THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

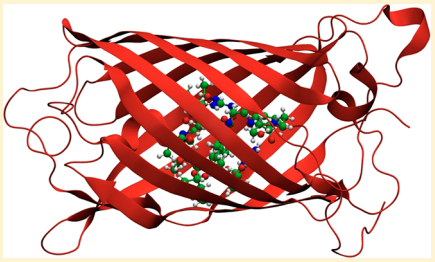
Peek at the Potential Energy Surfaces of the LSSmKate1 and LSSmKate2 Proteins

Carlos Randino,[†] Miquel Moreno,[†] Ricard Gelabert,*[,†] and José M. Lluch[†,‡]

[†]Departament de Química and [‡]Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

Supporting Information

ABSTRACT: To determine the energetic feasibility of the mechanisms involved in the generation of the fluorescent species in red fluorescent proteins LSSmKate1 and LSSmKate2 developed by Piatkevich et al. (*Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 5369−5374 and *J. Am. Chem. Soc.* **2010**, *132*, 10762−10770), a potential energy scan for the respective reaction coordinates was performed in large cluster models including the surroundings of the chromophores, based on the respective crystallographic structures, using DFT and TDDFT. The predicted absorption wavelengths agree to within 5 nm with experiment, thus confirming the accuracy of the calculational level and modeling done. In both proteins, it was found that the adiabatic electronic state with the largest oscillator strength at the Franck−Condon region was not the one from which fluorescence could occur in the products. A diabatization procedure was used to determine an approximate photoactive state, based on selecting the state with the largest oscillator strength throughout. For LSSmKate1, this led to a rather flat potential energy profile but still did not predict a minimum in the product side. It is suggested that relaxation processes, absent from the model, could bring about such a minimum. LSSmKate2, on the other hand, clearly displays a favorable exoergic process in the photoactive state, and its double-proton transfer can be described as concerted but highly asynchronous, involving a barrier in the transfer of the first proton. In this way, the model provides strong support for the mechanism proposed for LSSmKate2.



1. INTRODUCTION

The discovery and characterization of the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* in the 1970s was the beginning of a long series of publications about the original fluorescent protein and its newer variants.[1−4] The fluorescent proteins achieved huge popularity because of their extraordinary application as biofluorescent tags.[5−8] The encoding and gene expression of the amino acid sequence of wild-type green fluorescent protein (wt-GFP) along with the possibility of linking its genetic code to that of any other protein have allowed the monitorization and noninvasive tracking of cells, organelles and proteins in vivo.[9] wt-GFP absorbs strongly at 398 nm and fluoresces at 510 nm, in the green region of the electromagnetic spectrum, with a remarkable quantum yield of 0.8.[10] The GFP chromophore (*p*-hydroxybenzylideneimidazolinone)[11] is formed after an autocatalytic intramolecular cyclization reaction involving three amino acids and is the fundamental part responsible for the photophysical and photochemical properties of the protein. GFP has been extensively studied[12] and it is currently accepted that the fluorescent species is the deprotonated anionic chromophore generated through multiple proton transfers in the excited state promoted by enhanced acidity after the photoexcitation of the chromophore.[13,14] This reaction is extremely fast, with characteristic reaction times of a few picoseconds,[15,16] a characteristic shared by other fluorescent proteins.[17−19] The large quantum yield is linked to the rigidity of the protein, a compact barrel-shaped structure with 11 β-sheets containing the chromophore inside, tightly held in place; thus, it is very difficult for the protein, once photoexcited, to change conformation and relax nonradiatively to the ground state.

Despite its excellent qualities, GFP presents some disadvantages: The green fluorescence emission of GFP can affect surrounding cells, stimulating their own autofluorescence and making the detection of the target protein difficult. Moreover, fluorescent emission in high-energy regions of the electromagnetic spectrum (such as green-blue radiation) can cause damage to neighboring tissues. Consequently, in the past decade, much effort has been made by some research groups to identify natural alternatives or modify (mutate) some residues of the molecule, in order to obtain a set of fluorescent proteins with different features such as new colors, acceptable fluorescence intensity, improved photostability, and changes in the absorption and emission wavelengths.[20] The development of these new proteins has enabled several improvements to in vivo fluorescence imaging of cellular processes such as use of less energetic wavelengths. Among all the different kinds of fluorescent proteins, the red ones have a set of properties that alleviates the principal problems of GFP.[21]

Received: October 21, 2012
Revised: November 19, 2012
Published: November 30, 2012

dx.doi.org/10.1021/jp3104134 | *J. Phys. Chem. B* 2012, 116, 14302−14310

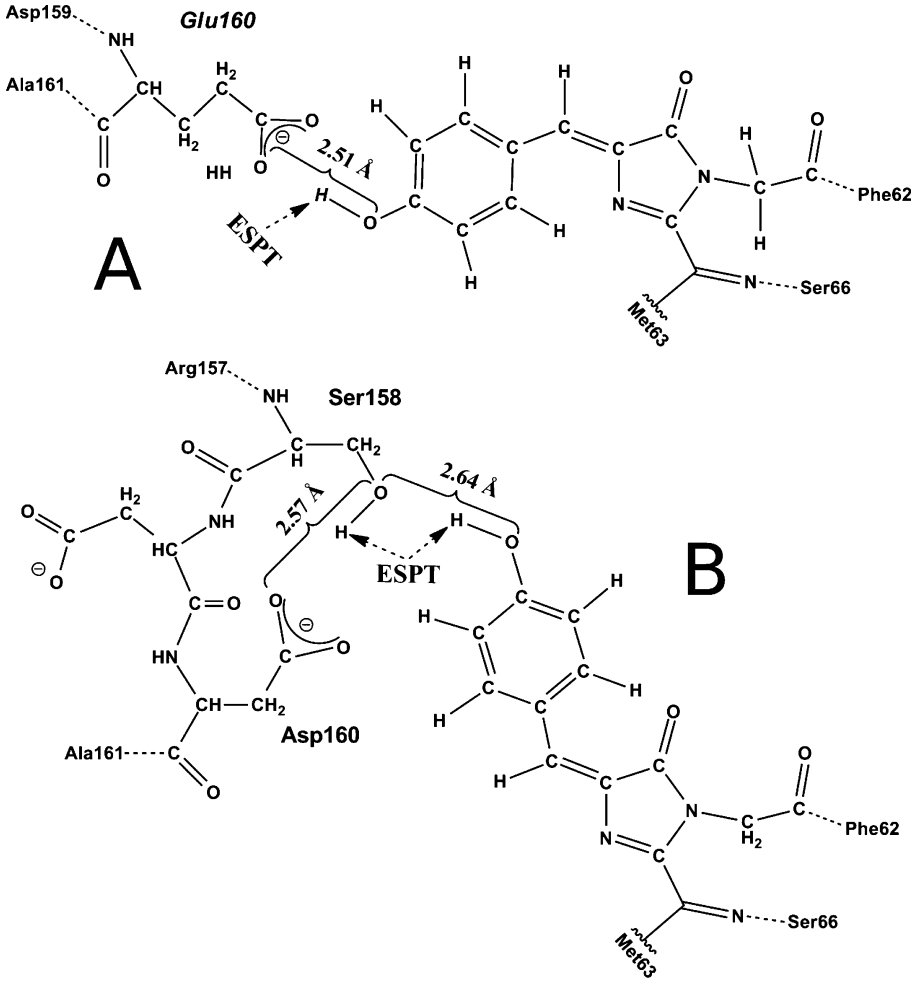


Figure 1. Schematic representation of the chromophores of (A) LSSmKate1 and (B) LSSmKate2 and their corresponding acceptor groups for the one-link and two-link proton wires that allow excited-state proton transfer (ESPT) for LSSmKate1 and LSSmKate2, respectively.

LSSmKate1 and LSSmKate2 are two recently developed far-red fluorescent proteins with a wide separation between the maximum absorption and emission wavelengths, a photochemical feature called a large Stokes shift (LSS). Both proteins were developed by Piatkevich et al.,[22,23] who introduced rational and random mutagenesis into the precursor monomeric far-red protein, mKate. The sequences of LSSmKate1 and LSSmKate2 differ from each other by a single amino acid at the position 160, and the two proteins exhibit absorption peaks at 463 and 460 nm and emission peaks at 624 and 605 nm, respectively. Their quantum yields are smaller than that of GFP, being 0.08 for LSSmKate1 and 0.17 for LSSmKate2. The LSSmKate proteins together with mKeima[19,24,25] form the group of fluorescent proteins with the largest Stokes shifts presently known.

The mechanism of formation of the fluorescent species is still unknown. Because each LSSmKate protein has a rigid structure similar to that of wt-GFP, in terms of both form and size, one could expect a similar mechanism for the LSSmKate proteins. Structural and spectroscopic analyses[22,23] have concluded that the LSSmKate proteins have a hydrogen-bonding network that would allow for proton transfer in the excited state (see Figure 1). LSSmKate1 fluorescence operates through a one-link proton wire from the hydrogen of the hydroxyphenyl group of the chromophore to the side chain of Glu160. LSSmKate2 fluorescence proceeds by means of a two-link proton wire that involves the hydroxyl group of the chromophore and the side chains of Ser158 and Asp160. The suggested photocycle of the LSSmKate proteins thus consists of a two-state model with the neutral and anionic chromophore species being responsible for absorption and emission, respectively.

Because of the recent characterization of LSSmKate1 and LSSmKate2, only one computational study on these red fluorescent proteins has been published.[26] In that work, Zhang et al. attempted a first-principles study of the one- and two-photon absorption (OPA and TPA, respectively) properties of the putative H-bonding complexes for LSSmKate1 and LSSmKate2, as inferred by Piatkevich et al. from X-ray diffraction data from crystallized proteins.[22,23] To this end, density functional theory (DFT) calculations were carried out using the B3LYP hybrid functional with the 6-311++G(d,p) basis set in the ground state to determine equilibrium geometries, and time-dependent DFT (TDDFT) calculations of the involved excited states were performed to compute the OPA properties and TPA cross sections. Their calculations were done on small clusters including only the chromophore and the residues proposed as proton acceptors for the excited-state proton-transfer (ESPT) process. The study was performed in the gas phase and in aqueous solution by means of the polarizable continuum model (PCM). With this theoretical model, the results concerning the computed magnitudes of the OPA and TPA parameters were found to be in reasonable

dx.doi.org/10.1021/jp3104134 | J. Phys. Chem. B 2012, 116, 14302−14310

agreement with the available experimental data. Interestingly, it was also reported that some aspects of the H-bonding complexes were sensitive to the amount of intermolecular charge transfer in the complex, especially in excited states with $\pi\pi^*$ character, which resulted in a red-shifted OPA band. In that work, nothing was reported on the proton-transfer processes that are supposed to lie at the heart of the ESPT process ending in red fluorescence from the excited state.

The purpose of our work was to focus on the topography of the potential energy surface of the ground and, especially, photoactive states of both proteins to determine whether the proposed mechanisms for the ESPT processes are energetically feasible and yield proton-transfer products in the excited state from which fluorescence is possible. In this regard, the energetics of the proton-transfer processes triggered by photoabsorption by the chromophore is expected to be quite sensitive to the environment of the chromophore: In the protein, the chromophore and residues that take part in the aforementioned proton transfers experience the interaction with a very structured environment composed of the inside part of the cavity that hosts the chromophore in these proteins, including water molecules that might be in the vicinity. This is different from bulk solvent environment. In fact, the protein environment exerts its influence by means of Coulombic interactions, but also by means of extended hydrogen bonding to residues nearby. Thus, to determine whether the proposed mechanism is feasible, it is necessary to introduce the environment, taking into account the residues involved with as much detail as possible. In this work, the energies of ground and excited states were determined with models that explicitly include all atoms in the vicinity of the groups supporting the proton transfers proposed by Piatkevich et al.[22,23]

Of course, this will imply models with hundreds of atoms. About the only affordable approach that is precise enough to produce quality energy values for excited states of systems of this size and kind is TDDFT,[27−31] and even in this case, large amounts of computer power are necessary. However, TDDFT poses another problem, as standard DFT hybrid functionals (such as B3LYP) can lead to severe underestimations of the excitation energies of excited states with large contributions of charge-transfer character.[28,29] The main interest lies not only in the computation of the absorption wavelength but also in the precise estimation of the exoergodicity of the proton-tranfer process in the photoactive state and the height of the potential energy barriers, if any. Therefore, the risk of underestimations of the excitation energy, which would lead to wrong results, is too large. To this end, as described in the Theoretical Methods section, we disregarded the use of standard hybrid functionals and instead chose one recent hybrid functional with correct asymptotic behavior to avoid problems in describing charge-transfer states.

2. THEORETICAL METHODS

To explore and locate the different electronic excited states that are linked to $\pi\pi^*$ transitions, time-dependent density functional theory (TDDFT)[27] calculations were successfully performed. For each calculation, 20 electronic excited states converged. This allowed us to construct the diabatic excited state as described later. The energies of the ground states were found using single-point DFT calculations of the crystallographic modeled structures. Moreover, the TDDFT method has been established as a suitable and proper way to study and analyze the different ESPTs such as those found in these systems.[28,30] TDDFT is also known to give very reliable results provided that the excited state is not of the charge-transfer type. In that case, TDDFT grossly underestimates the excitation energy (because of the inability of pure DFT methods to properly describe long-range exchange effects).[29,31] This system has been described as containing some charge-transfer character in the excited states.[26] To minimize as much as possible the poor description of this type of system, the Coulomb-attenuated CAM-B3LYP functional was used.[32] This functional varies the Hartree−Fock and DFT exchange-correlation contributions depending on the electron−nucleus distance. It has been recently shown that CAM-B3LYP performs extremely well for the excitation energies of GFP-like chromophores.[33]

The considered basis set was the 6-31+G(d,p) basis set, which includes polarization functions for all of the atoms and diffuse functions for heavy atoms.[34,35] Despite being a smaller basis set than that used by Zhang et al.,[26] previous works have concluded that the use of larger and more complex basis sets does not necessarily improve the quality of the results when working with DFT methods,[36] and in fact, we test this point at the beginning of the Results and Discussion section. Furthermore, it is necessary to take into account the huge size of the systems. A greater basis set would increase the number of basis functions, making the calculation more difficult and even unfeasible. For the 6-31+G(d,p) basis set, our LSSmKate1 and LSSmKate2 clusters reach 2378 and 2258 basis functions, respectively. For both ground and excited electronic states, the Gaussian 09 program was used.[37]

3. RESULTS AND DISCUSSION

Currently, the study of the structural and spectroscopic properties of the new LSS proteins is still in its infancy. To date, only very little information has been published, including a suggested mechanism based on the crystallographic structure and the absorption and emission experimental wavelengths from LSSmKate1 and LSSmKate2. The latter are the only data available for comparison with our research. This involved analysis of the ground and excited states of the proteins to find the energy difference between them and, thereby, the wavelength related with this transition. The best way to do this would be to run dynamical simulations to obtain proper structures that would share more similarities with the real proteins located in a suitable physiological environment. That would better represent the spectroscopic properties of the LSSmKate proteins and would allow the hypothetical mechanism to be accepted or refuted. However, such a treatment is quite lengthy and expensive, and it is currently underway in our laboratory. Because of these difficulties, it is customary in computational works to use very much reduced models of the protein and perform regular optimization approaches to obtain such structures and spectra.

3.1. Small Cluster Models. In the first part of our study, we consider such a reduced model and discuss the results obtained as compared with the "real" protein. As cited in the introduction, Zhang et al.[26] published the only theoretical study so far about the two recent red fluorescent proteins LSSmKate1 and LSSmKate2. In that work, they considered the isolated chromophores and two cluster models. These models represented the photoactive part of the protein incorporating the chromophore and reduced versions of the amino acids responsible for the ESPT. Nevertheless, the real protein contains a larger and more complex structure around the chromophore that includes amino acids other than the receptor

dx.doi.org/10.1021/jp3104134 | J. Phys. Chem. B 2012, 116, 14302−14310

ones, a specific distribution of charges among them, and some water molecules that stabilize these partial charges and contribute to the H-bonding network.

On the other hand, another important aspect to bear in mind in this study is the fact that the photoactive state (S_1 in the reduced cluster model of Zhang et al.[26]) is of the charge-transfer type. This character was already stated by Zhang et al. and is evidenced here by the large change of the permanent dipole moment upon electronic excitation. Table 1 presents

Table 1. Dipole Moment Components (in Debye) for Each Electronic State Together with the Norm of the Dipole Moment and the Norm of the Variation of the Dipole Moment for LSSmKate1 and LSSmKate2

		μ_x	μ_y	μ_z	$\lvert\vec{\mu}\rvert$	$\lvert\overrightarrow{\Delta\mu}\rvert$
		LSSmKate1				
CAM-B3LYP/6-31+G(d,p)	ground	11.65	2.26	−1.93	12.02	5.74
	excited	6.31	0.14	−1.91	6.59	
CAM-B3LYP/6-311++G(d,p)	ground	12.08	2.36	−1.89	12.45	5.75
	excited	6.69	0.36	−1.87	6.96	
B3LYP/6-311++G(d,p)	ground	9.94	1.78	−1.73	10.25	4.67
	excited	6.04	−0.78	−1.64	6.31	
		LSSmKate2				
CAM-B3LYP/6-31+G(d,p)	ground	12.09	2.67	1.63	12.49	5.42
	excited	7.40	−0.04	2.01	7.67	
CAM-B3LYP/6-311++G(d,p)	ground	12.24	2.54	1.65	12.61	5.27
	excited	7.67	−0.07	2.01	7.93	
B3LYP/6-311++G(d,p)	ground	10.50	1.47	1.80	10.76	4.94
	excited[a]	6.31	−1.13	2.09	6.74	

[a]Geometry obtained at the CAM-B3LYP/6-311++G(d,p) level used to avoid problems of convergence.

these variations in the dipole moment as calculated using our three chosen combinations of functional and basis set. A noticeable change in the dipole moment can be observed. Knowing the notorious drawback of the B3LYP functional when dealing with systems that present a charge-transfer character, the results of Zhang et al. should be revised.

To further discuss these models and some related issues, we tried to reproduce the results of Zhang et al. using the fragmentary information supplied in their article and the associated Supporting Information.[26] Table 2 compares the observed absorption wavelength (the only experimental data that can be readily compared to the theoretical calculations) with the results of applying the hybrid B3LYP and Coulomb-

Table 2. Absorption Wavelengths and Corresponding Oscillator Strengths Calculated Using Small Cluster Models for the Three Combinations of Functional and Basis Set Considered Together with the Experimental Absorption Wavelengths for LSSmKate1 and LSSmKate2

	LSSmKate1			LSSmKate2		
computational level		λ (nm)	f		λ (nm)	f
CAM-B3LYP/6-31+G(d,p)	S_1	443.49	1.00	S_1	414.99	1.22
CAM-B3LYP/6-311++G(d,p)	S_1	436.98	1.00	S_1	409.25	1.22
B3LYP/6-311++G(d,p)	S_1	508.94	0.82	S_1	468.51	1.01
experimental		463			460	

attenuated CAM-B3LYP functionals. For the latter, the effect of slightly diminishing the basis set from the 6-311++G(d,p) basis set used in the previous work to the slightly smaller 6-31+G(d,p) basis set was also analyzed.

By following the same procedure as used by Zhang et al.[26] but working with CAM-B3LYP as the functional and 6-31+G(d,p) as the basis set, the calculated absorption wavelengths were 443.49 nm for LSSmKate1 and 414.99 nm for LSSmKate2. Trying to improve these results, we obtained 436.98 nm for LSSmKate1 and 409.25 nm for LSSmKate2 using the same functional but the larger basis set proposed by Zhang et al. [6-311++G(d,p)].[26] All of these values correspond to shorter wavelengths than the experimental values. Alternatively, repeating the same calculations with the hybrid B3LYP functional and the larger basis set led to 508.94 and 468.51 nm for LSSmKate1 and LSSmKate2, respectively, very close to the values published by Zhang et al.[26] As expected, these calculated absorption wavelengths are overestimated with respect to the previous absorption wavelengths obtained using CAM-B3LYP (by approximately 60−70 nm), as CAM-B3LYP provides a better description of the charge-transfer excited states.

In their work, Zhang et al.[26] added an environment using the continuum PCM method to introduce the solvent effect (water, ε = 78.5 at 298 K) to largely improve the LSSmKate1 result. (The LSSmKate2 result is almost unaffected by the introduction of the solvent effect.) However, it should be noted that the results of a continuum method to introduce the solvent in a protein with the shape of LSSmKates is fully meaningless. Similarly to GFP, the two red LSS proteins have a structure with the chromophore embedded in a quite rigid (β-barrel) amino acid chain (see Figure 2), so there are only a few

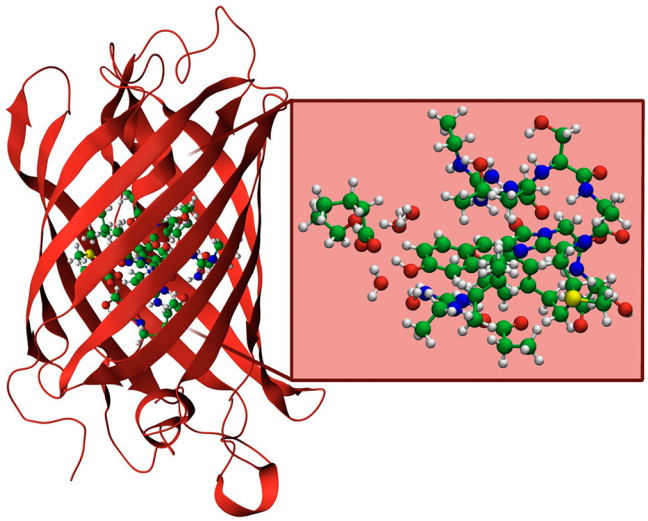


Figure 2. Picture showing the barrel structure of the whole protein and a closer view of its inside, the model considered for the different calculations. In this case, the figure presents the LSSmKate1 model. The model for LSSmKate2 is completely analogous.

isolated water molecules around the chromophore that could affect the excited-state proton-transfer reaction. Therefore, the final good fit of the absorption wavelengths obtained by Zhang et al.[26] has to be considered the result of a fortuitous error cancellation. At this point, there are two options for obtaining a good description of the absorption and emission wavelengths: (a) One can use CAM-B3LYP with PCM calculations for water as the solvent. However, we already commented that the

dx.doi.org/10.1021/jp3104134 | J. Phys. Chem. B 2012, 116, 14302−14310

protein environment cannot be represented at all by a continuum environment of water. In addition, the use of the CAM-B3LYP functional would probably impede the error cancellation of the values produced with B3LYP. (b) One can use CAM-B3LYP and a much more complete model of the protein, including the main residues near the chromophore. Indeed, this is the most suitable and realistic choice. The price to pay for the considerable enlargement of the molecular model is that the resulting size will prevent its geometrical optimization.

Given that our model considers the full molecular structure surrounding the chromophore, it contains more than 200 atoms. Therefore, the 6-311++G(d,p) basis set is directly prohibitive, reaching almost 3000 basis functions. However, when the slightly smaller 6-31+G(d,p) basis set was used with the small cluster model, we found that the calculated absorption wavelengths differed by less than 7 nm from those calculated with the larger basis set. Because the improvement is not really significant, the use of a lower basis set in our calculations is perfectly acceptable.

A final word of caution has to be made for the geometry optimization of highly reduced protein models as was done in the previous work by Zhang et al.,[26] because they come from molecular interactions that are unlikely to be found in the actual structure of the protein. Because of the lack of some linking amino acids, the optimization of reduced structures that do not contain all of the amino acids can easily lead to inadequate conformations and, consequently, to large errors in the electronic excited-state energies.

3.2. Large Cluster Models. The impossibility of geometrical optimizations for the full LSSmKate cluster models gave us no other option than to work with the crystallographic structures that can be found in the Protein Data Bank (PDB) as 3NT9 for LSSmKate1 and 3NT3 for LSSmKate2. It is reasonable to think that each crystallographic structure from the PDB represents a minimum of energy and that they could be used as good starting points to describe and represent the reactant side. From these structures, choosing a representative part of the whole protein that could describe the proper interactions and reproduce the spectroscopic features of the two fluorescent proteins has been an important issue. To build them, we included the chromophore surrounded by a first amino acid layer including the main amino acid acceptors of the different proton transfers, namely, Glu160 for LSSmKate1 and Ser158 and Asp160 for LSSmKate2; partial structures of the surrounding amino acids, namely, Ala59, Thr60, Ser61, Phe62, Ser66, Tyr67, Arg92, Ser158, Arg197, Leu199, and Glu215 for LSSmKate1 and Thr60, Ser61, Phe62, Ser66, Tyr67, Arg92, Glu145, Leu174, Ser176, and Arg197 for LSSmKate2; and crystallographic water molecules. For that, we established a uniform surrounding of 4 Å radius, positioning the origin of this radius in each atom of the chromophore. All dangling bonds were saturated with hydrogens. Because of the low resolution of X-ray diffraction, all hydrogen atoms were missing from the structure and had to be added. Special care was taken to correctly set the protonation state of titrable groups (assuming a value of pH 7), and specifically in this case, the included hydrogens were configured such that, when possible, hydrogen bonds between residues and/or crystallographic water molecules were correctly represented. The sheer size of these models compelled us to cut and saturate with hydrogen atoms every side chain extending beyond the 4 Å radius, trying to preserve as much as possible the nature of the chemical bonds as in the original molecule. To study the absorption and emission processes separately, we created two different models for each protein, modifying only the position of the transferable protons. These are bonded to their donor groups to reproduce the light absorption by the reactants. On the other hand, these protons bonded to their acceptor groups to represent the light emission by the products. In the end, the models reach the remarkable figures of 206 atoms for LSSmKate1 and 196 atoms for LSSmKate2 (see Figure 2).

Considering all of the calculated transitions between the ground and electronic excited states, we focused on those transitions with high oscillator strength values. Analyzing the orbital representation, one can see that the orbitals that participate in the absorption and emission processes have the same nature. More precisely, a nodal plane exists in the chromophore that allows us these transitions to be identified as $\pi \rightarrow \pi^*$. There are various excited states with different nonzero oscillator strengths. All of these are described in the TDDFT formalism as involving the same excited determinants with varying coefficients. Table 3 presents the states with the largest oscillator strengths at both the reactant and product sides.

Table 3. Excited States (S_n) with the Largest Oscillator Strength (f) at the Reactant and Product Regions for LSSmKate1 and LSSmKate2 Calculated Using Large Cluster Models

LSSmKate1					
reactant (λ_{exp} = 463 nm)[a]			product (λ_{exp} = 624 nm)[b]		
S_7	λ = 470.41 nm[c]	f = 0.23	S_2	λ = 554.48 nm[d]	f = 0.45
Excitations and Coefficients[e]			Excitations and Coefficients[e]		
363 → 377		−0.10	373 → 377		0.21
370 → 377		−0.33	374 → 377		−0.35
371 → 377		0.44	375 → 377		0.55
372 → 377		0.35			
LSSmKate2					
reactant (λ_{exp} = 460 nm)[a]			product (λ_{exp} = 605 nm)[b]		
S_3	λ = 456.48 nm[c]	f = 0.76	S_1	λ = 523.37 nm[d]	f = 0.91
Excitations and Coefficients[e]			Excitations and Coefficients[e]		
345 → 357		0.12	355 → 357		0.68
350 → 357		0.66	355 → 358		−0.11
353 → 357		0.13			

[a]Experimental absorption wavelength. [b]Experimental emission wavelength. [c]Calculated absorption wavelength. [d]Calculated emission wavelength. [e]Excitation transitions contributing to the excited state together with their corresponding coefficients. Only coefficients >0.1 are shown.

Our next point is to discuss the quality of our models and validate our method. To start, we could compare the absorption and emission wavelengths using the reactant and product values of our reaction coordinates with the experimental data obtained by Piatkevich et al.[22,23] Observing the calculated and experimental absorption wavelengths, one can see very good agreement despite the modified crystallographic structure (≈7 nm for LSSmKate1 in S_7 and ≈4 nm for LSSmKate2 in S_3). The models used for the reactants could be defined as very good representations of the chromophore and environment in the reactants. Although these crystallographic models have a relative closeness to the real geometry of the reactants, their use is limited by their minimal flexibility, especially the fluorescence emission where one does not have relaxed structures. This

dx.doi.org/10.1021/jp3104134 | J. Phys. Chem. B 2012, 116, 14302−14310

difference between our models and the real structure starts to become important at the product side, and the calculated wavelengths do not reproduce the experimental data well, differing by ∼70 nm for LSSmKate1 in S_2 and by ∼82 nm for LSSmKate2 in S_1. As expected, the correct representation of the product of proton transfer would require taking into account the relaxation processes occurring after photoexcitation and proton transfers. Therefore, this rigid model based on the crystallographic structure can only partially reproduce the effects of the proton transfers in the relative energy of the electronic states.

It seems clear that a crossing pattern along the ESPT exists. In fact, it is fairly impossible to stay in a unique adiabatic state because the oscillator strength of the initial excited state is dramatically decreased to 0.0026 for LSSmKate1 (S_7) and to 0.0012 for LSSmKate2 (S_3) at the end of the reaction. Hence, fluorescence would not be possible once the product formed. The question remaining now is the following: Is there a way to obtain a reasonable picture of the photoactive state energy profile? From our results, it seems that photoexcitation leaves the system in the reactant region in excited state S_7 for LSSmKate1 and S_3 for LSSmKate2, and from the above discussion, fluorescence can take place effectively only from a different adiabatic state, S_2 for LSSmKate1 and S_1 for LSSmKate2, because they are the only ones that have large enough values of the oscillator strength in the product region. With these data, it seems logical to conclude that the system must change the adiabatic state between photoexcitation and fluorescence. The next goal then is to obtain a potential energy profile of a diabatic state connecting the photoexcitation and fluorescence regions. A diabatic state connects different adiabatic states based on the properties of their electronic wave functions over a range of nuclear coordinates. A way of achieving this to obtain an approximate picture of the diabatic state connecting the photoexcitation and fluorescence regions, while paying attention to the similarity of the electronic properties, consists of joining those adiabatic state segments where the oscillator strength is highest (in itself a quantitative function of the transition dipole moment and, in turn, of the wave functions of the ground and excited electronic states involved). This is certainly just one of many possible criteria, but it provides an approximate rule on which to construct the diabatic state on reasonable grounds that is correct at least at the two ends of the reaction coordinate.

Following this criterion, we computed the reaction coordinate shifting the different protons from their donor to their acceptor groups, retaining all the remaining coordinates frozen. As before, we have placed each proton in such a way that it is collinear with its donor and acceptor atoms. Each reaction coordinate is defined as $d_{don} - d_{acc}$, where d_{don} is the distance between the transferred proton and the donor group and d_{acc} is the distance between the proton and the acceptor group. This choice is typical in transfer reaction coordinates because of its convenience: Negative values denote structures closer to reactants, whereas positive values are related to structures closer to products, and values close to zero are representative of the transition state of the transfer.

The single proton transfer of LSSmKate1 was calculated through 12 geometries starting with the proton bonded to the carboxyl group and ending bonded to the hydroxyl group of Glu160 (see Figure 1). All system coordinates were constrained except the movement of the labile proton. At every point, we analyzed the oscillator strengths along the excited-state proton-transfer reaction and determined which excited states had large values. In Figure 3, the oscillator strengths of the four excited

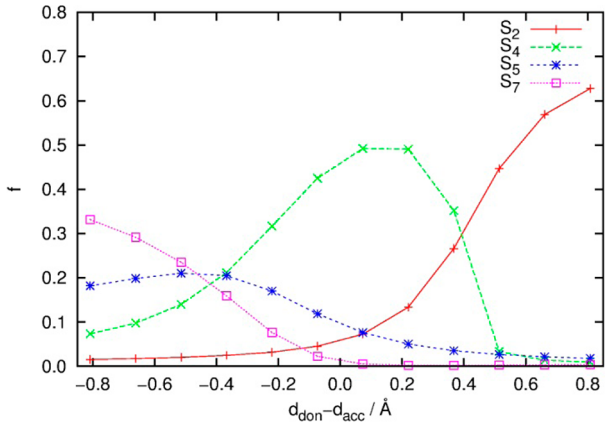


Figure 3. Values of the oscillator strength of the four most significant adiabatic excited states along the reaction coordinate for LSSmKate1.

states having the largest oscillator strengths are presented. The state with the largest *f* value moves basically along the proton-transfer reaction across these four adiabatic excited states: It begins at S_7, goes for a while to S_5, then moves to S_4, and finally ends up in S_2. So, there is an extended excited-state crossing pattern along the proton-transfer process.

The energies of the ground and photoactive excited states along the coordinate reaction are presented in Figure 4. The

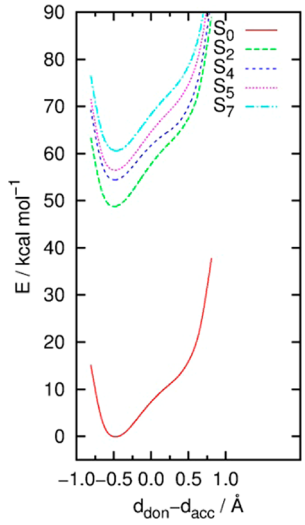


Figure 4. Potential energy profiles of ground and excited states with large oscillator strengths along the proton-transfer reaction coordinate.

ground state has a very well-defined minimum at $d_{don} - d_{acc} \approx -0.5$ Å (reactants) and an incipient shoulder at about $d_{don} - d_{acc} \approx 0.5$ Å (products), approximately 12 kcal mol^{-1} higher in energy.

A similar profile can be observed in the excited states. The molecular model used for LSSmKate1 works very well at the reactants, and each excited state shows a well-described energy minimum. However, the excited-state energy profiles in the product region show an interesting behavior: If the proton transfer took place in the S_7 adiabatic state, the attainment of the product would require a very large excess of energy relative to the Franck−Condon arrival point. Besides, fluorescence will

dx.doi.org/10.1021/jp3104134 | J. Phys. Chem. B 2012, 116, 14302−14310

be very slow owing to the vanishing value of the oscillator strength. Because of this, we assume the ESPT in this state must take place diabatically, ending up in the S_2 adiabatic state at the product region. That is the reason why it is necessary to diabatize.

Figure 5 shows the energy profile of the diabatic state constructed following the state with largest oscillator strength

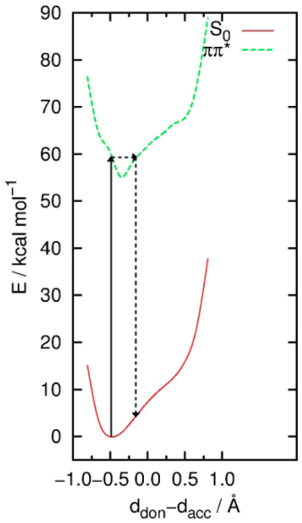


Figure 5. Diabatized photoactive excited state (see text) for the $\pi \rightarrow \pi^*$ transition along the proton-transfer reaction coordinate. The solid lines denote the Franck−Condon transition (vertical) and the corresponding energy level (horizontal). The vertical dashed line denotes the possible fluorescence from the region where $d_{don} - d_{acc} = 0$.

along the proton-transfer reaction coordinate. This state also has a potential energy minimum at the reactants but shifted toward less negative values of $d_{don} - d_{acc}$. It is remarkable that the energy at the Franck−Condon point and that at $d_{don} - d_{acc} = 0$ are almost the same. The diabatic profile does not contain a minimum in the product region of the reaction coordinate, even though the potential energy increases very slowly in that region. Nevertheless, we believe that our results are compatible with the putative mechanisms for LSSmKate1 proposed by Piatkevich et al.[22,23] for two reasons: First, the actual system is not rigid and can relax after the photoexcitation. Such process would bring down the product side of the potential energy profile. Second, fluorescence can occur for these parts of the wavepacket in the photoactive state over the product region. In fact, they should have a large overlap with excited vibrational states of the ground electronic state, and there are very large oscillator strengths in the product region (see Figure 3).

We now turn our attention to LSSmKate2, whose fluorescence is supposed to arise after a double proton transfer is complete. Now, the proton of the hydroxyphenyl group of the chromophore forms a hydrogen bond with the hydroxyl group of Ser158 (H1 coordinate), which, in turn, makes another hydrogen bond with the carboxylic group of Asp160 (H2 coordinate). In principle, the two protons might move independently from each other, revealing the process to be concerted or stepwise. A total of 7 × 7 = 49 points were computed here to provide a mesh from which the motion of the two protons can be analyzed.

Figure 6 shows the energy landscape of the ground state. As was the case in LSSmKate1, the proton-transfer reaction in the

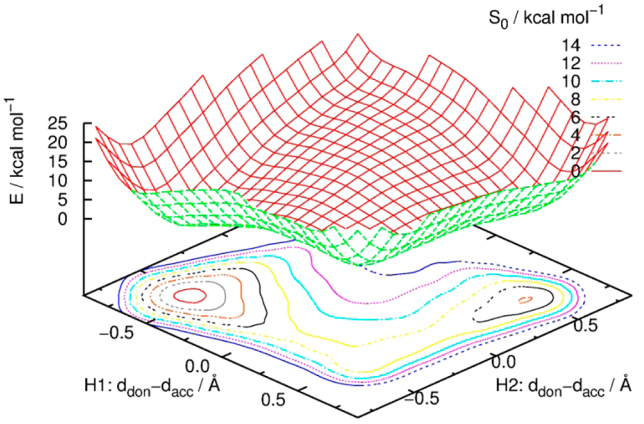


Figure 6. Representation of the potential energy surface for the ground state of LSSmKate2 along the motion of the two transferring protons.

ground state is an endoergic process. However, this time, there is an actual minimum-energy region in the product side. The transformation must go over a barrier of ∼7−8 kcal mol^{-1} and yields a product approximately 4 kcal mol^{-1} higher in energy. The process is concerted although highly asynchronous, initiated by the chromophore proton (H1). The reaction path visualized in Figure 6 shows that the second proton (H2) does not begin to transfer until the transfer of the first proton (H1) is almost completed.

Following the same strategy as used for LSSmKate1, the construction of the potential energy surface for the photoactive state of LSSmKate2 involves several steps. First, we identified the state with highest oscillator strength corresponding to a $\pi \rightarrow \pi^*$ transition on the chromophore. Then, picking the state with the highest oscillator strength for each point, we can represent the energy of each selected excited state to obtain a diabatic potential energy surface along the double proton-transfer reaction coordinate. Now, the excited-state proton transfer is an exoergic reaction. Therefore, the behaviors of the ground and excited state are opposed, making the displacement of the two protons in the excited states easier and causing a large shift between the absorption and fluorescent wavelengths. The fact that the process of double proton transfer is exoergic in the excited states predicts a substantial Stokes shift, even in this unrelaxed potential energy profiles (see Figure 7).

As in the ground state, the diabatic photoactive state does not involve a simultaneous movement of the two protons along the reaction path. The apparent absence of an intermediate minimum in the $\pi \rightarrow \pi^*$ state at coordinates H1 ≈ 0.5 Å and H2 ≈ −0.5 Å rules out a strictly stepwise mechanism. The reaction mechanism is described as concerted movement of the two protons but completely asynchronous. Initially, the diabatic ESPT starts with the motion of the H1 proton. This moves across three different adiabatic excited states, starting on S_3, going past S_2, and ending in S_1. Once the first proton has been transferred, H2 begins its movement, but it remains at every moment on S_1 until H2 reaches the carboxylic group of Asp160 (see Figure 1). Energetically, the process in the photoexcited state involves a potential energy barrier, about 4 kcal mol^{-1}, with respect to the Franck−Condon structure. This barrier occurs during the transfer of the first proton. The product is then formed in an energetically downhill process ending approximately 3 kcal mol^{-1} below the Franck−Condon excitation energy. From there, fluorescence can occur. This

dx.doi.org/10.1021/jp3104134 | J. Phys. Chem. B 2012, 116, 14302−14310

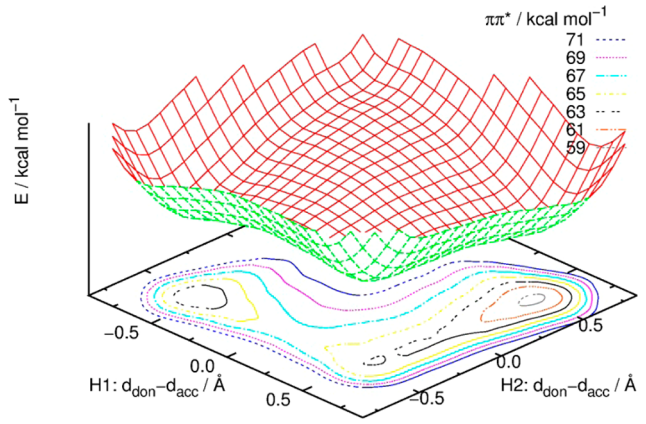


Figure 7. Representation of the potential energy surface for the diabatic photoactive excited state of LSSmKate2 along the motion of the two transferring protons.

result provides strong support for the mechanism proposed by Piatkevich et al.[22,23] for LSSmKate2.

4. CONCLUSIONS

In this article, we have presented the results of an initial exploration of the potential energy landscape governing the proton-transfer reactions in the red fluorescent proteins with large Stokes shifts LSSmKate1 and LSSmKate2, as suggested by Piatkevich et al. from crystallographic data.[22,23] This was done on the ground and excited states, with the aim of determining the feasibility of the proposed mechanisms. Special stress has been placed on including the effect of the close protein environment in the description of the energetics of the proton-transfer reactions. To this end, complete cluster models of the chromophore and its surroundings within roughly 4 Å were used. However, because of the huge size of these structures, all of the geometries remained fixed. On these cluster models, DFT (for the ground state) and TDDFT calculations (for as many as 20 excited electronic states) were computed. The calculations indicate that the chromophore undergoes a $\pi \rightarrow \pi^*$ transition at wavelengths that agree very well with available data of peaks of absorption, being in error by about 4−7 nm, which validates the accuracy of the calculational level and model used. Fluorescence wavelengths display a larger error, and this is attributed to the inability of these rigid cluster models to include the effects of a possible relaxation of the environment of the chromophore after photoexcitation and proton transfer. In both proteins, it was found that the adiabatic electronic state with the largest oscillator strength at the Franck−Condon region was not the one from which fluorescence could occur in the products.

The feasibility of the proton-transfer reactions was studied. In the case of LSSmKate1, a single proton transfer has been proposed, so an unrelaxed potential energy profile was computed by shifting the position of the proton of the chromophore toward the acceptor atom in Glu160 along the straight line linking the donor and acceptor atoms. It was found that the ground state displays a single minimum in the reactant region. As for the excited states, the adiabatic state with the largest oscillator strength at the reactants displays a vanishingly small oscillator strength at the products and has an energy that is too high to be affordable after photoexcitation. Conversely, other states lower in energy at the product side display large oscillator strengths. This suggests a nonadiabatic reaction. Therefore, we studied an approximate diabatic state in which, at each configuration, the adiabatic state with the largest oscillator strength was picked. This tentative "photoactive" state still displays a single minimum on the reactant side further shifted toward the products, but if geometrical relaxation had been performed, it would probably bring down the product side of the proton transfer.

LSSmKate2 has been proposed to undergo a two-link proton-wire transfer. The same procedure as discussed in the preceding paragraph was followed, now leading to a two-dimensional potential energy surface. The ground state displays a deep minimum in the absolute reactants. The proton-transfer process in the ground state is unfavorable, but would correspond to a concerted and very asynchronous mechanism where the chromophore's proton moves first, followed by that of Ser158. The excited-state description of LSSmKate2 in this cluster model is similar to that of LSSmKate1. However, the bundle of electronic states with important oscillator strengths is present only during the portion of the surface related to the transfer of the chromophore's proton. Afterward, only the lowest excited state displays an appreciable oscillator strength as the second proton transfers. The same approximate diabatization procedure as in LSSmKate1 was followed, and this renders a photoactive state with two minima, one at the absolute reactants and the other at the final products. The overall process in the photoactive state is exoergic by approximately 3 kcal mol^{-1} with respect to the Franck−Condon point and presents a single potential energy barrier of about 4 kcal mol^{-1}, encountered during the first proton transfer. The process is, thus, concerted, highly asynchronous—the first proton moves first—and favorable even in the absence of geometrical relaxation.

Summarizing, the theoretical study presented herein has revealed the surprising possibility that the proton-transfer processes in the photoactive states for LSSmKate1 and LSSmKate2 are nonadiabatic in nature. The next step in this detailed study for both proteins and the proton-transfer processes would require the consideration of the full environment.

■ ASSOCIATED CONTENT

Supporting Information

Complete Cartesian coordinates of the large cluster models of LSSmKate1 and LSSmKate2, including the initial and final positions of the studied ESPT; discussion on the effects of optimization and environment on the structures used; and complete lists of authors of refs 21 and 37. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ricard.gelabert@uab.cat. Tel.: (+34) 935811669. Fax: (+34) 935812477.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the "Ministerio de Economía y Competitividad" through Project CTQ2011-24292 and by the "Generalitat de Catalunya" through Project 2009SGR409. Use of computational facilities at the "Centre de Serveis Científics i Acadèmics de Catalunya (CESCA)" is gratefully acknowledged.

C.R. also acknowledges the bestowing of a fellowship within the FPU program by the "Ministerio de Economía y Competitividad".

REFERENCES

(1) Chalfie, M. *Angew. Chem., Int. Ed.* **2009**, *48*, 5603−5611.
(2) Shimomura, O. *Angew. Chem., Int. Ed.* **2009**, *48*, 5590−5602.
(3) Tsien, R. Y. *Angew. Chem., Int. Ed.* **2009**, *48*, 5612−5626.
(4) Remington, S. J. *Protein Sci.* **2011**, *20*, 1509−1519.
(5) Dooley, C. T.; Dore, T. M.; Hanson, G. T.; Jackson, W. C.; Remington, S. J.; Tsien, R. Y. *J. Biol. Chem.* **2004**, *279*, 22284−22293.
(6) Remington, S. J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 714−721.
(7) Remington, S. J.; Wachter, R. M.; Yarbrough, D. K.; Branchaud, B.; Anderson, D. C.; Kallio, K.; Lukyanov, K. A. *Biochemistry* **2005**, *44*, 202−212.
(8) Shcherbakova, D. M.; Hink, M. A.; Joosen, L.; Gadella, T. W.; Verkhusha, V. V. *J. Am. Chem. Soc.* **2012**, *134*, 7913−7923.
(9) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509−544.
(10) Zimmer, M. *Chem. Rev.* **2002**, *102*, 759−781.
(11) Shimomura, O. *FEBS Lett.* **1979**, *104*, 220−222.
(12) Sanders, J. K. M.; Jackson, S. E. *Chem. Soc. Rev.* **2009**, *38*, 2821−2822.
(13) Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Prendergast, F. G.; Cormier, M. J. *Gene* **1992**, *111*, 229−233.
(14) Yang, F.; Moss, L. G.; Phillips, G. N. *Nat. Biotechnol.* **1996**, *14*, 1246−1251.
(15) Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8362−8367.
(16) van Thor, J. J.; Zanetti, G.; Ronayne, K. L.; Towrie, M. *J. Phys. Chem. B* **2005**, *109*, 16099−16108.
(17) Erez, Y.; Gepshtein, R.; Presiado, I.; Trujillo, K.; Kallio, K.; Remington, S. J.; Huppert, D. *J. Phys. Chem. B* **2011**, *115*, 11776−11785.
(18) Henderson, J. N.; Gepshtein, R.; Heenan, J. R.; Kallio, K.; Huppert, D.; Remington, S. J. *J. Am. Chem. Soc.* **2009**, *131*, 4176−4177.
(19) Kogure, T.; Karasawa, S.; Araki, T.; Saito, K.; Kinjo, M.; Miyawaki, A. *Nat. Biotechnol.* **2006**, *24*, 577−581.
(20) Shaner, N. C.; Patterson, G. H.; Davidson, M. W. *J. Cell. Sci.* **2007**, *120*, 4247−4260.
(21) Shcherbo, D.; Merzlyak, E. M.; Chepurnykh, T. V.; Fradkov, A. F.; Ermakova, G. V.; Solovieva, E. A.; Lukyanov, K. A.; Bogdanova, E. A.; Zaraisky, A. G.; Lukyanov, S.; Chudakov, D. M. *Nat. Methods* **2007**, *4*, 741−746.
(22) Piatkevich, K. D.; Hulit, J.; Subach, O. M.; Wu, B.; Abdulla, A.; Segall, J. E.; Verkhusha, V. V. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 5369−5374.
(23) Piatkevich, K. D.; Malashkevich, V. N.; Almo, S. C.; Verkhusha, V. V. *J. Am. Chem. Soc.* **2010**, *132*, 10762−10770.
(24) Henderson, J. N.; Osborn, M. F.; Koon, N.; Gepshtein, R.; Huppert, D.; Remington, S. J. *J. Am. Chem. Soc.* **2009**, *131*, 13212−13213.
(25) Violot, S.; Carpentier, P.; Blanchoin, L.; Bourgeois, D. *J. Am. Chem. Soc.* **2009**, *131*, 10356−10357.
(26) Zhang, M. Y.; Wang, J. Y.; Lin, C. S.; Cheng, W. D. *J. Phys. Chem. B* **2011**, *115*, 10750−10757.
(27) Dreuw, A.; Head-Gordon, M. *Chem. Rev.* **2005**, *105*, 4009−4037.
(28) Sobolewski, A. L.; Domcke, W. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3065−3072.
(29) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2004**, *120*, 8425−8433.
(30) Aquino, A. J. A.; Lischka, H.; Hattig, C. *J. Phys. Chem. A* **2005**, *109*, 3201−3208.
(31) Handy, N. C. *Mol. Phys.* **2004**, *102*, 2399−2409.
(32) Yanai, T.; Tew, D.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51−57.
(33) List, N. H.; Olsen, J. M.; Rocha-Rinza, T.; Christiansen, O.; Kongsted, J. *Int. J. Quantum Chem.* **2012**, *112*, 789−800.
(34) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654−3665.
(35) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta.* **1973**, *28*, 213−222.
(36) Papajak, E.; Zheng, J. J.; Xu, X. F.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*, 3027−3034.
(37) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N. ; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C; Jaramillo, J.; Gomperts, R. E.; Stratmann, O.; Yazyev, A. J.; Austin, R.; Cammi, C.; Pomelli, J. W.; Ochterski, R.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, version C.01; Gaussian, Inc.: Wallingford, CT, 2009.