Excited States of GFP Chromophore and Active Site Studied by the SAC-CI Method: Effect of Protein-Environment and Mutations

JUN-YA HASEGAWA,[1,2] KAZUHIRO FUJIMOTO,[1] BEN SWERTS,[1] TOMOO MIYAHARA,[1,2] HIROSHI NAKATSUJI[1,2]
[1]*Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Kyoto-Daigaku-Katsura, Nishikyo-ku, Kyoto 615-8510, Japan*
[2]*Quantum Chemistry Research Institute (QCRI), Kyodai Katsura Venture Plaza, Goryou Oohara 1-36, Nishikyo-ku, Kyoto 615-8245, Japan*

Received 25 August 2006; Revised 2 November 2006; Accepted 24 November 2006
DOI 10.1002/jcc.20667
Published online 23 August 2007 in Wiley InterScience (www.interscience.wiley.com).

Abstract: Excited states of fluorescent proteins were studied using symmetry-adapted cluster-configuration interaction (SAC-CI) method. Protein-environmental effect on the excitation and fluorescence energies was investigated. In green fluorescent protein (GFP), the overall protein-environmental effect on the first excitation energy is not significant. However, glutamine (Glu) 94 and arginine (Arg96) have the red-shift contribution as reported in a previous study (Laino et al., Chem Phys 2004, 298, 17). The excited states of GFP active site (GFP-W22-Ser205-Glu222-Ser65) were also calculated. Such large-scale SAC-CI calculations were performed with an improved code containing a new algorithm for the perturbation selection. The SAC-CI results indicate that a charge-transfer state locates at 4.19 eV, which could be related to the channel of the photochemistry as indicated in a previous experimental study. We also studied the excitation and fluorescence energies of blue fluorescent protein, cyan fluorescent protein, and Y66F. The SAC-CI results are very close to the experimental ones. The protonation state of blue fluorescent protein was determined. Conformation of cyan fluorescent protein indicated by the present calculation agrees to the experimentally observed structure.

 J Comput Chem 28: 2443–2452, 2007

Key words: fluorescent proteins; GFP; BFP; CFP; Y66F; SAC-CI

Introduction

Green fluorescent protein (GFP) is involved in the jellyfish, *Aequorea Victoria*,[1–6] and has very efficient emission property. It is now widely used as an excellent molecular marker in various fields of biology.[6,7] Recent developments realized variety of GFP mutants having different fluorescence colors.[6–10] Structure of three representative GFP mutants are shown in Figure 1. Blue fluorescent protein (BFP) has histidine (imidazole group) instead of Tyr66 (*p*-hydroxyphenyl group) of the GFP chromophore. This replacement shifts the absorption and fluorescent peaks to the higher-energy region.[10] Replacements to tryptophane (indole group) and to phenylalanin (phenyl group) lead cyan fluorescent protein (CFP)[8] and Y66F,[9] respectively. These fluorescent proteins are also used for donor and acceptor in the fluorescence resonance energy transfer[11,12] Therefore, fluorescent proteins have become one of the key molecules in the field of biomolecular imaging.

Most of the fluorescent proteins were created by genetic cloning[13] and mutation technique.[7] The aim of these experiments is to improve its efficiency (quantum yield and intensity) and color-variation. Since fluorescence is a transient photochemical event, contributions from theoretical studies would be very important to this issue. For the better understanding and the rational design of the photochemical reaction in the fluorescent proteins, it is necessary to understand the electronic structure, spectroscopy, excited-state dynamics, mechanism of the color-determination, and the role of the protein environment. There are theoretical studies investigating the spectroscopy of GFP chromophore model compound.[14–21] Regarding the potential surface of the excited state of GFP chromophore, a complete active space self-consistent field method and second-order perturbation theory (CASPT2) study showed the two-dimensional potential surface to explain the fate of the excited state.[22] On the protein environment, there are only a few studies.[19,23–26] The emitting state of GFP I-state was studied by quantum mechanical (QM) (CASPT2)/molecular mechanics (MM) (CHARMM) approach.[23] The protein effect to the absorption spectrum of BFP[24] and

Correspondence to: H. Nakatsuji; e-mail: h.nakatsuji@qcri.or.jp

Contract/grant sponsors: Ministry of Education, Culture, Sports, Sciences, and Technology of Japan; Matsuo Foundation; Computing Service Group, ACCMS and IIMC, Kyoto University

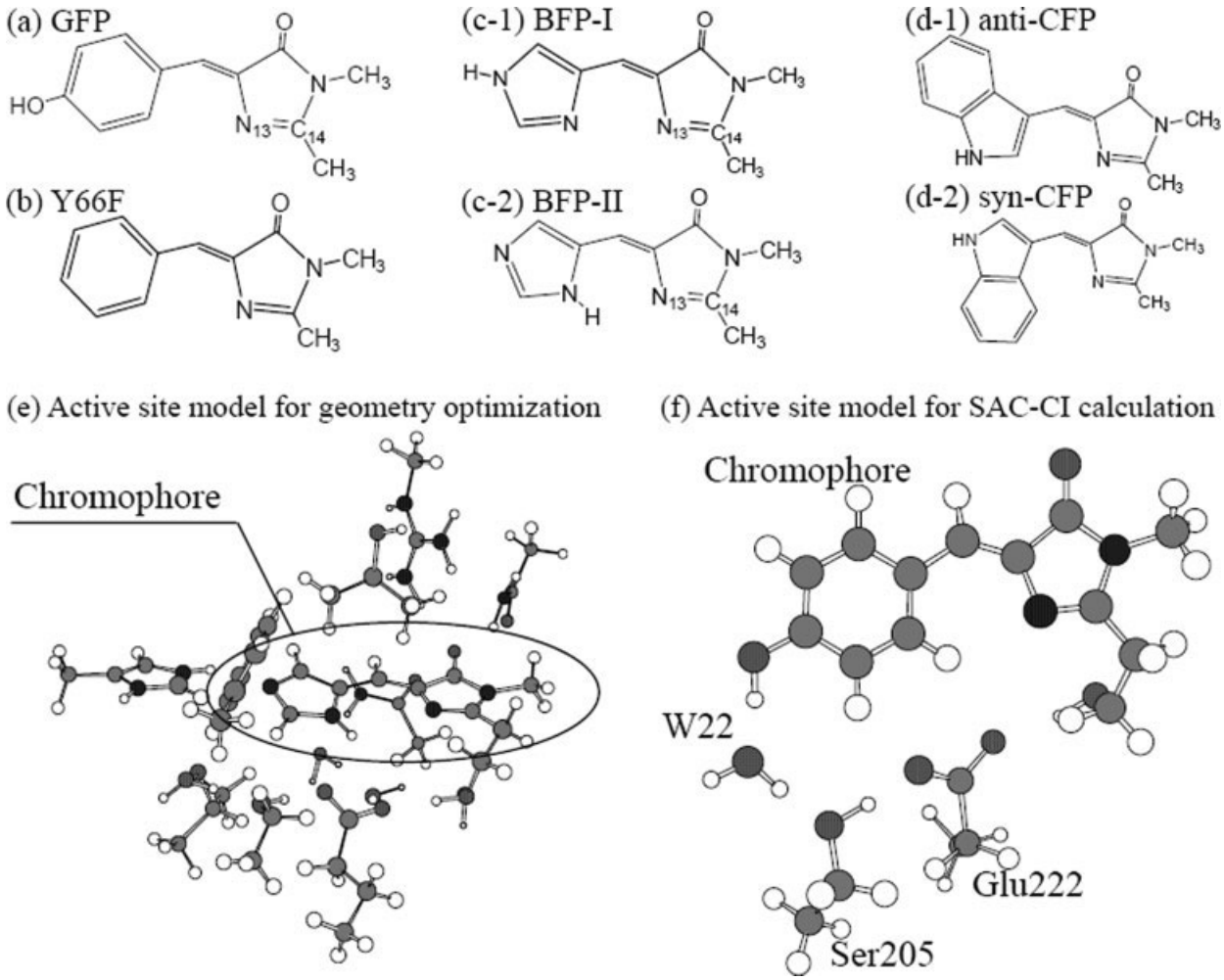

Figure 1. Computational models. (a–d) Chromophores of GFP and its mutants. (e) Active site model of BFP for the geometry optimization. (f) Active site model of GFP for the SAC-CI calculations.

CFP[25] were also studied at time-dependent density functional theory (DFT) level. These studies reported that the protein-environmental effect causes small shift in the absorption energy. In the present study, we also studied the environmental effect on the absorption and fluorescence energies of GFP, BFP, and Y66F.

Protein environment may contribute to the photochemistry of GFP not only as environment but also as the donor in the electron-transfer. Radiating UV (254 nm, 4.9 eV) or visible (390 nm, 3.2 eV) lights induce photochemical conversion of the GFP active site.[7,27,28] Recently, van Thor et al. revealed the X-ray structure of GFP_{UV} (phototransformed GFP by irradiating UV light) and showed that decarboxylation occurs at Glu222 (see Fig. 1f) during the photo process. A hypothetical mechanism proposed[28] is that the radiation causes charge-transfer (CT) excitation from Glu222 to the GFP chromophore, and the Glu222 is decarboxylated by Kolbe reaction. However, there is so far neither experimental nor theoretical evidence for the CT excitation. To investigate intermolecular CT excited states, we performed the symmetry-adapted cluster-configuration interaction (SAC-CI) calculations for the GFP active site including the chromophore, Water 22, Ser205, Glu222, and Ser65.

Protonation state of the chromophore becomes very important, when the excited-state proton transfer occurs. Although X-ray structure of GFP[29,30] and BFP[10] was obtained, it is difficult to determine the protonation state of the chromophore only from the X-ray results. In the case of GFP, it has reached to a consensus: the A and B forms are neutral and anionic forms, respectively.[17,21,31,32] Our previous SAC-CI result also supported these protonation states.[21] In the case of BFP, there are two possibilities as indicated in (c-1) and (c-2) of Figure 1. In this study, we performed the SAC-CI calculation for these two protonation states. On the basis of the theoretical-spectroscopic point of view, we assess the protonation state of the BFP chromophore.

We have already reported a SAC-CI theoretical study on the excited states of the GFP chromophore in its various protonation states.[21] The SAC-CI excitation and fluorescent energies showed nice agreement with the experimental data for the GFP chromophore in various protonation states. The present study extends the previous study. First, the absorption and fluorescence ener-

gies of GFP chromophore are calculated by the SAC-CI method, and the protein-environmental effect is discussed. Next, we investigated the excited states of the GFP active site, and theoretical evidence of the CT from Glu222 to the chromophore has been obtained. We further extended our study to the GFP mutants, BFP, CFP, and Y66F, and the electronic structure of the excited chromophores are explained. In large-scale SAC-CI calculations, one of the time-consuming steps was the selection of the two-electron excitation operators. In our recent development, we improved the routines for the perturbation selection of the doubles. The calculation of the GFP active site in the present study includes 447 active orbitals in the SAC-CI calculation. The program improvement is briefly described after the computational details.

Computational Details

Figures 1a–1d show computational models for the chromophores of the GFP and other three mutants. For GFP, only the neutral form was focused on. Two protonation states were considered for BFP [BFP-I (c-1) and BFP-II (c-2)]. Two conformational isomers were examined for CFP [anti-CFP (d-1) and syn-CFP (d-2)].

We used three types of computational models, Model I, II, and III. The Model I is the chromophore in the gas phase, and the structures are shown in Figures 1a–1d. In Model II, only the chromophore was treated by QM method, but the protein electrostatic effect was included by the point-charge model. Model III additionally includes the hydrogen-bonding network connecting to the chromophore. As shown in Figure 1f, the chromophore, Water22, serine (Ser) 205, glutamate (Glu) 222, and Ser65 were treated by the QM calculation, and the rest of the protein-electrostatic effect was taken into account by the point-charge model.

For the Model I, the molecular geometry was optimized by DFT with B3LYP[33,34] functional for the ground state and CI Singles for the excited states. The basis sets used were the 6-31 g* sets[35,36] for the DFT calculations. The D95(d)[37] sets were used for the CIS calculations. In the Model II, the ground-state structures were optimized by using DFT (B3LYP). The computational model used for the optimization was composed of QM and MM regions. The QM region for BFP case was shown in Figure 1e. Equivalent QM region was taken for GFP and Y66F. Total nine amino residues and four water molecules were included in the QM region, and the rest of the protein environment was treated by the point-charge model. Positions of the QM atoms in the backbone and the MM atoms were fixed to the reference X-ray structure during the energy minimization. A MM force field, AMBER96,[38] was used. For the excited-state structure of Model II, we used QM(CIS/6-31 g*)/MM(AMBER96[38]) method[39] for the energy minimization. The structures shown in Figures 1a and 1b were used for the QM region, and the rest of the proteins were treated by the MM force field. Positions of all the atoms were optimized by the QM/MM method. The initial structures of the GFP and BFP were taken from Protein Data Bank (1EMB[29,30] for GFP and 1BFP[10] for BFP). Model III was employed only for GFP, and the geometry used was the same as that in the Model II.

The detail of the SAC-CI method can be found elsewhere in a previous review.[40] The basis sets used for the C, N, and H atoms were of double-ζ plus polarization quality.[37] To describe the CT excitation from Glu222 to the chromophores, single p-type anion functions ($\alpha = 0.059$)[37] were augmented for the atoms in the GFP chromophore in the SAC-CI calculation for the GFP active site (Model III). The active space for the SAC-CI calculation included all the valence orbitals. The 1s orbitals and their corresponding virtual orbitals were treated as the frozen orbitals. All the single-excitation and selected double-excitation operators were included in the SAC/SAC-CI wave functions. The perturbation-selection method[41] was used for selecting doubles. The energy threshold for the ground and excited states are 1×10^{-5} and 1×10^{-6} au, respectively. For the calculation of the GFP active site, we used improved set of the threshold, 5×10^{-6} and 5×10^{-7} au, for the ground and excited states, respectively. The CI-singles wave functions were used for the reference states in the perturbation selection.

All calculations were carried out with the development version of Gaussian03 program system.[42] We have improved the computational algorithm for the perturbation selection of the two-electron operators (See the next section). This development realized the large-scale SAC-CI calculation of the GFP active site (Fig. 1f) having 447 molecular orbitals (MOs) in the active spaces. This development has been released via Gaussian03 revision D01.

Improved Algorithm for the Perturbation Selection of the Double-Excitation Operators

For selecting important double-excitation operators to be included in the SAC/SAC-CI wave functions, we use the second-order perturbation energy as the criteria (perturbation selection).[41] In the case of SAC and SAC-CI SD-*R* methods, Hartree-Fock (HF) and CI-Singles (CIS) wave functions are used for the reference function, respectively.

$$E^{(2)}_{iajb} = \frac{|\langle \Psi^{(0)}_{\mathrm{ref}} | \hat{H} | \Phi^{ab}_{ij} \rangle|^2}{E^{(0)}_{iajb} - E^{(0)}_{\mathrm{ref}}} \tag{1}$$

The indices $i, j, \ldots$ and $a, b, \ldots$ refer to occupied and unoccupied orbitals, respectively. Configuration Φ^{ab}_{ij} is included in the SAC/SAC-CI wave function, if the second-order energy $E^{(2)}_{iajb}$ is larger than a given threshold. In large-scale calculations, the most time-consuming step is the computation of the numerator, $\langle \Psi^{(0)}_{ref} | \hat{H} | \Phi^{ab}_{ij} \rangle$. When the CIS wave function for the singlet state, $\langle \Psi_{\mathrm{CIS}}| = \Sigma_{ia} C_{ia} \langle \Phi_{ia}|$, is used for the reference function, the explicit expression is as follows.

$$\begin{aligned}
\langle \Psi_{\mathrm{CIS}} | \hat{H} | \Phi^{ab}_{ij} \rangle &= \sum_{ia} C_{ia} \langle \Phi_{ia} | \hat{H} | \Phi^{ab}_{ij} \rangle \\
&= -2^{-1/2} \Bigg[\sum_k C_{ka} \{2(ik|jb) - (ib|jk)\} \\
&+ \sum_k C_{kb} \{2(jk|ia) - (ja|ik)\} - \sum_c C_{ic} \{2(ca|jb) - (cb|ja)\} \\
&- \sum_c C_{jc} \{2(cb|ia) - (ca|ib)\} \Bigg]
\end{aligned} \tag{2}$$

Table 1. Timing Data (CPU time) for the Perturbation Selection.

	CPU time (HP DS25)	
	Integral sorting	Selection
Singlet ground states		
Previous	None	3m 25s
Present	1m30s	48s
Singlet excited states		
Previous	None	1h 53m 10s
Present	1m38s	6m 7s
Triplet states		
Previous	None	6h 47m 53s
Present	1m37s	11m 48s

Cyan fluorescent protein, $C_{15}H_{15}N_3O_2$ (C_1-symmetry), with DZP level of basis sets. The 1s core and corresponding virtual orbitals were frozen. Total number of active space is 290 (51 occ. & 239 unocc.). Computation was performed HP DS25 workstation.

Here, we assume that HF orbitals are used, and the occupied–unoccupied block of Fock matrix is zero. With this equation, number of operation count for the selection is proportional to $O(N^5)$. To reduce the computational effort, we only included the configuration whose coefficient is larger than 0.01 in the CIS wave function. This approximation reduces the number of operation to $O(MN^3)$, where M is the number of the configuration adopted in the reference CIS wave function. Since the coefficient vector becomes sparse after the cut-off, the matrix multiplication between the coefficient vector and molecular integrals is not straightforward. We adopted a list-driven algorithm. In the case of the first term of eq. (2), we first count number of the selected C_{ka} in each index a, Nk(a). Next, we construct the list of index k for each index a, and the multiplication is performed for the fixed i, j, a, and b.

In Table1, we show the timing data. The present algorithm is compared with the previous one adopted in the Gaussian 03 rev. C02. The system is a chromophore of CFP, $C_{15}H_{15}N_3O_2$ (C_1-symmetry). The basis sets of DZP level[37] was used, and total 290 active orbitals (51occupied and 239 unoccupied orbitals) were correlated in the SAC/SAC-CI calculation. The number of the reference states was eight in the selection. The comparison shows that the CPU time was remarkably reduced for singlet and triplet excited states. The present selection algorithm has been released in the Gaussian03 rev. D01.

Results and Discussion

Excitation and Fluorescent Energies of the GFP Chromophore

In Table2, SAC-CI results are compared with the experimental data. For GFP, we focused only on the A form (neutral form). The anion form has already been discussed in our previous paper.[21] The SAC-CI method nicely reproduced the experimental peak positions of GFP and Y66F. The experimental absorption and emission originate from the first excited state in all the fluorescent proteins under the study. The nature of the states is one-electron excitation from the highest-occupied molecular orbital (HOMO, π) to the lowest-unoccupied molecular orbital (LUMO, π^*) in all the chromophores.

In Figure 2, frontier orbitals of the GFP chromophore are analyzed by the orbital-correlation diagram. MOs of the chromophore are characterized in terms of the phenol (Tyr) and imidazolinone (Im) groups. HOMO of GFP is composed of HOMOs of Tyr and Im units with the anti-bonding interaction, while LUMO is the bonding combination between next-LUMO of Tyr and LUMO of Im. The HOMO and LUMO are delocalized over the chromophore. The first excited state is therefore characterized as local exciton state of GFP chromophore. On the other hand, the next HOMO and next LUMO of the chromophore are localized within the Tyr unit. This is because these Tyr MOs have node on the carbon atom which connects to the Im unit.

Table 2. Excitation and Fluorescent Energies of the Chromophore of GFP and Y66F Mutants.

	Excitation energy (eV)				Fluorescence energy (eV)		
	SAC-CI				SAC-CI		
	Model I	Model II	Model III		Model I	Model II	
Fluorescent protein	Chro.[b]	Chro.[b] + PC[c]	Active site[d] + PC[c]	Exptl.	Chro.[b]	Chro.[e] + PC[c]	Exptl.
GFP[a]	3.23	3.21	3.27	3.12[f]	2.70	2.73	2.7[f]
Y66F	3.33	3.32		3.44[f,g]	2.86	2.83	2.8[f,g]

The results are compared among several computational models.
[a]Green fluorescent protein in A-form (neutral).
[b]Choromophore model shown in Figures 1a and 1b.
[c]Point charge model.
[d]Active site model shown in Figure 1f.
[e]Chromophore model shown in Figures 1a and 1b. A water is also included in the quantum mechanical region.
[f]Ref. 7.
[g]Ref. 9.

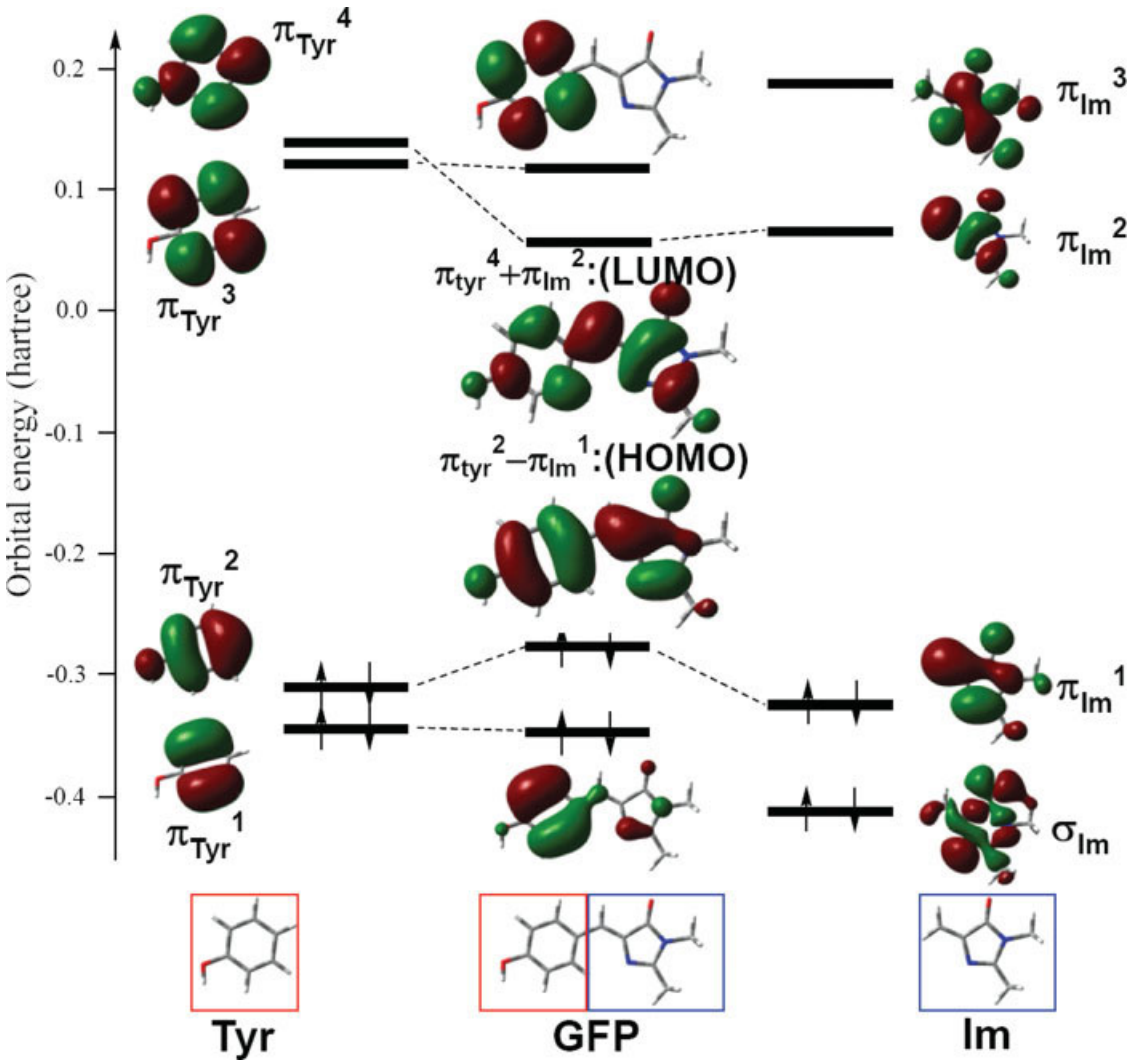

Figure 2. Orbital correlation diagram of GFP chromophore. MOs of the chromophore are characterized in terms of the phenol (Tyr) and imidazolinone (Im) groups. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Table 2 shows the SAC-CI results obtained with the three computational models. Model I is the chromophore in the gas phase, and the computational model is shown in Figures 1a and 1b. In Model II, only the chromophore was treated by QM method, but the protein-electrostatic effect was taken into account by the point-charge model. Model III additionally includes the hydrogen-bonding network connecting to the chromophore. As shown in Figure 1f, GFP chromophore, Water22, Ser205, Glu222, and Ser65 were treated by quantum-mechanical calculation, and the rest of the protein-electrostatic effect was taken into account by the point-charge model.

As seen in Table 2, the protein-environmental effect to the excitation and fluorescence energies of GFP is −0.02 in Model II and +0.04 eV in Model III. The gas phase model (Model I) gives the excitation and fluorescence energies close to the experimental ones, even though the protein-environmental effect was not considered.

One of the reasons is that the character of the excited states is a local excited state without CT character as seen in Figure 2. Some important HF MOs obtained with the Model III are also shown in Figure 3. The 103th and 107th orbitals are related to the transition in the first excited state. These orbitals are clearly localized on the GFP moiety. Consequently, change in the Mulliken atomic charges upon the excitation is small as shown in Figure 4. Maximum change is −0.09 at the C8 atom, and RMS of the change is 0.03. As shown in Table 4, the same trend is also observed in the Model III results. The Tyr unit decreases its charge by only 0.02, and the bridging methin (—CH—) unit and the imidazolinone group increases by only 0.10 and 0.02 eV, respectively.

In Figure 4, we also show the protein-electrostatic potential at each atomic center. The electrostatic potential on the GFP chromophore does not show any specific change in the chromophore binding site. The excitation-energy shift caused by the

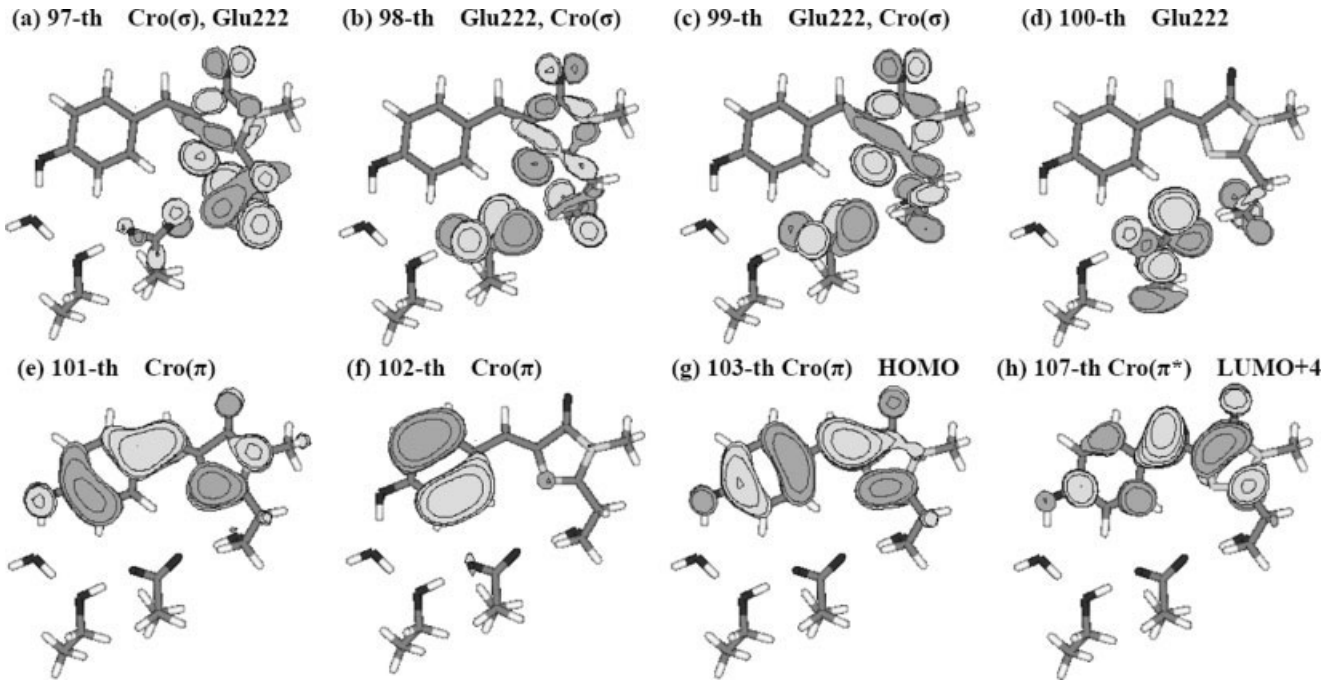

Figure 3. Some important molecular orbitals of the GFP active site. Character of the MOs is also shown herewith. Graphical representation of MOs were carried out with Molden software.

protein effect can be roughly estimated by $\Sigma(Q_{ex} - Q_g) \times V_{ESP}$, where Q_{ex} and Q_g are atomic charge in the excited and ground states, respectively. V_{ESP} is the protein-electrostatic potential. With this estimation, the protein-electrostatic effect is blue-shift of 0.05 eV compared with the gas phase. This value is very small compared with the retinal proteins in which the excitation energy shifts about 0.4 eV by the electrostatic environment.[43] This difference is basically relevant to the electronic structure of the excited state. The excited state of retinal Schiff base has CT character, while that of GFP has a local exciton character.

In a previous study of GFP,[19] the effect of the neighboring residues was studied in detail. In addition to the chromophore, they included Gln94 and Arg96 in the computational model. In their result, the Gln94 and Arg96 decrease the excitation energy of the GFP chromophore by 0.30–0.36 eV in a semiempirical CAS-CI calculations and by 0.16–0.17 eV in DFT HOMO-LUMO gap.[19] We also evaluated the effect of these residues using the classical treatment shown above. These residues reduce the excitation energy by about 0.1 eV, which qualitatively agrees with the previous analysis.[19] This indicates that the rest of the protein-electrostatic effect increases the excitation energy and diminishes the red-shift effect of the Gln94 and Arg96.

Excited States of the GFP Active Site

A recent study[28] has solved the crystal structure of GFP_{UV}. On the basis of the structure, a hypothetical mechanism was proposed for the generation of GFP_{UV}: Radiation of UV (254 nm, 4.9eV) or visible (390 nm, 3.2 eV) lights causes CT excitation from the Glu222 to the chromophore, and this CT excitation triggers the decarboxylation at Glu222. However, there have been neither experimental nor theoretical evidence on the existence of the CT state.

To investigate the existence of the CT excited state, the computational model was extended and included the chromophore, Glu222, Ser205, Ser65, and one water molecule. These residues constitute hydrogen-bonding network in this system (see Fig. 1f). The rest of the protein effect was treated by the point-charge model.

Figure 3 shows MOs that are important in the excited states. The 97th–99th MOs have amplitude both on Glu222 and the chromophore. These MOs are composed of π orbital of glutamate, n(lone-pair) orbital of Ser65, and $(\sigma + n)$-orbital of imidazolinone ring. The 100th orbital is mostly localized on Glu222. The 101–103(HOMO)-th MOs are π-orbitals of the chromophore, and they are clearly localized within the chromophore moiety. In the unoccupied orbitals, there are a few low-lying diffuse orbitals. Their characters are regarded as mixed diffuse-σ* and diffuse-π* orbitals. The 107th orbital (LUMO + 4) has valence π*-character localized within the chromophore. The shapes of these π-orbitals are almost the same as those obtained in the gas-phase calculations of the chromophore.

In Table3, we show the results of the SAC-CI calculation for the low-lying 10 excited states both for singlet and triplet spin multiplicities. The excitation energy of the 2^1A state is calculated to be 3.27 eV. This corresponds well to the experimentally observed peak position at 3.12 eV, showing that the present calculation have reasonable accuracy.

Based on the SAC-CI result, we point out that the target states of the experimental radiation have π–π* character within the chromophore, not the CT excitation from Glu222 to the chromophore. The radiations of 390 nm (3.2 eV) and 254 nm

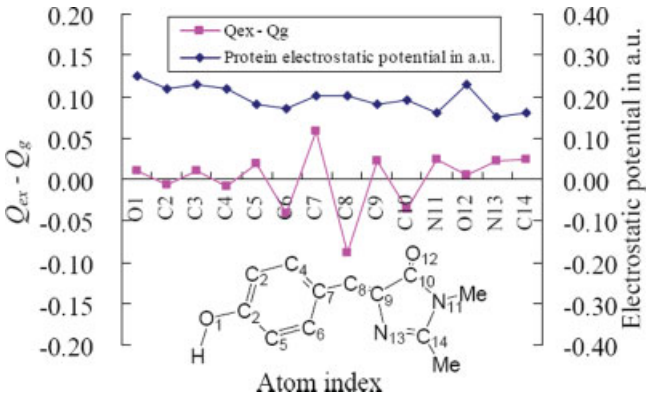

Figure 4. Change in the atomic charge upon transition from the ground to the first excited state ($Q_{ex}-Q_g$). Protein electrostatic potential on each atomic center is also shown. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

(4.9 eV) light correspond to 2^1A (3.27 eV) and 8^1A (4.85 eV) states, respectively. The 2^1A and 8^1A states are excitations from HOMO to 107th MO and from 101th MO to 107th MO, respectively. As seen in Figure 3, these MOs are localized within the chromophore. In Table4, Mulliken population analysis for several important excited states is shown. The result also indicates that the excitations do not significantly affect the electron population of the chromophore and Glu222.

However, we found that the 5^1A state calculated at 4.18 eV has the CT character. The main configuration is the excitation from 97th–99th MOs to 107th MO. The population analysis clearly shows the transfer of around −0.5 charges from Glu222 to the chromophore. There is no other CT state within the singlet states below 5.5 eV in the present SAC-CI calculation. We also investigated triplet excited states. As shown in Table 3, there is a triplet CT state, 5^3A state, calculated at 4.09 eV. The character is very close to that of the 5^1A state, indicating that the 5^3A state is a triplet counter-part of the 5^1A state. The amount of the charge transfer is very similar to the 5^1A state. These states would be the possible candidate for the CT state proposed in the previous experimental study.[28]

The previous paper[28] assumed that the imidazolinone ring of the chromophore is positively charged in the ground state, which stabilizes the CT state. However, the ring is just slightly charged (+0.04) in the SAC-CI result.

Based on the present result, there is a possibility that the radiation of 254 nm (4.9 eV) light leads the CT excitation by through the relaxation process among the excited states. It might be interesting to examine illumination of 295 nm (4.2 eV) light, to directly generate the CT state (5^1A state at 4.18 eV), even though the oscillator strength of the state is very small. On the other hand, the radiation of 390 nm light gives the 2^1A state. However, there is no CT state below the 2^1A state. If the decarboxylation at Glu222 occurs by the radiation of 390 nm light, there should be another mechanism that leads to the decarboxylation. There is also a possibility that the chromophore is excited to the states around 6.4 eV (3.2 × 2) by the two-photon processes, since GFP has large two-photon absorption cross section.[44,45] In this case, the charge transfer excited state is able to generate via the relaxation process among the excited states.

Table 3. Singlet and Triplet Excited States of the Green Fluorescent Protein Active Site.

	SAC-CI				Exptl.
State	Main configuratiors (C > 0.3)	Character	E_{ex} (eV)[a]	Osc. (au)[b]	E_{ex} (eV)
1^3A	−0.89(103 → 107)	Cro π → Cro π^*	1.77	–	
2^1A	0.90(103 → 107)	Cro π → Cro π^*	3.27	0.56	3.12
2^3A	0.56(101 → 107)−0.36(103 → 121)	Cro π → Cro π^*	3.71	–	
3^3A	0.79(103 → 104)	Cro π → Cro Ryd.	3.96	–	
3^1A	−0.90(103 → 104)	Cro π → Cro Ryd.	3.98	4.0×10^{-3}	
4^3A	0.43(103 → 104)−0.37(103 → 105)−0.33(102 → 107)−0.31(103 → 110)	Cro π → Cro Ryd.	4.05	–	
5^3A	0.61(99 → 107)+0.47(98 → 107)+0.42(97 → 107)	Cro σ, Glu222 → Cro π^*	4.09	–	
4^1A	0.84(103 → 106)−0.38(103 → 105)	Cro π → Cro Ryd.	4.11	1.7×10^{-3}	
5^1A	−0.61(99 → 107)−0.47(98 → 107)−0.42(97 → 107)	Cro σ, Glu222 → Cro π^*	4.18	2.7×10^{-2}	
6^3A	0.72(103 → 106)	Cro π → Cro Ryd.	4.24	–	
6^1A	0.65(103 → 105)+0.35(103 → 106)	Cro π → Cro Ryd.	4.34	1.1×10^{-2}	
7^3A	−0.56(103 → 105)−0.33(101 → 107)+0.33(102 → 110)	Cro π → Cro Ryd.	4.47	–	
8^3A	0.60(103 → 105)+0.36(103 → 106)	Cro π → Cro Ryd.	4.54	–	
7^1A	0.48(103 → 105)−0.47(103 → 110)+0.34(102 → 107)	Cro π → Cro Ryd.	4.56	1.2×10^{-2}	
8^1A	0.72(101 → 107)−0.33(103 → 108)	Cro π → Cro π^*	4.85	0.15	
9^1A	−0.75(103 → 108)−0.31(103 → 109)	Cro π → Cro Ryd.	4.95	6.8×10^{-3}	
9^3A	0.72(102 → 107)−0.36(103 → 110)	Cro π → Cro π^*	4.96	–	
10^1A	0.84(103 → 109)	Cro π → Cro Ryd.	5.17	1.0×10^{-2}	
10^3A	0.66(95 → 107)	Cro π → Cro π^*	5.35	–	
11^1A	0.81(102 → 106)	Cro π → Cro Ryd.	5.58	8.9×10^{-2}	

[a]Excitation energy in eV unit.
[b]Oscillator strength in atomic unit.

Table 4. Mulliken Population Analysis for the Several Important States of the GFP Active Site.

	Singlet states				Triplet states	
Group	X^1A (Ground)	2^1A	5^1A	8^1A	1^3A	5^3A
Chromophore	−0.17	−0.17	−0.66	−0.09	−0.10	−0.67
(Phe[a])	(0.01)	(−0.01)	(−0.55)	(−0.35)	(0.06)	(−0.54)
(Bridge[b])	(−0.25)	(−0.15)	(0.24)	(−0.21)	(−0.18)	(−0.25)
(Im[c])	(0.04)	(0.06)	(0.29)	(0.49)	(0.05)	(−0.04)
(Ser[d])	(0.02)	(−0.02)	(0.17)	(−0.02)	(−0.03)	(0.16)
Glu222	−0.73	−0.76	−0.23	−0.79	−0.78	−0.21
Ser205	−0.14	−0.25	−0.16	−0.16	−0.16	−0.16
Wat22[d]	0.04	0.04	0.04	0.04	0.04	0.04

[a]Phenyl group.
[b]Bridge (−CH−) unit.
[c]Imidazolinone group.
[d]Ser65 unit, namely HO−CH_2−CH_2− group.

Excitation and Fluorescent Energies of the GFP Mutants, Y66F, BFP, and CFP

The Y66F mutant has phenyl-group instead of the *p*-hydroxyphenyl-group as shown in Figure 1b. The SAC-CI excitation and fluorescence energies with the Model I (gas phase) are 3.33 and 2.81 eV, respectively. After including the protein-environmental effect with the Model II, the results obtained (3.32 eV for excitation and 2.83 eV for fluorescence) were almost the same as the gas phase ones. This indicates the environmental effect on the excitation and emission energies is very small in Y66F mutant. These SAC-CI results are very close to the experiment[7] as shown in Table 2.

For BFP, there are two possibilities in the protonation state of the chromophore as shown in Figure 1c (referred as "BFP-I" and "BFP-II" structures). As shown in Table5, the results for the BFP-II structure gives better agreement with the experimental one both in the excitation and fluorescent energies. The SAC-CI excitation energy with the Model II (in protein) is 3.16 and 3.22 eV for BFP-I and BFP-II structures, respectively. The excitation energy for BFP-II is closer to the experimental value of 3.25 eV. The fluorescence energies calculated for BFP-I and BFP-II are 2.92 and 2.77 eV, respectively, and the BFP-II one is closer to the experimental one (2.78 eV). In addition, the ground state of BFP-II structure is energetically more favorable than that of BFP-I. The SAC and DFT (B3LYP) results with Model I show that the former is more stable than the latter by 4.6 and 10.9 kcal/mol, respectively.

Figure 5 shows the optimized structure of the BFP-II calculated by the large-active site cluster model. The BFP chromophore is more or less planer and the imidazolate-group forms the hydrogen-bonding with the neighboring His 148. On the other hand, the chromophore become nonplaner in the BFP-I structure in protein. The present result would be a new proposal for the protonation state of the BFP chromophore, since the BFP-I structure seems to be assumed in the previous X-ray study.[10]

CFP chromophore substitutes the *p*-hydroxyphenyl-group in GFP by the indole-group. There are two conformational isomers regarding the C−C single bond of the bridging unit. We performed SAC-CI calculations for the two conformers, and the result was compared to the experimental one. The X-ray study[46] showed that the CFP chromophore has anti-conformation.

In the SAC-CI result in the gas phase (Model I), the calculated excitation energy (2.99 eV) for the anti-conformer is closer to the experiment (2.84 eV) than that for the syn-conformer (3.15 eV). The calculated fluorescence energy of the anti-conformer (2.54 eV) is also closer to the experiment (2.56 eV) than that of the syn-conformer. We have also compared the ground-state energy. The anti-CFP is more stable than syn-CFP by 3.3 and 1.3 kcal/mol in the SAC and DFT(B3LYP) methods, respectively. Therefore, all the theoretical results suggested the anti-

Table 5. Excitation and Fluorescent Energies of the GFP, Y66F, BFP, and YFP Chromophores Calculated by the SAC-CI Method.

	Excitation energy (eV)			Fluorescence energy (eV)	
	SAC-CI			SAC-CI	
Molecule	Model I Chro[b]	Model II Chro[b]+PC[c]	Exptl.	Model I Chro[b]	Exptl.
GFP[a]	3.23	3.21	3.12[d]	2.70	2.70[d]
Y66F	3.33	3.32	3.44[d,e]	2.81	2.81[d,e]
BFP-I	3.41	3.16		2.92	
BFP-II	3.11	3.22	3.25[d,f]	2.77	2.78[d,f]
syn-CFP	3.15			2.68	
anti-CFP	2.99		2.84[d,g]	2.55	2.56[d,g]

[a]GFP chromophore in A-form (neutral).
[b]Choromophore model shown in Figure 1.
[c]Point charge model.
[d]Ref. 7.
[e]Ref. 9.
[f]Ref. 10.
[g]Ref. 8.

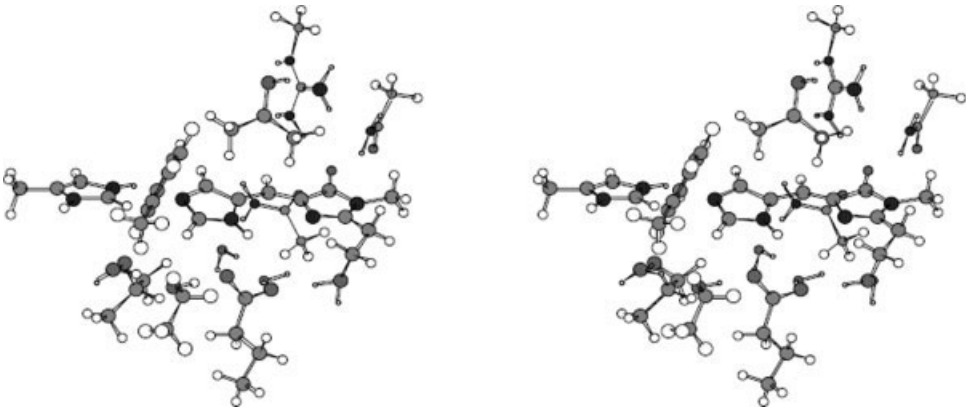

Figure 5. Optimized structure of the BFP active site (stereo view).

conformer in the CFP chromophore, which is in agreement with the X-ray experimental structure.[46]

On the C14-N13 Length in the X-ray Structures and its Relation to the Excitation Energy

Before closing, we mention about the X-ray geometry of the GFP and BFP chromophores. Comparing the 1BFP structure with the optimized one, there is a significant deviation in the C14-N13 bond length (see Figs. 1a–1d): 1.440 Å in 1BFP and 1.303 Å in the optimized structure. The other X-ray structures are relatively close to the optimized one. For example, the C14-N13 lengths in 2EMD, 2EMN, and 2EMO structures are 1.320, 1.332, and 1.339 Å, respectively. The longer bond length causes much smaller excitation energy, since the ground state become unstable. This is because MOs on the C14-N13 bond in HOMO and LUMO are bonding and anti-bonding characters, respectively. With the 1BFP structure, the SAC-CI excitation energy becomes 2.44 eV, which was by about 0.8 eV smaller than that obtained using the optimized structure. There is the similar problem in the GFP, and the relationship between structure and optical properties were already pointed out.[19] The bond lengths in 1EMB and 1EMG structures are 1.456 and 1.416 Å, respectively, while those in the optimized structure and the other X-ray structures are around 1.30–1.34 Å. It is therefore desirable for the X-ray experiments to refer quantum-mechanically determined geometry to refine the atomic coordinate.

Conclusion

Excited states of the chromophores in the GFP and its mutants were studied. The SAC-CI calculations were performed for the chromophores and the active site of the fluorescent proteins. Geometry was optimized by the QM/MM models where the rest of the protein environment (MM region) was treated by the point-charge model. In order to study the excited states of the active site of GFP, we introduced an improved algorithm for the perturbation selection of the double-excitation operators. This development has been released in the Gaussian 03 revision D01.

First, we investigated the protein environmental effect to the excited state of GFP by using several computational models. The results show that the protein-electrostatic effect on the excitation energy is -0.02 to $+0.04$ eV in total. However, some neighboring residues have specific contributions. In our analysis, the Gln94 and Arg96 decrease the excitation energy of the GFP chromophore, which agrees with the previous analysis.[19] Such environmental effect in GFP is in contrast to the case of retinal proteins which are used for the visual pigment to recognize red, green, and blue lights.

On the mechanism for the photoconversion to GFP_{UV}, a hypothesis raised by van Thor et al.[28] was examined theoretically by large scale SAC-CI calculations of the GFP active site. We found a CT state in singlet and tri plet states at 4.18 and 4.09 eV, respectively. These CT states might be responsible to the van Thor's mechanism. However, the experimental photoillumination was applied to the local excited states of GFP chromophore. To confirm the hypothetical mechanism, we therefore propose that illumination of 4.2 eV light. If the mechanism is correct, this experiment is expected to increase the quantum yield for generating GFP_{UV}.

The SAC-CI excitation and fluorescence energies show fine agreement to the experimental values of GFP, Y66F, Blue-FP, and Cyan-FP. Based on the excitation energy, fluorescence energy, and total energy, we propose that the protonation state of the BFP chromophore to be the BFP-II structure. The SAC-CI spectroscopic results are close to those of anti-CFP structures, which agrees with the X-ray structure.[46]

The existing X-ray structures, 1BFP, 1EMB, and 1EMG have elongated C14-N13 bond in the imidazolinone ring. Using this geometry, the excitation energy was significantly underestimated. If the structures were optimized by the modern accurate correlated methods, the bond-length becomes shorter, and the calculated excitation energy reasonably agrees with the experimental values. It is therefore desirable for the X-ray experiments to refer quantum-mechanically determined geometry to refine the atomic coordinate.

Acknowledgments

A part of the computations was carried out at RCCS, Okazaki, Japan.

References

1. Shimomura, O.; Johnson, F. H.; Saiga, Y. J Cell Comp Physiol 1962, 59, 223.

2. Morin, J. G.; Hastings, J. W. J Cell Physiol 1971, 77, 313.
3. Morise, H.; Shimomura, O.; Johnson, F. H.; Winant, J. J Biochem 1974, 13, 2656.
4. Ward, W. W. Photochem Photobiol Rev 1979, 4, 1.
5. Inouye, S.; Tsuji, F. I. FEBS Lett 1994, 341, 277.
6. Zimmer, M. Chem Rev 2002, 102, 759.
7. Tsien, R. Y. Annu Rev Biochem 1998, 67, 509.
8. Heim, R.; Prasher, D. C.; Tsien, R. Y. Proc Natl Acad Sci USA 1994, 91, 12501.
9. Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y. Trends Biochem Sci 1995, 20, 448.
10. Wachter, R. M.; King, B. A.; Heim, R.; Kallio, K.; Tsien, R. Y.; Boxer, S. G.; Remington, S. J. Biochemistry 1997, 36, 9759.
11. Miyawaki, A.; Tsien, R. Y. Methods Enzymol 2000, 327, 472.
12. Sato, M.; Hida, N.; Ozawa, T.; Umezawa, Y. Anal Chem 2000, 72, 5918.
13. Matz, M. V.; Fradokov, A. F.; Labas, Y. A.; Savitsky, A. P.; Zaraisky, A. G.; Markelow, M. L.; Lukyanov, S. A. Nat Biotechnol 1999, 17, 969.
14. Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. Proc Natl Acad Sci USA 1996, 93, 8362.
15. Voityuk, A. A.; Michel-Beyerle, M.-E.; Rosch, N. Chem Phys 1998, 231, 13.
16. Voityuk, A. A.; Michel-Beyerle, M.-E.; Rosch, N. Chem Phys Lett 1998, 296, 269.
17. Voityuk, A. A.; Kummer, A. D.; Michel-Beyerle, M.-E.; Rosch, N. Chem Phys 2001, 269, 83.
18. Weber, W.; Helms, V.; McCammon, J. A.; Langhoff, P. W. Proc Natl Acad Sci USA 1999, 96, 6177.
19. Laino, T.; Nifosi, R.; Tozzini, V. Chem Phys 2004, 298, 17.
20. Helms, V.; Winstead, C.; Langhoff, P. W. J Mol Struct 2000, 506, 179.
21. Das, A. K.; Hasegawa, J.; Miyahara, T.; Ehara, M.; Nakatsuji, H. J Comput Chem 2003, 24, 1421.
22. Martin, M. E.; Negri, F.; Olivucci, M. J Am Chem Soc 2004, 126, 5452.
23. Sinicropi, A.; Andruniow, T.; Ferre, N.; Basosi, R.; Olivucci, M. J Am Chem Soc 2005, 127, 11534.
24. Lopez, X.; Marques, M. A. L.; Castro, R. A. R. J Am Chem Soc 2005, 127, 12329.
25. Demachy, I.; Ridard, J.; Laguitton-Pasquier, H.; Durnerin, E.; Vallverdu, G.; Archirel, P.; Levy, B. J Phys Chem B 2005, 109, 24121.
26. Marques, M. A. L.; López, X.; Varsano, D.; Castro, A.; Rubio, A. Phys Rev Lett 2003, 90, 258101.
27. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. Science 1994, 263, 802.
28. van Thor, J. J.; Gensch, T.; Hellingwerr, K. H.; Johnson, L. N. Nat Struct Biol 2002, 9, 37.
29. Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. Science 1996, 273, 1392.
30. Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormoe, M.; Remington, S. J. Proc Natl Acad Sci USA 1997, 94, 2306.
31. Bell, A. F.; He, X.; Wachter, R. M.; Tonge, P. J. Biochemistry 2000, 39, 4423.
32. Nielsen, S. B.; Lapierre, A.; Andersen, J. U.; Pedersen, U. V.; Tomita, S.; Andersen, L. H. Phys Rev Lett 2001, 87, 228102.
33. Lee, C.; Yang, W.; Parr, R. G. Phys Rev B 1988, 37, 785.
34. Becke, A. D. J Chem Phys 1993, 98, 5648.
35. Hehre, W. J.; Ditchfield, R.; Pople, J. A. J Chem Phys 1972, 56, 2257.
36. Hariharan, P. C.; Pople, J. A. Theor Chim Acta 1973, 28, 213.
37. Dunning, T. H. Jr.; Hay, P. J. In Modern Theoretical Chemistry; Schaefer, H. F., III, Ed.; Plenum: New York, 1976; pp. 1–28.
38. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, J. K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179.
39. Swerts, B.; Van Droogenbroeck, J.; Peeters, A.; Van Alsenoy, C. J Phys Chem A 2002, 106, 4245.
40. Nakatsuji, H. In Computational Chemistry—Reviews of Current Trends; Leszczynski, J., Ed.; World Scientific: Singapore, 1997; pp. 62–124.
41. Nakatsuji, H. Chem Phys 1983, 75, 425.
42. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian Inc.: Pittsburgh, PA, 2003.
43. Fujimoto, K.; Hasegawa, J.; Hayashi, S.; Kato, S.; Nakatsuji, H. Chem Phys Lett 2005, 414, 239.
44. Volkmer, A.; Subramaniam, V.; Birch, D. J. S.; Jovin, T. M. Biophys J 2000, 78, 1589.
45. Xu, C.; Zipfel, W.; Shear, J. B.; Williams, R. M.; Webb, W. W. Proc Natl Acad Sci USA 1996, 93, 10763.
46. Bae, J. H.; Rubini, M.; Jung, G.; Wiegand, G.; Seifert, M. H. J.; Azim, M. K.; Kim, J.; Zumbusch, A.; Holak, T. A.; Moroder, L.; Huber, R.; Budisa, N. J Mol Biol 2002, 328, 1071.