp_{out}/Met_{in}$ conformation of BLUF). The corresponding crystal structure is described as 2IYG in

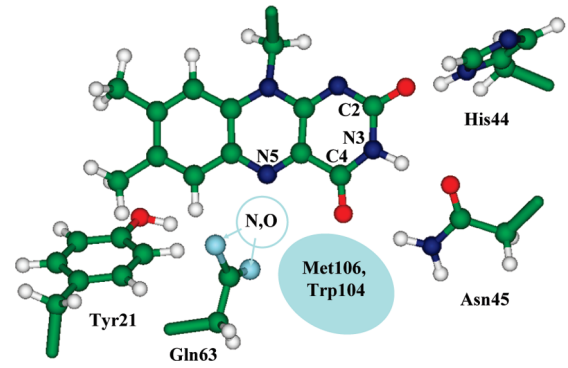


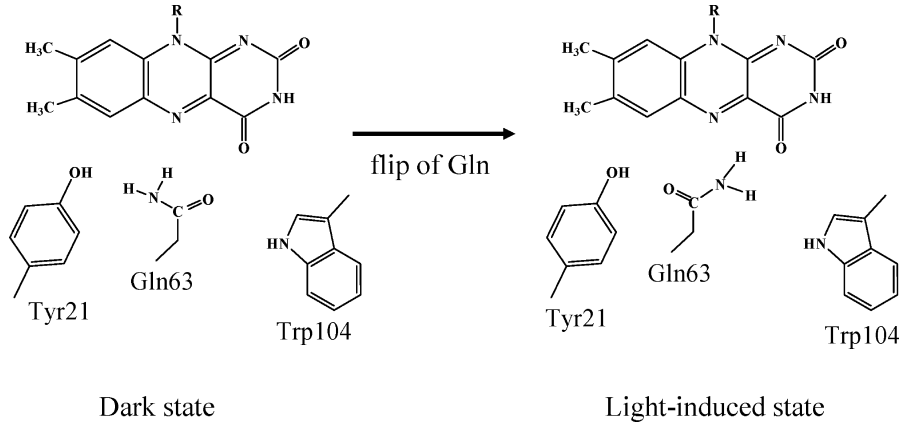

***Figure 1.*** A fragment of the chromophore-containing pocket of AppA BLUF. Here, and below, green, red, and blue colors are reserved for carbon, oxygen, and nitrogen, respectively. The functional head of the side chain of Gln63 is shown schematically as explained in the text. The debated location of either Met106 or Trp104 residues is also schematically indicated.

the PDB archive.[18] In this structure, the side chain of Met106 is hydrogen bonded to Gln63. On the contrary, Anderson et al.[6] assigned the structure (PDB code 1YRX) with the side chain of Trp104 close to the chromophore and with Met106 outside the region ($Trp_{in}/Met_{out}$) to the dark state. Here, Trp104 forms a hydrogen bond to Gln63. Transition from one structure to another apparently requires considerable conformational changes of the protein.

Our recent paper[12] argues in favor of the $Trp_{out}/Met_{in}$ structure (2IYG) for the dark state and proposes that the ($Trp_{in}/Met_{out}$) conformation of AppA corresponding to the 1YRX crystal structure should be assigned to the light-induced state. The calculations and careful analysis of the crystal structures indicated that the tautomeric imidic form of Gln63 is present in the light-induced state. The conclusions were supported by quantum chemical calculations of the structures and electronic and vibrational spectra for a series of clusters (i.e., the chromophore and the surrounding residues) as well as quantum mechanical/molecular mechanical (QM/MM) calculations with a relatively low-level description of the QM subsystem. Since then, new theoretical calculations were published,[7,8,19] however, some of their conclusions only partly coincided with our proposal. To provide more solid support to the mechanism formulated in ref 12, we performed here new higher-level calculations of the equilibrium geometries, vibrational spectra, and electronic transitions for the model systems mimicking the BLUF domain of the flavoprotein AppA.

From the computational point of view, accounting for such a small difference in excitation energy between the dark and light states (0.08 eV) poses a great challenge. The effect is due to a subtle reconstruction of the hydrogen-bond network near the chromophore, and, apparently, the model molecular clusters for excited states calculations should be large enough. They should include not only the chromophore but also the nearby amino acid side chains, some of which (Tyr, His, and Trp) donate their $\pi$ electrons to the model system. Using time-dependent density functional theory (TD-DFT), which is a tempting choice for excited-state calculations in large systems, is questionable owing to the well-known difficulties

***Scheme 2.*** The Mechanism of Light Activation in AppA by Anderson et al[6]




Dark state Light-induced state

with an unphysical description of the charge transfer (CT) states, which may contaminate the valence excited states.[20,21] The work of Sadeghian et al.[8] vividly demonstrated the problematic behavior of TD-DFT precisely for the systems considered in the present work. The application of the complete active space self-consistent field (CASSCF)-based approaches (multiconfigurational second-order perturbation theory, CASPT2, and multiconfigurational quasi-degenerate perturbation theory, MCQDPT2), which are shown to be accurate and efficient for calculating valence transitions in small- and medium-size organic chromophores,[22−32] may also be problematic because of the difficulties with the selection of active space that yields balanced description of the states of interest in such large model systems.

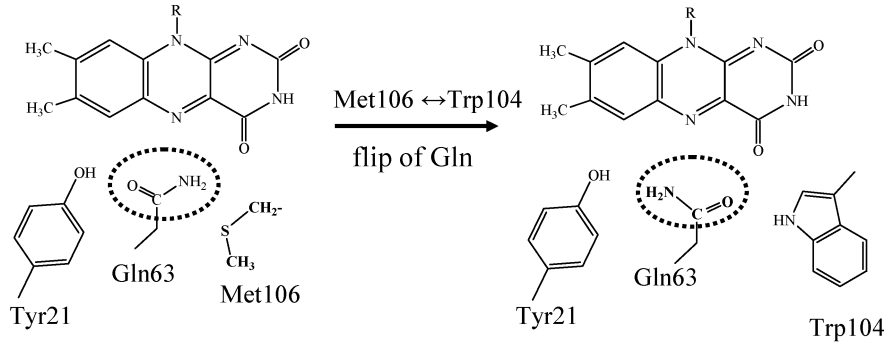
In this context, the scaled opposite-spin configuration interaction with single substitutions SOS-CIS(D) method that has been recently developed[33,34] is an attractive alternative to a more computationally expensive techniques. The approach is closely related to perturbatively corrected configuration interaction singles (CIS), CIS(D),[35] which can be viewed as an approximation of more rigorous equation-of-motion coupled cluster singles and doubles (EOM-CCSD).[36,37] The CIS wave functions are qualitatively correct for the excited states dominated by single-electron excitations including Rydberg, valence, and charge transfer (CT) states as well as interacting states of a mixed character. Due to single-reference formulation and size-intensive properties of the model, the relative energy differences between the excited states and the changes in excitation energies in homological series or isomers are reproduced much better than the absolute values of excitation energies (for which errors often exceed 1 eV). Moreover, the perturbative inclusion of double excitations in CIS(D) yields considerable improvement in accuracy, i.e., the CIS(D) excitation energies for well-behaved cases are very close to that of the EOM-CCSD ones. Overall, CIS(D) is similar to second-order approximate coupled cluster (CC2). The scaling of the method is also $N^5$, although the computational cost is less than that of CC2 owing to the noniterative character of the correction. Recently, CIS(D) was implemented using the SOS variant of second-order perturbation theory[33,34] in which only the opposite-spin correlation energy is computed, and the same-spin correlation is estimated from the opposite-spin one using an empirical scaling factor. For the ground-state calculations, SOS-MP2 demonstrated more robust behavior and higher accuracy than MP2,[38] suggesting a moderate improvement in accuracy for the excited-state variant. Moreover, the SOS implementation along with RI enabled the reduction of computational scaling to $N^4$. Thus, SOS-CIS(D) calculations can be performed for relatively large systems. In our recent studies of the green fluorescent protein, we observed an excellent performance of the SOS-CIS(D) method for excitation energies.[39]

## Review of the Previous Quantum Chemical Calculations

As mentioned above, different mechanisms of the photocycle were proposed based on the results of experimental and computational studies. Several research groups support the assignment of the $Trp_{in}$ conformation originated from the 1YRX crystal structure[6] to the dark state. An initial suggestion of Anderson et al.[6] was that photoactivation might induce the flip of the Gln63 side chain leading to a new hydrogen bond of Gln with the flavin's O4 (Scheme 2). The recent study from this group[15] concluded that light exposure did not cause noticeable change of Trp104 from $Trp_{in}$ to $Trp_{out}$ conformations.

The structural model[6] illustrated in Scheme 2 was used for interpretation of the fast spectroscopic studies[14,40−42] as well as of NMR studies.[9] Along with the results of mutagenesis and spectroscopy studies, Unno et al.[43] presented B3LYP/6-31G** calculations of the series of molecular clusters. The 7,8-dimethyl-10-glycerylisoalloxazine molecule was used as a model of the chromophore. In addition, water, acetamide, and 4-methylimidazole were included in the calculations to reproduce hydrogen bonds with the active-site residues of Gln63, Asn45, and His44. The fragments were arranged based on the crystal structure of AppA BLUF suggested by Anderson et al.[6] and subsequently optimized. The authors calculated the frequencies for the C4═O stretch in flavin and the electronic excitation energies. No convincing support for the dark protein structure denoted here as $Trp_{in}$ have emerged from these calculations. However, the important conclusion from this work[43] is that there is no hydrogen bond between Gln63 and flavin's C4═O in the dark state, and, therefore, the Gln side chain rotates to make a new hydrogen bond with C4═O in the signaling state. We note

***Scheme 3.*** The Mechanism of Light Activation in AppA by Jung et al[17]




that this conclusion may apply to the ($Trp_{out}/Met_{in}$)[17] conformation as well.

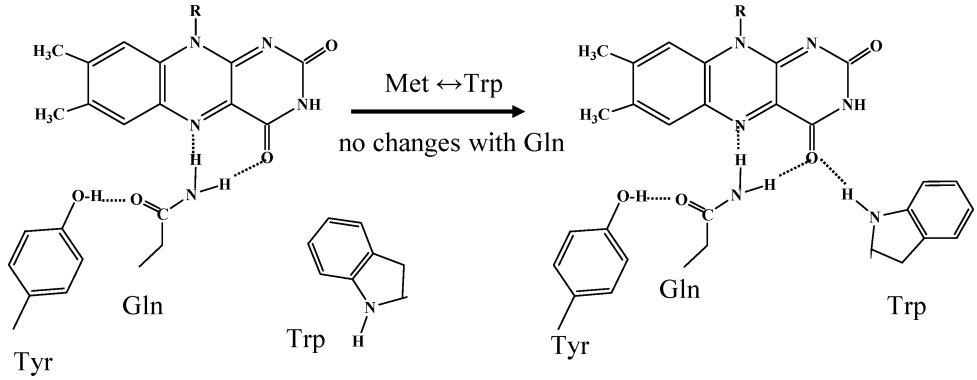
Götze and Saalfrank[19] performed quantum chemical calculations for the cluster composed of the model flavin molecule surrounded by the nearest 10 amino acid side chains (Tyr21, Ser23, Ser41, His44, Asn45, Gln63, Met76, His85, Trp104, and Met106). That was the largest model system (165 atoms) treated at the QM level among all previously considered models.[7,8,12,43] The equilibrium structures were optimized by B3LYP/6-31G* following classical molecular dynamics (MD) simulations initiated from the NMR structure PDBID: 2BUN.[10] Although the latter structure is poorly resolved and cannot be unambiguously assigned to either $Trp_{in}$ or $Trp_{out}$ conformation, the mechanism ultimately advocated by Götze and Saalfrank apparently assumes that the dark state corresponds to the $Trp_{in}$ conformation. The TD-DFT (B3LYP/6-31G*) method was used to calculate excitation energies. The authors applied the following strategy: they calculated the excitation energies for the 20 available NMR structures of AppA BLUF[10] (without MD) and averaged the results. This corresponds to an ensemble consisting of these structures with equal weights. The calculated average UV−vis spectrum has peak maxima at 341 and 450 nm, whereas the experimental spectrum has maxima at 374 and 443 nm for the $S_0 \rightarrow S_2$ and $S_0 \rightarrow S_1$ transitions, respectively. The authors reported a systematic dependence of the computed spectra on the positions of conserved residue Ser41 which, according to their suggestion, is actively involved in the photoactivation.

Now we turn to the supporters of the model in which the dark state is associated with the ($Trp_{out}/Met_{in}$) conformation. Initially, this idea was formulated by Jung et al.[17] who proposed the photoactivation mechanism illustrated in Scheme 3. According to this model, light induces the exchange between Trp104 and Met106 to form the $Trp_{in}$ conformation in the signaling state. Thus, the interaction of Met106 with Gln63 changes supporting a rotation of the side chain of the latter.

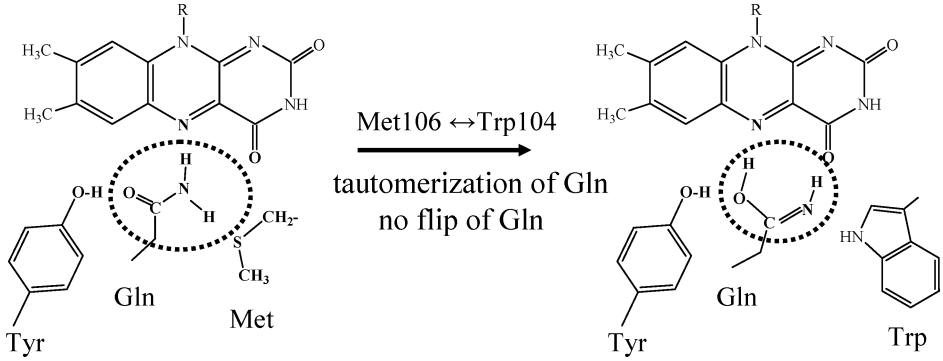
Calculations supporting this model were presented in ref 12 and included QM(RHF/6-31G)/MM(AMBER) optimizations of the equilibrium structures of the model system starting from crystal structures 2IYG and 1YRX. The optimized structure obtained in the calculation initiated using the 2IYG coordinates of heavy atoms agreed well with the parent crystal structure. The minimum energy structure obtained starting from 1YRX was also consistent with the respective crystal structure, however, it suggested the tautomerized form of the Gln63 side chain. This imidic form of Gln63 allows for an additional hydrogen bond of Gln63 with O4 of flavin, thus accounting for the small red shift in absorption maximum of the $S_0 \rightarrow S_1$ transition as well as for the downshift of the C4═O stretch of flavin. The red shift of about 10 nm in absorption was estimated using a relatively low-level approach, QM(CASSCF/6-31)/MM(AMBER). A series of quantum chemical calculations for molecular clusters representing the chromophore-binding pocket yielded qualitatively similar conclusions as the QM/MM simulations. In the cluster calculations, the chromophore was represented by the lumiflavin molecule, and the Tyr21, Gln63, Met106, Trp104 side chains were modeled by phenol, acetamide, dimethylthioether, and pyrrole molecules, respectively. The equilibrium geometry parameters were optimized with B3LYP/6-31G(d). Energies and wave functions of the five lowest singlet states were calculated by CISD in the active space of the 32 highest occupied and 32 lowest virtual Hartree−Fock orbitals. The vertical electron transition energies were obtained using CASSCF/6-31G(d) state averaged over the five lowest roots with the active space chosen based on the CISD solutions. The CASSCF active space consisted of three bonding $\pi$, one nonbonding np, and one antibonding $\pi^*$ molecular orbitals (MOs) localized on the lumiflavin and phenol molecules. The CASSCF energies were corrected by MCQDPT2. The resulting excitation energies, 2.65 (467) and 2.59 eV (477 nm), were in excellent agreement with the experimental results, e.g., the computed red shift was 10 nm, as compared to the experimental value of 13 nm. About 40 $cm^{-1}$ downshift for the C4═O stretch of flavin was obtained.

In another theoretical study, Obanayama et al.[7] also critically analyzed the models shown in Schemes 2[6] and 3.[17] The authors carried out quantum chemical calculations for the cluster composed of the lumiflavin molecule and the truncated side chains of Tyr21, His44, Asn45, Gln63, and Trp104. Geometry was optimized with B3LYP/6-31G** following the MD simulations of the BLUF domains from proteins AppA, BlrB, and Tll0078. The excitation energies were computed using TD-DFT (B3LYP/6-31G**). The calculations supported the model with the $Trp_{out}$ conformation assigned to the dark state and with the $Trp_{in}$ conformation

***Scheme 4.*** The Mechanism of Light Activation in AppA by Obanayama et al[7]




***Scheme 5.*** The Mechanism of Light Activation in BLUF Proteins by Sadeghian et al[8]




to the signaling state. The authors suggested the light-induced transformations illustrated in Scheme 4 assuming that movement of Trp104 from its 'out' to 'in' conformation is sufficient to strengthen the hydrogen bonds between flavin and Gln and Trp to explain the spectroscopic results. However, they failed to reproduce the major features of the signaling state: the computed absorption maximum was 9 nm blue shifted (instead of being red shifted), and the C4═O frequency was blue-shifted by 11 $cm^{-1}$ (instead of being red shifted).

The theoretical work of Sadeghian et al.[8] suggested the mechanism of light activation of BLUF that has common features of the proposal formulated by Domratcheva et al.[12] The simulations were initiated from the crystal structure of another BLUF containing protein BlrB (PDB code 2BYC).[5] Correspondingly, a different numbering system was used for the residues. After the MD relaxation of the model structures, the QM/MM ground-state geometry optimization was performed starting from the representative MD snapshots. The B3LYP and local MP2 methods were employed in the QM subsystem, and the CHARMM force field was used for the MM part. The excited-state potential energy surfaces were described by TD-DFT with various functionals and selected single-point coupled-cluster calculations using the CC2 method. The authors clearly illustrated difficulties in computing excitation energies with the standard functionals (BP, B3LYP, and BHLYP) due to the self-interaction error, which leads to the gross underestimation of the excitation energies of the charge-transfer (CT) states, i.e., Tyr → flavin CT states. Only the relatively expensive CC2 method allowed the authors to obtain the correct ordering of the excited states in the dark state.

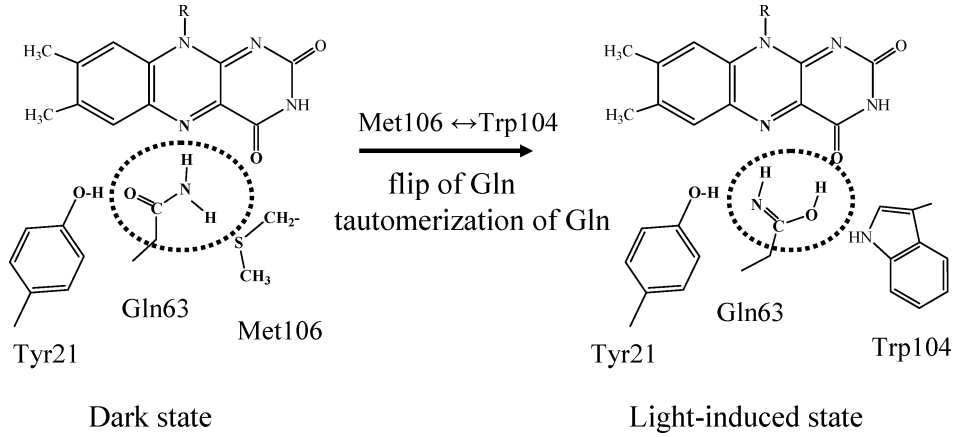
Sadeghian et al.[8] concluded that the ($Trp_{out}$/$Met_{in}$) protein structure should be associated with the dark state. For the light-induced state they suggested a variant of the ($Trp_{in}$/$Met_{out}$) conformation, in which Gln is present in the tautomerized imidic from, but no flipping of the Gln side chain was assumed (Scheme 5). The best values for the shift in the absorption maximum between the dark and signaling states has been obtained at the CC2 level: 10−15 nm to the red, if one ignores the wrong order of the local excitation and charge-transfer states in the assumed light-induced form (Table 3 of ref 8). The agreement with the experimental shift was deemed fortuitous owing to observed strong dependence of the excitation energies on point charges in the MM subsystem. A small red shift (8 $cm^{-1}$) in the computed C4═O vibrational frequency was obtained.

Scheme 6 illustrates the model[12] which receives more rigorous support from high-level quantum calculations described below.

## Calculations and Results

Classical MD trajectory calculations initiated from the corresponding X-ray structures were executed to provide starting points for optimization of the geometry parameters in the QM/MM calculations. By performing such optimization, coordinates of both QM and MM subsystems were included in calculations. Thus obtained equilibrium geometry parameters were used to compute vibrational frequencies. Quantum chemical calculations of the excited electronic states were performed only for the molecular clusters comprising the QM subsystems at the geometries obtained by the QM/MM optimization.

***Scheme 6.*** The Mechanism of Light Activation in AppA by Domratcheva et al[12]




All MD calculations were performed using the NAMD 2.6 program.[44,45] The initial coordinates of the protein and the flavin mononucleotide (FMN) species were generated from crystal structures PDBIDs: 2IYG and 1YRX to model the dark and light states, respectively. The general AMBER force field (GAFF) parameters[46,47] were used for FMN. The CHARMM22 force field[48] for the protein atoms and the TIP3P model parameters for all water molecules were employed. Parameterization of the imidic form of Gln63 was verified on the basis of ab initio calculations. The protein was solvated using TIP3P water in rectangular 61 × 61 × 66 Å$^3$ and 62 × 65 × 74 Å$^3$ boxes in the calculations based on structures 2IYG and 1YRX, respectively. Finally, the systems were neutralized by adding four sodium ions. Periodic boundary conditions were assumed. All long-range electrostatic interactions were computed by using the particle mesh Ewald method.[44] The constant temperature MD simulations were performed for the NVT ensemble at 300 K by using the Langevin thermostat. The simulations were carried out with a 1 fs integration step following the 1500 step energy minimization. No restrictions were imposed on the coordinates of all atoms in the trajectory calculations. The sets of relatively short trajectories (about 2 ns) preceded longer runs of 10 ns. The VMD program[49] was used for the visualization of the results.

The goal of MD simulations was to provide starting coordinates of atoms for optimization of the geometry parameters in the QM/MM calculations. This strategy to employ the MD frames at the beginning and then to proceed to expensive QM/MM optimization is often applied in modeling properties of biomolecular systems, e.g., ref 50. If a MD trajectory does not cover large conformational changes of the protein (as in this application), then selection of a particular frame is not a critical step; the resulting coordinates of QM/MM optimization, at least for the relatively small QM subsystems, are essentially the same for different starting MD frames. As explained below, for our model of the dark state of AppA, we actually obtained four slightly different sets of equilibrium geometry coordinates in QM/MM minimization depending on the selection of MD frames.

The QM/MM geometry optimizations and vibrational analysis were performed within the electronic embedded

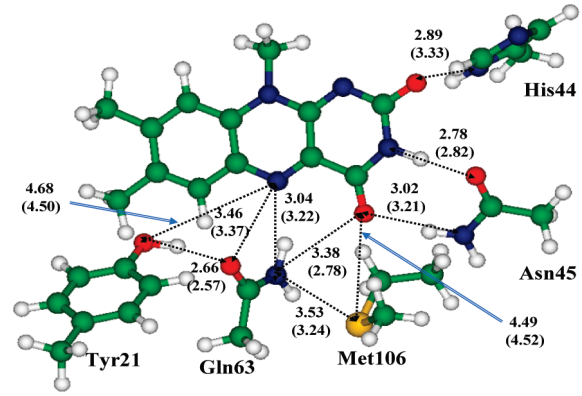



***Figure 2.*** The molecular cluster corresponding to the ($Trp_{out}$/$Met_{in}$) conformation: a model for the dark state. The distances between heavy atoms are given in Å. Top: the computed values and bottom (in parentheses): values from the crystal structure 2IYG.[17]

cluster approximation[51] using the NWChem program package.[52] Energies and forces in the QM subsystem were calculated with B3LYP/cc-pVDZ. The MM part was described with the AMBER force field parameters. The cutoff radius of the MM zone allowed to interact with the quantum region both electrostatically and through the van der Waals interactions was 9 Å. For the presumably dark state originated from the 2IYG crystal structure, the quantum subsystem included the chromophore represented by a lumiflavin molecule and the side chains of Tyr21, His44, Asn45, Gln63, and Met106. For the presumably signaling state originated from the 1YRX crystal structure, the QM part consisted of the chromophore and the side chains of Tyr21, His44, Asn45, Gln63, and Trp104. The remaining protein and solvent water molecules were assigned to the MM part.

Vertical excitation energies were calculated by SOS-CIS(D)/cc-pVDZ using the Q-Chem program.[53] These computations were performed for molecular clusters (Figures 2 and 3) comprising the QM subsystems of the preceding QM/MM calculations at the optimized equilibrium geometries.

The Cartesian coordinates of the systems shown in Figures 2 and 3 are given in the Supporting Information to this paper. In these figures, we compare several key intermolecular distances (more precisely, several distances between the

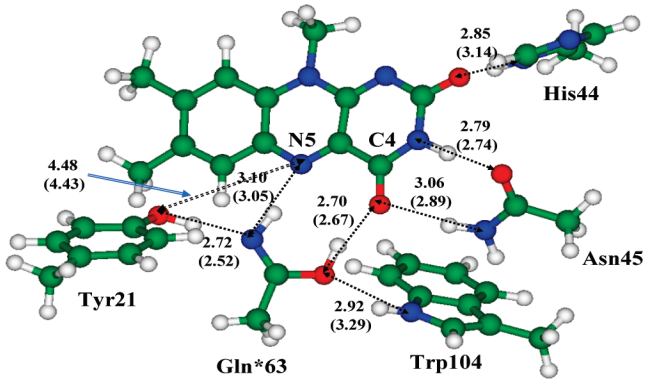



***Figure 3.*** The molecular cluster corresponding to the ($Trp_{in}$/$Met_{out}$) conformation: a model for the light state. The distances between heavy atoms are given in Å. Top: the computed values and bottom (in parentheses): values from the crystal structure 1YRX.[6]

heavy atoms from different molecular groups) computed by B3LYP(cc-pVDZ)/MM(AMBER) with those resolved in crystal structures 2IYG[17] and 1YRX.[6] We note that the agreement between the calculated and measured distances is fairly good. Of a special importance is a good agreement of the theoretical model (Figure 3) that assumes a tautomerized form of Gln63 with the experimental data of Anderson et al.,[6] who considered a normal amidic form of this side chain. We repeat the arguments presented in ref 12 that the original assumption of Anderson et al. on the direct contact of the −C═O group from Gln with the −C4═O group from flavin (see the left side of Scheme 2) contradicts the measured distance of 2.7 A. Such a short distance apparently requires a proton between the corresponding oxygen atoms, as illustrated in Figure 3.

We present in the Supporting Information the results of vibrational analysis using the QM/MM. Here we focus on the prominent band experimentally assigned to the C4═O vibration of flavin (Figure 1). According to ref 4, the FTIR difference spectrum of the AppA BLUF domain showed a sharp band at 1707(−)/1684(+) $cm^{-1}$. Consistent with the FTIR data, a change in the Raman spectrum upon formation of the signaling state is approximately 20 $cm^{-1}$ in the C4═O stretch (observed at 1706 $cm^{-1}$).[16] Therefore, the experimental red shift in the C4═O frequency upon passing from the dark to the light-induced state in AppA BLUF was measured as 20−23 $cm^{-1}$. The computed frequencies for the dark (Figure 2) and signaling (Figure 3) states, 1700 and 1675 $cm^{-1}$, yield a shift of 25 $cm^{-1}$ which is in excellent agreement with the experimental value.

Analysis of MD trajectories for the presumably dark state showed that we could distinguish two conformations of the Met106 side chain with smaller (around 53°) and larger (around 114°) values of the CB−CG−S−CE dihedral angles occurring along the MD run. We also noted oscillations occurring along MD trajectories between two conformations of the His44 side chain with respect to the flavin moiety. Therefore, we took respective four MD snapshots for QM/MM optimization of molecular clusters for which the excitation energies were calculated (Figure 2 illustrates geometry parameters corresponding to the lowest energy). We observed only a slight dependence of the computed excitation energies on the parent MD frame; the respective SOS-CIS(D)/cc-pVDZ values of the $S_0-S_1$ excitation energies with the oscillator strengths 0.66−0.69 were 2.89, 2.92, 2.92, and 2.91 eV. The corresponding wavelengths are 429, 425, 425, and 426 nm. For the presumably light-induced state, the MD trajectories did not show noticeable changes in the configurations in the QM subsystem (Figure 3), hence, we considered the only QM/MM optimized structure for which the excitation energy 2.81 eV (with the oscillator strength 0.63) was obtained corresponding to 441 nm for the wavelength. Therefore, the computed red shift in the absorption band maxima between the dark and light-induced states is 12÷16 nm, to be compared with the 13 nm shift measured for AppA BLUF.[4]

It is important to note that positions of the CT states corresponding to the Tyr → flavin electronic redistribution are much higher in both structures (3.7−3.9 and 3.5 eV for the dark and light-induced forms, respectively), and there were no difficulties in computing the excitation energies for the $S_0-S_1$ transition. It is also worth noting the magnitude of the SOS-CIS(D) corrections to the CIS excitation energies: Without these corrections, the computed $S_0-S_1$ energy gaps are about 1.1 eV higher.

## Discussion and Conclusion

It is instructive to compare the present results with those from the previous simulations, especially with our earlier results of Domratcheva et al. for AppA BLUF[12] and those of the most recent QM/MM calculations of Sadeghian et al.[8] initiated from the crystal structure of the BLUF domain from another flavoprotein BlrB (PDB code 2BYC).

In ref 12, the CASSCF-based MCQDPT2 method for simpler molecular clusters mimicking the dark and light-induced forms of AppA BLUF was employed. The model clusters shown in Figure 3 of ref 12 illustrate the same general trend: The ($Trp_{out}$/$Met_{in}$) conformation of BLUF is characterized by a shorter wavelength and represents the dark state, while the ($Trp_{in}$/$Met_{out}$) conformation exhibits a longer wavelength and represents the light state. The results of the MCQDPT2 calculations agreed well with the experimental data. The theoretical and experimental shifts were 10 and 13 nm, respectively. These calculations involved tremendous efforts including manual inspection of the configuration state functions and the orbitals of the underlying CASSCF solutions.

The CC2 calculations at the QM/MM optimized geometries by Sadeghian et al.[8] also reproduced the red shift of 9 nm in absorption maximum when passing from the dark to light-activated form within their model illustrated in Scheme 5 (ignoring the wrong ordering of local excitation and CT states in the latter form). A very small red shift (8 $cm^{-1}$) in the C4═O vibrational frequency was obtained as well.

Unfortunately, it is difficult to compare the structures obtained in this work with those of Sadeghian et al.[8] since only limited information on the geometries can be extracted from their paper. Considering the dark-state model (Figure 4), we note quite reasonable agreement between the results of the two different theoretical approaches. However, for the only parameter that can be directly compared to the corre-

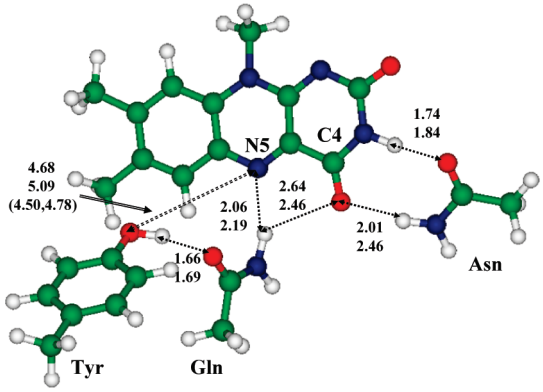



***Figure 4.*** The comparison of selected distances in the dark-state structure obtained in this work (upper values for distances in Å) and by Sadeghian et al.[8] (lower values). For the distance between oxygen from Tyr and N5 from flavin, the bottom values in parentheses refer to the crystal structures 2IYG (AppA) and 2BYC (BlrB).

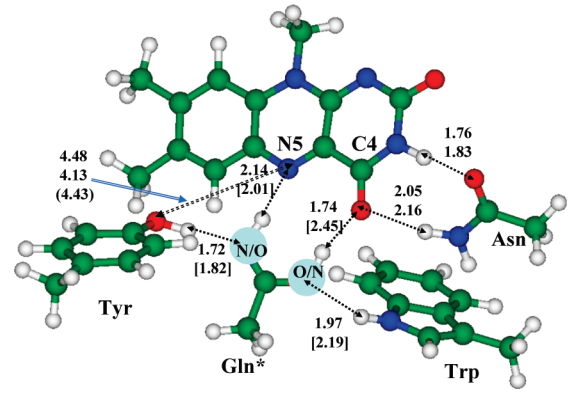



***Figure 5.*** The comparison of selected distances in the light-induced structure obtained in this work (upper values for distances in Å) and by Sadeghian et al.[8] (lower values). For the distance between oxygen from Tyr and N5 from flavin, the bottom value in parentheses refers to the crystal structure 1YRX (AppA). We do not specify O or N atoms in the head of Gln*.

sponding values in the crystals, i.e., the distance between oxygen from Tyr and N5 from flavin, we note a large discrepancy between the computed[8] value (5.09 Å) and those measured in AppA (4.50 Å) and BlrB (4.78 Å). The agreement between our results (4.68 Å) and the crystal data is much better.

The models for the light-induced state suggested by Domratcheva et al.[12] (which is confirmed by this work) and the one suggested by Sadeghian et al.[8] agree that the tautomerized form of Gln is present in this form, however, the two models differ with respect to the orientation of the Gln functional side chain. In our model, nitrogen from Gln participates in hydrogen bonding with Tyr and (Gln*)NH with N5 of flavin, while (Gln*)OH is hydrogen bonded to O4 from flavin. In the model of Sadeghian et al.,[8] oxygen from Gln is involved in hydrogen bonding with Tyr and (Gln*)OH with N5 of flavin, while (Gln*)NH is hydrogen bonded to O4 from flavin. Similarly to the dark state (Figure 4), we note better agreement of our result with the crystal data for the distance between oxygen from Tyr and N5 from flavin; our computed distance of 4.48 Å is closer to the experimental value of 4.43 Å than the distance of 4.13 Å reported by Sadeghian et al.[8] More important is the distance between O4 from flavin to hydrogen-bound carbon from Gln* (1.74 Å in our model) or nitrogen from Gln* (2.45 Å in the model of Sadeghian et al.).[8] Since the result of light activation (production of red-shifted transient form) is believed to be a formation of a new hydrogen bond with O4 from flavin, the distance from O4 to hydrogen is a key issue. The hydrogen-heavy atom distance of 2.45 Å is apparently too long to account for a noticeable (10−15 nm) red shift in the absorption spectrum or a downshift (~20 $cm^{-1}$) in the vibrational spectrum. In fact, the authors reported very small computed vibrational shift of 8 $cm^{-1}$.[8]

A significant conclusion of ref 8 is that the results of standard TD-DFT calculations of the excited states for molecular clusters composed of the flavin chromopore and the surrounding aromatic amino acid residues are not reliable due to the contamination of the valence states by the CT ones. Interestingly, Götze and Saalfrank[19] and Obanayama et al.[7] who applied TD-DFT for such model systems did not report difficulties due to artificial CT states. Obanayama et al.[7] wrote that the obtained shifts in $S_0-S_1$ absorption maximum and the C4═O frequency were of the opposite sign relative to the experimental values.

In this work, we applied the modern tools of molecular modeling including molecular dynamics, QM/MM, and quantum chemistry methods to characterize the two structures of AppA BLUF corresponding to the dark and signaling states. The SOS-CIS(D) results are in excellent agreement with the experimental data. The computational cost of SOS-CIS(D) is slightly higher than that of TDDFT (which is not reliable in this case) but considerably lower than that of the CASSCF-based approaches or CC2. The errors in absolute values of the vertical excitation energy (2.89÷2.92 eV in the calculations for the molecular cluster versus 2.80 eV observed experimentally for the protein in the dark state and 2.81 versus 2.72, respectively, in the light state) are also quite small. The computed band shift of +(12÷16) nm is perfectly consistent with the experimental value of +13 nm. No preliminary work in selecting or rearranging the orbitals for configuration interaction calculations was involved in these calculations. Vibrational calculations performed at the QM/MM level for single experimentally reliable bands corresponding to the C4═O mode also show an excellent agreement with the experiment.

Comparing the present calculations with the previous work,[12] we note that despite large differences in computational protocols, the main qualitative result is the same. According to the calculations, the ($Trp_{out}$/$Met_{in}$) conformation of AppA BLUF should be associated with the dark state and the ($Trp_{in}$/$Met_{out}$) conformation with the light-induced state.

The detailed discussion of the reaction photocycle of AppA BLUF taking into consideration the results of fast spectroscopy studies and molecular modeling will be presented elsewhere. Here we concentrated mostly on quantum chemical aspects of modeling and showed that a careful selection of appropriate computational methods could provide firm validation of the mechanism of photoinduced reactions.

**Acknowledgment.** We thank Prof. Ilme Schlichting for generous help and valuable discussions. This work is supported by the grant from the U.S. Civilian Research and Development Foundation (Project no. RUC1-2914-MO-07) and the Russian Foundation for Basic Research (Project no. 10-03-00139). The Russian team thanks the SKIF-GRID program for providing resources at the SKIF-MSU, SKIF-SIBERIA, and UGATU computational facilities. M.K. acknowledges partial support from the grant MK-64815.2010.4. This work was conducted under the auspices of the iOpen-Shell Center for Computational Studies of Electronic Structure and Spectroscopy of Open-Shell and Electronically Excited Species http://iopenshell.usc.edu supported by the National Science Foundation through the Grant no. CRIF: CRF CHE-0625419_0624602_0625237 as well as through Grant no. CHE-0616271 grant (A.K.).

**Note Added after ASAP Publication.** This article was published ASAP on July 1, 2010. Initials have been added to all author names. The correct version was published on July 15, 2010.

**Supporting Information Available:** Cartesian coordinates of the systems shown in Figures 2 and 3. Results of vibrational analysis using the QM/MM. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Gomelsky, M.; Klug, G. *Trends Biochem. Sci.* **2002**, *27*, 497–500.

(2) van der Horst, M. A.; Hellingwerf, K. J. *Acc. Chem. Res.* **2004**, *37*, 13–20.

(3) Masuda, S.; Bauer, C. E. *Cell* **2002**, *110*, 613–623.

(4) Masuda, S.; Hasegawa, K.; Ono, T. *Biochemistry* **2005**, *44*, 1215–1224.

(5) Jung, A.; Domratcheva, T.; Tarutina, M.; Wu, Q.; Ko, W.-H.; Shoeman, R.; Gomelsky, M.; Gardner, K. H.; Schlichting, I. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12350–12355.

(6) Anderson, S.; Dragnea, V.; Masuda, S.; Ybe, J.; Moffat, K.; Bauer, C. *Biochemistry* **2005**, *44*, 7998–8005.

(7) Obanayama, K.; Kobayashi, H.; Fukushima, K.; Sakurai, M. *Photochem. Photobiol.* **2008**, *84*, 1003–1010.

(8) Sadeghian, K.; Bocola, M.; Schütz, M. *J. Am. Chem. Soc.* **2008**, *130*, 12501–12513.

(9) Grinstead, J. S.; Avila-Perez, M.; Hellingwerf, K. J.; Boelens, R.; Kaptein, R. *J. Am. Chem. Soc.* **2006**, *128*, 15066–15067.

(10) Grinstead, J. S.; Hsu, S. T. D.; Laan, W.; Bonvin, A.; Hellingwerf, K. J.; Boelens, R.; Kaptein, R. *ChemBioChem* **2006**, *7*, 187–193.

(11) Laan, W.; Gauden, M.; Yeremenko, S.; van Grondelle, R.; Kennis, J. T. M.; Hellingwerf, K. J. *Biochemistry* **2006**, *45*, 51–60.

(12) Domratcheva, T.; Grigorenko, B. L.; Schlichting, I.; Nemukhin, A. V. *Biophys. J.* **2008**, *94*, 3872–3879.

(13) Wu, Q.; Gardner, K. H. *Biochemistry* **2009**, *48*, 2620–2629.

(14) Dragnea, V.; Waegele, M.; Balascuta, S.; Bauer, C.; Dragnea, B. *Biochemistry* **2005**, *44*, 15978–15985.

(15) Dragnea, V.; Arunkumar, A. I.; Yuan, H.; Giedroc, D. P.; Bauer, C. E. *Biochemistry* **2009**, *48*, 9969–9979.

(16) Unno, M.; Sano, R.; Masuda, S.; Ono, T.; Yamauchi, S. *J. Phys. Chem. B* **2005**, *109*, 12620–12626.

(17) Jung, A.; Reinstein, J.; Domratcheva, T.; Shoeman, R. L.; Schlichting, I. *J. Mol. Biol.* **2006**, *362*, 717–732.

(18) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(19) Götze, J.; Saalfrank, P. *J. Photochem. Photobiol., B* **2009**, *94*, 87–95.

(20) Dreuw, A.; Weisman, J. L.; Head-Gordon, M. *J. Chem. Phys.* **2003**, *119*, 2943–2946.

(21) Magyar, R. J.; Tretiak, S. *Chem. Theory Comput.* **2007**, *3*, 976–987.

(22) Molina, V.; Merchán, M. *Proc. Nat. Acad. Sci. U.S.A.* **2001**, *98*, 4299–4304.

(23) Martin, M. E.; Negri, F.; Olivucci, M. *J. Am. Chem. Soc.* **2004**, *126*, 5452–5464.

(24) Coto, P. B.; Strambi, A.; Ferré, N.; Olivucci, M. *Proc. Nat.Acad. Sci. U.S.A.* **2006**, *103*, 17154–17159.

(25) Frutos, L. M.; Andruniów, T.; Santoro, F.; Ferré, N.; Olivucci, M. *Proc. Nat. Acad. Sci. U.S.A.* **2007**, *104*, 7764–7769.

(26) Strambi, A.; Coto, P. B.; Frutos, L. M.; Ferre', N.; Olivucci, M. *J. Am. Chem. Soc.* **2008**, *130*, 3382–3388.

(27) Bravaya, K.; Bochenkova, A.; Granovsky, A.; Nemukhin, A. *J. Am. Chem. Soc.* **2007**, *129*, 13035–13042.

(28) Bravaya, K. B.; Bochenkova, A. V.; Granovsky, A. A.; Savitsky, A. P.; Nemukhin, A. V. *J. Phys. Chem.* **2008**, *112*, 8804–8810.

(29) Tokmachev, A. M.; Boggio-Pasqua, M.; Bearpark, M. J.; Robb, M. A. *J. Phys. Chem. A* **2008**, *112*, 10881–10886.

(30) Filippi, C.; Zaccheddu, M.; Buda, F. *J. Chem. Theory Comput.* **2009**, *5*, 2074–2087.

(31) Valsson, O.; Filippi, C. *J. Chem. Theory Comput.* **2010**, *6*, 1275–1292.

(32) Olsen, S.; Smith, S. C. *J. Am. Chem. Soc.* **2008**, *130*, 8677–8689.

(33) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095–9102.

(34) Rhee, Y. M.; Head-Gordon, M. *J. Phys. Chem. A* **2007**, *111*, 5314–5326.

(35) Head-Gordon, M.; Rico, R. J.; Oumi, M.; Lee, T. J. *Chem. Phys. Lett.* **1994**, *219*, 21–29.

(36) Stanton, J. F.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 7029–7039.

(37) Krylov, A. I. *Annu. Rev. Phys. Chem.* **2008**, *59*, 433–462.

(38) DiStacio, R. A.; Head-Gordon, M. *Mol. Phys.* **2007**, *105*, 1073.

(39) Epifanovsky, E.; Polyakov, I.; Grigorenko, B.; Nemukhin, A.; Krylov, A. I. *J. Chem. Theory Comput.* **2009**, *5*, 1895–1906.

(40) Gauden, M.; van Stokkum, I. H.; Key, J. M.; Luhrs, D.; van Grondelle, R.; Hegemann, P.; Kennis, J. T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10895–10900.

(41) Gauden, M.; Grinstead, J. S.; Laan, W.; van Stokkum, I. H. M.; Avila-Perez, M.; Toh, K. C.; Boelens, R.; Kaptein, R.; van Grondelle, R.; Hellingwerf, K. J.; Kennis, J. T. M. *Biochemistry* **2007**, *46*, 7405–7415.

(42) Stelling, A. L.; Ronayne, K. L.; Nappa, J.; Tonge, P. J.; Meech, S. R. *J. Am. Chem. Soc.* **2007**, *129*, 15556–15564.

(43) Unno, M.; Masuda, S.; Ono, T. A.; Yamauchi, S. *J. Am. Chem. Soc.* **2006**, *128*, 5638–5639.

(44) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(45) *NAMD 2.6*; Theoretical and Computational Biophysics Group, University of Illinois at Urbana-Champaign: Urbana, IL, 2006; http://www.ks.uiuc.edu/Research/namd/ (accessed June 2010).

(46) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(47) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics* **2006**, *25*, 247–260.

(48) *CHARMM22*; Harvard University: Cambridge, MA, 1998; http://mackerell.umaryland.edu/CHARMM_ff_params.html (accessed June 2010).

(49) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol.Graphics.* **1996**, *14*, 33–38.

(50) Sanchez-Garcia, E.; Doerr, M.; Hsiao, Y.-W.; Thiel, W. *J. Phys. Chem. B* **2009**, *113*, 16622–16631.

(51) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

(52) Straatsma, T. P.; Aprà, E.; Windus, T. L.; Bylaska, E. J.; de Jong, W.; Hirata, S.; Valiev, M.; Hackler, M.; Pollack, L.; Harrison, R.; Dupuis, M.; Smith, D. M. A.; Nieplocha, J.; Tipparaju, V.; Krishnan, M.; Auer, A. A.; Brown, E.; Cisneros, G.; Fann, G.; Früchtl, H.; Garza, J.; Hirao, K.; Kendall, R.; Nichols, J.; Tsemekhman, K.; Wolinski, K.; Anchell, J.; Bernholdt, D.; Borowski, P.; Clark, T.; Clerc, D.; Dachsel, H.; Deegan, M.; Dyall, K.; Elwood, D.; Glendening, E.; Gutowski, M.; Hess, A.; Jaffe, J.; Johnson, B.; Ju, J.; Kobayashi, R.; Kutteh, R.; Lin, Z.; Littlefield, R.; Long, X.; Meng, B.; Nakajima, T.; Niu, S.; Rosing, M.; Sandrone, G.; Stave, M.; Taylor, H.; Thomas, G.; van Lenthe, J.; Wong, A.; Zhang, Z. *NWChem, A Computational Chemistry Package for Parallel Computers*, version 4.6; Pacific Northwest National Laboratory: Richland, WA, 2004 (accessed June, 2010).

(53) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neil, D. P.; Distasio, R. A., Jr.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Bird, E. F. C.; Daschel, H.; Doerksen, R. J.; Drew, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G. S.; Khalliulin, R. Z.; Klunziger, P.; Lee, A. M.; Liang, W. Z.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock III, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Herhe, W. J.; Schaefer, H. F., III; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

CT100179P