Effect of Protein Environment on pK_a Shifts in the Active Site of Photoactive Yellow Protein

Masaki Yoda,[†] Yoshio Inoue,[†] and Minoru Sakurai*[,‡]

Department of Biomolecular Engineering, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan, and Center for Biological Resources and Informatics, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan

Received: August 13, 2003; In Final Form: October 12, 2003

During the photocycle of photoactive yellow protein (PYP), the pK_a values of the chromophore (*S*-methyl *p*-coumarate) and Glu46 undergo drastic changes. In the dark state (pG), the chromophore is deprotonated and Glu46 is protonated. In the blue-shifted intermediate called pB, their protonation states are inverted. These pK_a changes are thought to be closely related to the mechanism of signal transduction leading to bacterial phototaxis. In this study, we investigate the physical origin of the above pK_a shifts using quantum chemical calculations at the level of ab initio Hartree−Fock (HF), semiempirical AM1, and density functional theory (DFT). Here, *S*-methyl *p*-coumarate and acetic acid (propionic acid) are chosen as models of the chromophore and Glu46, respectively. First, the protein environment surrounding these residues is approximated by a dielectric continuum. The HF and DFT calculations coupled with a continuum solvent model (SCIPCM or COSMO) reveal that the relative pK_a (ΔpK_a) of acetic acid to the chromophore model has a positive value in low dielectric media, meaning that the chromophore is deprotonated and the acetic acid is protonated. However, with increasing dielectric constant (ϵ), the ΔpK_a value gradually decreases and its sign is inverted from plus to minus at about $\epsilon = 2-4$. Consequently, in higher dielectric media ($4 < \epsilon$), the chromophore is protonated and Glu46 is deprotonated. To obtain more detailed information about the mechanism of the ΔpK_a shifts, we perform hybrid quantum mechanics/molecular mechanics (QM/MM) calculations in which the chromophore and Glu46 are treated quantum chemically and the protein environment is approximated by an assembly of partial atomic charges, obtained from linear-scaling molecular orbital calculations (known as MOZYME) for the entire PYP. These calculations also indicate that the sign of the ΔpK_a is inverted from plus to minus on going from pG to pB. Also, detailed analysis indicates that this inversion is mainly caused by the effect of slight conformational differences between pG and pB, as observed by X-ray crystallographic studies.

Introduction

Photoactive yellow protein (PYP) is a small soluble protein found in *Ectothiorhodospira halophila*.[1] PYP is thought to be a light sensor that mediates the negative phototactic response to blue light.[2] Upon absorbing light (absorption maximum at 446 nm), the ground state of PYP (pG) enters a photocycle involving at least two intermediates, which exhibit red- and blue-shifted absorption maxima relative to that of pG, respectively.[3,4] Although the signal transduction pathway is not fully known, the blue-shifted intermediate (pB) is thought to be a signaling state that interacts with another protein.

PYP consists of 125 amino acid residues and a *p*-coumaric acid chromophore binding to Cys69 via thiol ester linkage.[5,6] This chromophore interacts with the surrounding protein matrix, and the absorption maximum shifts remarkably from that of 284 nm in the free state to 446 nm in pG. A main origin of this spectral shift is the deprotonation of the chromophore in the protein.[7,8] This deprotonated state persists in the range of pH <3, while the pK_a of the chromophore in the free state is about 9.0.[6,9] Thus the chromophore−protein interaction also causes a large pK_a shift in pG, where the negative charge on the deprotonated chromophore may be stabilized via a hydrogen-bonding network, involving Tyr42, Glu46, and Thr50, and by the positive charge on Arg52.[10] Additionally, Glu46, located near the phenol group of the chromophore, has an unusually high pK_a value (>7). Thus, in pG, the chromophore is deprotonated and Glu46 is protonated. Interestingly, the relative order of their pK_a values is inverted during the photocycle of PYP: in pB the chromophore is protonated while Glu46 is deprotonated. Therefore, the acid−base equilibrium between these residues is an important factor to understanding the formation of the signaling state of PYP.

The chromophore−protein interaction in PYP could consist of specific interactions, such as the above hydrogen-bonding network, and nonspecific interactions, such as a dielectric response of the protein matrix. The dielectric response effects in proteins are difficult to evaluate experimentally, and hence their biological significance has been substantiated mainly by computational studies.[11] A frequently used pK_a calculation method for proteins is based on an electrostatic model where the protein is treated as an assembly of point charges embedded in a dielectric continuum.[12,13] Such a model has been already applied to PYP.[14] According to that study, the experimental pK_a value of the chromophore in the pG state is reproduced if the protein interior is assumed to be a low dielectric medium. However, only partial agreement between the experimental and theoretical data was obtained for the pB state based on the crystal structure.[15]

The above electrostatic type of pK_a calculation can take into account effects of the protein structure and the surrounding solvent at a uniform theoretical level, but the physical description

* Corresponding author. E-mail: msakurai@bio.titech.ac.jp.
[†] Department of Biomolecular Engineering.
[‡] Center for Biological Resources and Informatics.

10.1021/jp0364102 CCC: $25.00
Published on Web 12/03/2003

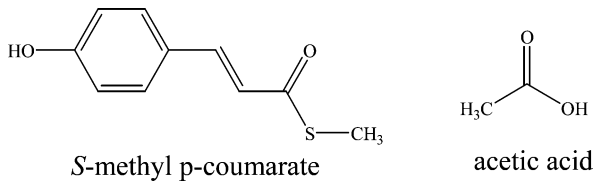

Figure 1. Structure of System 1. *S*-Methyl *p*-coumarate and acetic acid are infinitely separated.

of residues of interest is rather poor. Thus, another approach to be able to quantum chemically describe the important residues may be required for better understanding of the pK_a problem in PYP. In this study, first the relative pK_a values of Glu46 to the chromophore are investigated using quantum chemical calculations (ab initio Hartree−Fock (HF) and density functional theory (DFT)) coupled with continuum solvent models. These results provide the basic information to understanding how the acid−base equilibrium between the chromophore and Glu46 is influenced by the polarizable protein environment. Next, to more explicitly take into account the protein environment, we perform hybrid quantum mechanics/molecular mechanics (QM/MM) calculations in which the chromophore and Glu46 are treated quantum chemically and the protein environment is approximated by an assembly of partial atomic charges. A feature of our method is that the MM charges are obtained from the linear-scaling molecular orbital calculations (known as MOZYME), which means that the polarizable effect of the protein environment is explicitly taken into account. The present results indicate that the inversion of the relative pK_a between the chromophore and Glu46 occurs on going from pG to pB, consistent with the corresponding experimental results.

Methods

Calculations of Medium Effects Using Continuum Solvent Models. *S*-Methyl *p*-coumarate and acetic acid (or propionic acid) were selected as model compounds of the chromophore and glutamic acid (Glu), respectively. The use of acetic acid (or propionic acid) seems to be reasonable because the experimental pK_a value (4.7) of acetic acid is almost equal to that (4.3) of Glu.[16] Figure 1 shows that these molecules exist without any interactions to each other. This noninteracting system will be called "System 1". Figure 2 shows other models where *S*-methyl *p*-coumarate and propionic acid are placed in accordance with their spatial arrangements in the X-ray structures of pG[10] and pB.[15] Hereafter, the models for pG and pB are called System 2G and System 2B, respectively.

To examine the nonspecific medium effect on the energies of these model systems, we carried out the solvent effect calculation using representative continuum solvent models, SCIPCM[17] and COSMO,[18] which are implemented in the Gaussian 98 program.[19] Then, the solvent dielectric constant (ϵ) was chosen to be 1, 2, 4, 10, 40, or 80. Such solvent effect calculations were carried out at the HF/6-31+G** or B3LYP/6-31+G** level.

Unless otherwise noted, the geometry of each compound in System 1 was obtained from the energy minimization calculation for the gas state and it was also used for the solvent effect calculations mentioned above. On the other hand, the geometries of Systems 2G and 2B were obtained from the results of linear-scaling molecular orbital calculations as described below.

Linear-Scaling MOZYME Calculations. First, the atomic coordinates of pG and pB were taken from the corresponding X-ray structures registered as 2PHY[10] and 2PYP[15] in the Protein Databank, respectively. Since these crystal structures lack the

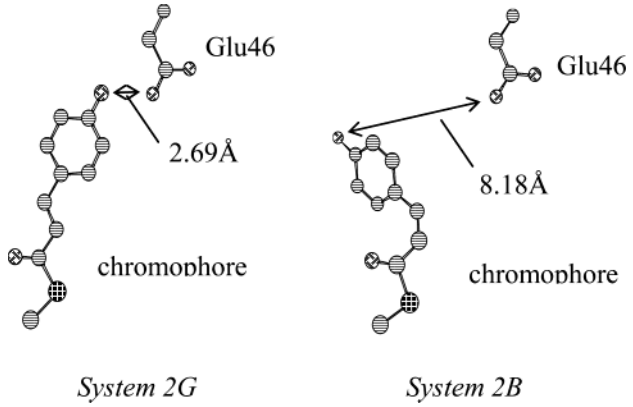

Figure 2. Structures of Systems 2G and 2B. *S*-Methyl *p*-coumarate and propionic acid were arranged in accordance with the spatial arrangement of the chromophore and Glu46 in the crystalline states of pG and pB. The geometric parameters of these compounds were determined by the MOZYME calculation (see Methods). The distances between the phenolate oxygen of the chromophore and the side chain oxygen of Glu46 were obtained from the X-ray structures of pG (2PHY) and pB (2PYP). The corresponding values after the geometry optimization by MOZYME are in Table 2.

coordinates of hydrogen atoms, we added them by using the LEaP module of the AMBER 4.1 program.[20] Then, for pG, the chromophore was deprotonated and Glu46 was protonated, and for pB the protonation states of these residues were inverted. Also, the resulting net charge of the entire protein was −6. Next, the positions of all the hydrogen atoms and the heavy atoms of the chromophore and the side chain of Glu46 were optimized by a linear-scaling molecular orbital method called MOZYME,[21] which is implemented in the MOPAC2000 program.[22] Then, the positions of the backbone heavy atoms were fixed to be the initial X-ray values. In the MOZYME calculations, the AM1 Hamiltonian[23] was used. Furthermore, to evaluate the relative pK_a values between the chromophore and Glu46, we carried out additional optimization calculations for hypothetical structures in which the protonation states of these residues are inverted with respect to those in the original pG and pB states. Consequently, for pG and pB, the two geometric structures corresponding to the two different protonation states of the protein were obtained from the MOZYME calculations. From these structures, we picked up the chromophore−Cys69 moiety and Glu46 to construct Systems 2G and 2B (Figure 2), where the $\alpha\angle\beta$ bonds in these fragments were cut and the β-positions were terminated with methyl groups. Then, the positions of these methyl hydrogen atoms were optimized by the AM1 level calculations.

Calculation of Relative pK_a. In solution, the equilibrium between a Brønsted acid AH and a base B

$$AH + B \leftrightarrow A^- + BH^+ \tag{1}$$

is governed by the free energy ΔG_{medium} [24]

$$\Delta G_{medium} = \Delta G + \Delta\Delta G_m \tag{2}$$

where ΔG corresponds to the free energy change of reaction 1 in vacuo. $\Delta\Delta G_m$ is given as follows:

$$\Delta\Delta G_m = \Delta G_m(A^-) - \Delta G_m(AH) + \Delta G_m(BH^+) - \Delta G_m(B) \tag{3}$$

where $\Delta G_m(X)$ represents the solvation free energy of species X; it can be evaluated by using the continuum approximations as described above. ΔG_{medium} (given in kJ mol^{-1}) can be related

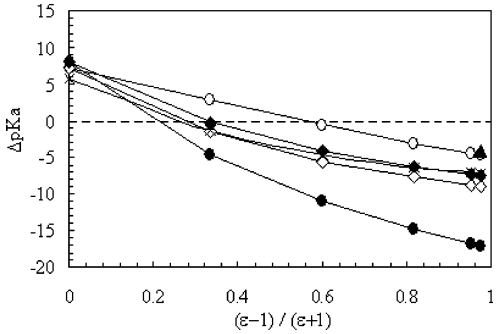

Figure 3. Dielectric constant dependence of the ΔpK_a of acetic acid to *S*-methyl *p*-coumarate in System 1. ΔpK_a was evaluated using eq 5, and it has positive sign when the pK_a of acetic acid is larger than that of pCAH. ○, HF(SCIPCM); ◇, HF(COSMO); ●, B3LYP(SCIPCM); ◆, B3LYP(COSMO). The symbols × indicate the results from the HF-(SCIPCM) calculation with the geometry optimized in each dielectric medium. The symbol ▲ shows the experimental ΔpK_a value in water.

to the pK_a of the acid AH through the thermodynamic equilibrium constant K of eq 1

$$\Delta G_{\text{medium}} = -2.303RT \log K = 5.71\ \text{p}K_a + \text{const} \quad (4)$$

where temperature T was taken to be 298.15 K, and "const" depends on the chemical nature of B. When AH is *S*-methyl *p*-coumarate (hereafter abbreviated as pCAH) or acetic acid (CH_3COOH) and B is a common base to them, the const term can be deleted by subtracting the equation of ΔG_{medium} for CH_3-COOH from that for pCAH. This procedure avoids the contamination of errors arising from evaluation of the const term. Consequently, the relative pK_a (ΔpK_a) of CH_3COOH to pCAH is obtained as follows:

$$\Delta\text{p}K_a = \frac{\Delta G_{\text{medium}}(\text{CH}_3\text{COOH}) - \Delta G_{\text{medium}}(\text{pCAH})}{5.71} =$$

$$\frac{G(\text{CH}_3\text{COO}^-) - G(\text{CH}_3\text{COOH}) - G(\text{pCA}^-) + G(\text{pCAH})}{5.71} +$$

$$\frac{\Delta G_{\text{m}}(\text{CH}_3\text{COO}^-) - \Delta G_m(\text{CH}_3\text{COOH}) - \Delta G_{\text{m}}(\text{pCA}^-) + \Delta G_{\text{m}}(\text{pCAH})}{5.71} \quad (5)$$

where G(X) is the free energy of species X in vacuo; it is obtained by adding thermodynamic corrections to the total energy of HF or DFT calculation for X. Such corrections were obtained according to an ordinary protocol of the Gaussian 98 program. Then, the HF/6-31+G**//HF/6-31+G** or B3LYP/6-31+G**//B3LYP/6-31+G** level of theory was used. As a result, the values of the thermodynamic corrections were determined to be −0.9 and 1.8 pK_a units from the HF and DFT calculations, respectively. By adding these contributions, the pK_a difference between pCAH and acetic acid in vacuo (the first term on the right-hand side of eq 5) was determined to be 7.3 from the HF calculation and 8.1 from the B3LYP one. In the next section, these values are implicitly used in the pK_a calculation based on eq 5.

Results

The pK_a Difference between Free Chromophore and Free Glutamic Acid. Figure 3 shows the dielectric constant dependence of ΔpK_a (eq 5) for System 1, where pCAH and acetic acid are infinitely separated. Irrespective of the computational methods used, two characteristic features are found: (1) ΔpK_a is positive in vacuo, that is, $(\epsilon - 1)/(\epsilon + 1) = 0$ for vacuum; (2) ΔpK_a almost linearly decreases with an increase in $(\epsilon - 1)/(\epsilon + 1)$, leading to negative ΔpK_a values in high dielectric media. These features indicate that in vacuo the pK_a of acetic acid is larger than that of the chromophore, but their relative order is inverted in high dielectric media like water ($\epsilon = 80$). Table 1 summarizes the data for $\Delta G_{\text{m}}(CH_3COO^-)$ and $\Delta G_{\text{m}}(\text{pCA}^-)$. The former more steeply decreases with increasing dielectric constant than the latter. Based on eq 5, this means that the pK_a of acetic acid is more steeply lowered than that of the latter, resulting in the tendency of item 2. Items 1 and 2 can also be interpreted from the size effect on electrostatic energy. We consider the Born model given by

TABLE 1: Results for ΔG_{m} (in kJ mol^{-1}) and ΔpK_a

	dielectric constant					
	1	2	4	10	40	80
			ΔG_{m}[a]			
S-methyl *p*-coumarate	0	−79.2	−96.8	−115.5	−125.4	−127.1
acetic acid	0	−104.3	−136.5	−170.8	−188.0	−190.9
			ΔpK_a[b]			
System 1	7.3	2.9	−0.6	−2.4	−4.4	−4.6
System 2G	0.7	−2.2	−4.0	−5.0	−5.6	−5.7
System 2B	7.2	1.8	−0.8	−2.2	−2.8	−2.9

[a] The dielectric constant dependence of the solvation energy of each solute in System 1. These data were obtained using the HF(SCIPCM) method. [b] The dielectric constant dependence of ΔpK_a's of the three systems shown in Figures 1 and 2. These data were obtained using the HF(SCIPCM) method.

$$\Delta G_m = \frac{1}{2}\left(\frac{1}{\epsilon_{\text{high}}} - \frac{1}{\epsilon_{\text{low}}}\right)\frac{Q}{R} \quad (6)$$

where Q and R are the charge and radius of a charged sphere, respectively, and ϵ_{high} and ϵ_{low} are high and low dielectric constants, respectively. This equation gives the free energy change associated with transfer of the charged particle from a low to high dielectric medium. Obviously, the smaller the radius is, the larger the magnitudes of the self-energy $[(1/2)Q/\epsilon R]$ and ΔG_{m} are. The negative charge of pCA^- (deprotonated *S*-methyl *p*-coumarate) is widely delocalized over the entire molecule through the π-conjugated system, while that of deprotonated acetic acid is localized on the carboxyl group. Thus, it is said that, at least in vacuo, pCA^- is more stable than deprotonated acetic acid, but in dielectric media the latter undergoes a larger stabilization by solvation.

In aqueous medium ($\epsilon = 80$), the ΔpK_a values obtained from HF(SCIPCM), HF(COSMO), B3LYP(SCIPCM), and B3LYP-(COSMO) are −4.6, −9.0, −17.1, and −7.6, respectively. On the other hand, the experimental pK_a values of pCAH and acetic acid in water are 8.8[9] and 4.7,[16] respectively, and thus the experimental ΔpK_a is −4.1. Thus, the result from the HF-(SCIPCM) calculation (symbol ○ in Figure 3) is in better agreement with the experimental one. As described under Methods, the above four types of calculations were performed using the geometries optimized for the gaseous solutes. It may be necessary to check how the solvent-induced geometric changes affect the calculated ΔpK_a values. For this purpose, we performed geometry optimization calculations for each solute of System 1 embedded in the medium of $\epsilon = 1$, 2, 4, 10, or 80. The ΔpK_a's were calculated using the resultant geometries. The results from the HF(SCIPCM) method are shown in Figure 3 (symbol ×). The resultant ΔpK_a value at $\epsilon = 80$ was −7.2.

Thus, the geometric effect is not so large compared with the result (○) for the in vacuo geometries.

Interestingly, when we adopted the SCIPCM model, the ΔpK_a vs $(\epsilon - 1)/(\epsilon + 1)$ curves were significantly affected depending on the theoretical level, that is, HF or B3LYP. Then, the overall ΔpK_a changes, on going from $\epsilon = 1$ to 80, obtained from these two methods were 12 and 25 pK_a units, respectively. On the other hand, when COSMO was used as a solvent model, the ΔpK_a vs $(\epsilon - 1)/(\epsilon + 1)$ curves were hardly affected by the theoretical levels and gave a ΔpK_a change ($\epsilon = 1 \rightarrow 80$) of ca. 16. Such different behavior between the SCIPCM and COSMO models might be caused due to the different definition of a cavity accommodating the solute molecule. In the SCIPCM model, the cavity surface is determined self-consistently from an isodensity surface (the density value was taken to be 0.0004 au in this study). Thus, the size and shape of the cavity depend on the theoretical level for the solutes (HF or B3LYP) and the dielectric constant used. In the COSMO method, the size and shape of the cavity are determined only from the solute geometry and atom types constructing the solute. Thus, the size and shape of the cavity do not depend on these factors.

In summary, the most important finding from Figure 3 is that the pK_a of pCAH (*S*-methyl *p*-coumarate) is smaller in low dielectric media than that of acetic acid, but their relative order is inverted in high dielectric media. The inversion of the relative pK_a occurs around $\epsilon = 2-4$.

Effect of the Interaction between the Chromophore and Glu46. To examine the effect of interaction between the chromophore and Glu46 on their relative pK_a values, we performed the following calculations for Systems 2G and 2B, where the side chain of Glu46 is replaced by propionic acid (Figure 2). First, eq 5 was rearranged as follows:

$$\Delta pK_a = \frac{(G(CH_3CH_2COO^-) + G(pCAH)) - (G(CH_3CH_2COOH) + G(pCA^-)}{5.71} + \frac{(\Delta G_s(CH_3CH_2COO^-) + \Delta G_s(pCAH)) - (\Delta G_s(CH_3CH_2COOH) + \Delta G_s(pCA^-))}{5.71} \quad (7)$$

From this equation, the relative pK_a of propionic acid relative to pCAH in their complex state can be written by

$$\Delta pK_a = \frac{(G(CH_3CH_2COO^- - pCAH)) - (G(CH_3CH_2COOH - pCA^-))}{5.71} + \frac{(\Delta G_s(CH_3CH_2COO^- - pCAH)) - (G_s(CH_3CH_2COOH - pCA^-))}{5.71} \quad (8)$$

where the first term is approximately evaluated from the total energies for the two protonated states of the complex in vacuo, and the second term represents the solvation energy difference between the two different protonation states. In the evaluation of eq 8, we did not take into account the contribution of the thermodynamic correction because it was expected to be small (see the result for the first term of eq 5 in the previous section). The second term of eq 8 was evaluated using the HF(SCIPCM) method because it gave the best result for the ΔpK_a of System 1. The results for Systems 2G and 2B are shown in Figure 4. As similar to the results for System 1, the ΔpK_a becomes smaller with increasing dielectric constant. For System 2G, the overall ΔpK_a change, on going from $\epsilon = 1$ to 80, is −6.4. The sign of the ΔpK_a was inverted from plus to minus at a low dielectric medium ($\epsilon = 1.2$). For System 2B, the ΔpK_a ($\epsilon = 1 \rightarrow 80$) is about 10, which is larger than that for System 2G and closer to that for System 1. This result is reasonable from the fact that pCAH is more largely separated from the propionic acid in System 2B than in System 2G (see Figure 2).

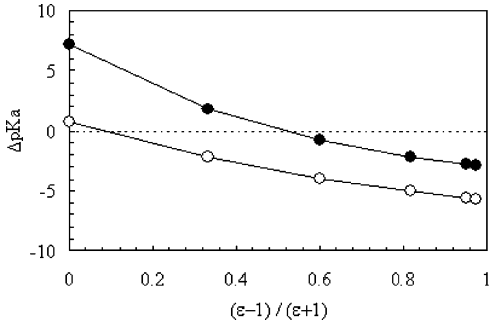

Figure 4. Dielectric constant dependence of the ΔpK_a of propionic acid to *S*-methyl *p*-coumarate in Systems 2G and 2B. ΔpK_a was evaluated using eq 8, and it has positive sign when the pK_a of propionic acid is larger than that of pCAH. All the calculations were carried out using the HF(SCIPCM) method. ○, System 2G; ●, System 2B.

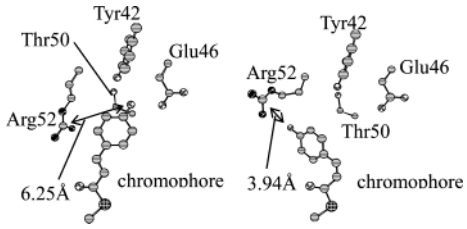

Figure 5. Hydrogen-bonding networks surrounding the chromophore in pG (left) and pB (right). Several characteristic atom−atom distances are summarized in Table 2.

TABLE 2: Several Atom−Atom Distances (Å) Pertinent to the Hydrogen-Bonding Network Surrounding the Chromophore

	hydrogen bond location			
states	pCA-Tyr42	pCA-Glu46	pCA-Thr50	pCA-Arg52
pG(exp)[a]	2.71	2.69	3.84	6.25
pG(A)[b]	2.73	2.68	3.85	6.26
pG(B)[b]	2.83	2.87	4.20	6.62
pB(exp)[a]	5.69	8.18	5.15	3.94
pB(A)[b]	4.71	7.72	4.83	3.39
pB(B)[b]	5.25	8.26	4.27	4.56

[a] Obtained from the X-ray structures of pG (2PHY) and pB (2PYP). [b] Obtained from the optimized geometries by the MOZYME calculations. "(A)" indicates that the chromophore is deprotonated and Glu46 is protonated, and "(B)" indicates the inverted protonation state (see Methods).

Effects of Protein Charge Distribution on ΔpK_a. The continuum approximations equalize the effects of orientational and electronic polarizations over the entire protein; hence they are incapable of appreciating the contributions of individual amino acid residues. In addition, System 2G is not necessarily a better model for the active site of PYP, because the chromophore in the protein is interacting not only with Glu46 but also with several polar residues as shown in Figure 5, where the chromophore forms a hydrogen-bonding network with Thr50, Tyr42, and Arg52. A similar situation is also found in the case of pB. In Table 2 the experimental atom−atom distances pertinent to this network are summarized together with the optimized values by the MOZYME calculations. In view

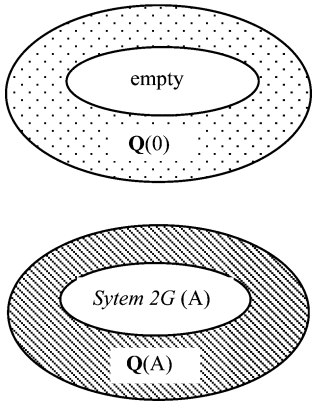

Figure 6. Schematic representation of the models assumed for the QM/MM calculations. The large and small ellipsoids represent the entire protein and the active site (for example, System 2G in this figure), respectively. The latter corresponds to region I described in the text, and the shaded region corresponds to region II. **Q**(0) and **Q**(A) represent the charge distributions over region II. The charge distribution **Q**(0) was obtained from the MOZYME calculation for a hypothetical mutant in which the chromophore and Glu46 were deleted. **Q**(A) was obtained from the MOZYME calculation for the entire protein under the condition that the protonation state of the active site is "(A)" type (see text). Similarly, **Q**(B) was obtained for the "(B)" type protonation state. In the text, we used the same nomenclature, **Q**(A, B, or 0), for the charge distribution of System 2B. In these calculations, the initial geometries of pG and pB were taken from the corresponding X-ray data (see Methods).

of these structural features of the active sites, the medium effect calculation should have been performed for these hydrogen-bonding cluster as solutes, but such calculation brought about severe convergence problems due to the large size and complicated shape of the solute parts.

Instead, we evaluated the ΔpK_a using another model system shown in Figure 6. Namely, the protein was partitioned into two regions: region I was composed of the chromophore and Glu46 and region II is the residual part of the protein. Region I was treated quantum chemically, and region II was replaced by a classical electrostatic model consisting of an atomic charge distribution. Region I is the same as System 2G or System 2B (Figure 2). For each of them, two different protonation states are allowed according to which residue, chromophore or Glu46, is protonated. Hereafter, "(A)" or "(B)" is used to discriminate the different protonation states: for example, System 2G(A) indicates that the chromophore is deprotonated and Glu46 is protonated, and System 2G(B) is the inverted protonation state with respect to these residues. To obtain the partial atomic charges of region II, we performed full atomic molecular orbital calculations for the entire protein in the pG and pB states (see Methods and Figure 6). Next, the partial charges assigned to the atoms of the chromophore and Glu46 were deleted from the charge distributions over the whole protein, because these residues are treated quantum chemically. The resultant atomic charge distributions are denoted as **Q**(A) and **Q**(B) for the above A and B types of protonation states.

On the other hand, to obtain the charge distribution of region II which is not perturbed from region I, we constructed a hypothetical mutant in which both Cys69 and Glu46 were replaced by Gly's. Then, we optimized all hydrogen atoms of this mutant using the MOZYME calculation and obtained a new atomic charge distribution **Q**(0) (see Figure 5). This mutant does not contain the chromophore and Glu46 residues, and thus **Q**(0) is independent of the electric perturbation arising from the protonation state changes in region I. Thus, the difference between **Q**(A) and **Q**(0) corresponds to the electronic reorganization arising from the presence of System 2G(A) (or System 2B(A)) and the difference between **Q**(B) and **Q**(0) is also regarded as the electronic reorganization due to System 2G(B) (or System 2B(B)).

TABLE 3: QM/MM Energies in the Presence of Different Charge Distributions in Region II

	energy/kcal mol^{-1}		
region I	**Q**(0)[a]	**Q**(A)[a]	**Q**(B)[a]
System 2G(A)	−4219.2	−4228.5	
System 2G(B)	−4210.3		−4224.5
System 2B(A)	−4175.7	−4190.0	
System 2B(B)	−4173.2		−4191.5

[a] The environments surrounding region I. Their physical meanings are described in the text and Figure 6.

TABLE 4: Results for Relative ΔpK_a Estimated from the Point Charge Approximation

	environments		
	in vacuo[a]	**Q**(0)[b]	**Q**(A)/**Q**(B)[c]
	System 2G		
ΔpK_a	8.3	6.5	2.9
$\Delta\Delta pK_a$	0	−1.8	−5.4
	System 2B		
ΔpK_a	11.9	1.8	−1.1
$\Delta\Delta pK_a$	0	−10.1	−13.0

[a] ΔpK_a calculated for the isolated state of System 2G or System 2B. [b] ΔpK_a calculated for System 2G (or System 2B) surrounded by the **Q**(0) charge distribution. The energy data were taken from the second column of Table 3. [c] ΔpK_a calculated for System 2G (or System 2B) surrounded by the **Q**(A) or **Q**(B) charge distribution according to its protonation states. The energy data were taken from the third and fourth columns of Table 3.

According to the usual QM/MM methodology, the electrostatic interactions between regions I and II were explicitly taken into account in the QM calculation of region I. The calculations were performed using the mopac97 program that was modified as to treat a background charge distribution.[26] Then, the AM1 Hamiltonian was used in accordance with the above MOZYME calculations. Such QM/MM calculations were carried out for the following combined systems: (System 2G(A) + **Q**(A)) and (System 2G(B) + **Q**(B)), (System 2G(A) + **Q**(0)), and (System 2G(B) + **Q**(0)). The corresponding calculations were also performed for System 2B. The total energies of these combined systems are summarized in Table 3.

The ΔpK_a values were obtained by using the first term on the right-hand side of eq 8, where the free energy G was replaced by the total QM/MM energy given in Table 3. In Table 4, the second column represents the data for Systems 2G and 2B in vacuo. The third and fourth columns list the ΔpK_a values evaluated in the presence of **Q**(0) and **Q**(A)/**Q**(B), respectively. Comparison between the second and fourth columns indicates that on going from in vacuo to the native protein environment (given by **Q**(A)/**Q**(B)), the ΔpK_a values for System 2G and System 2B decrease by 5.4 and 13.0, respectively. As a result, the ΔpK_a values for pG and pB are estimated to be 2.9 and −1.1, respectively. This indicates that in pG the chromophore is deprotonated and Glu46 is protonated, but in pB the inverted protonation states are realized.

Discussion

In this study, we focused on the relative pK_a of Glu46 to the chromophore rather than their absolute pK_a values, because most of the current pK_a calculation methods have not yet succeeded in reproducing the absolute pK_a of a molecule. Another reason

is that the information about the relative pK_a alone is sufficiently helpful for understanding the formation of the signaling state pB. The simplest models used here (Figures 1 and 3) indicated that in a low dielectric medium ($\epsilon < 2-4$) the chromophore model (pCAH) is deprotonated and acetic acid is protonated but in a high dielectric medium like water the inverted protonation states are realized. As explained with the aid of the Born model, such ΔpK_a inversion is essentially ascribed to the fact that the charge distribution of the chromophore is more delocalized than that of the acid species. According to Figure 4, the ΔpK_a inversion also occurs when the chomophore is interacting with Glu46. These results provide the universal tendency for the medium-dependent acid−base equilibrium between pCAH and its partner carboxylic acid.

The present QM/MM calculations successfully reproduced the ΔpK_a inversion occurring between pG and pB in the PYP photocycle. In view of the results from the above continuum model, such a drastic ΔpK_a change should be relevant to the change in polarity and/or polarizablity of the protein environment. According to the X-ray structure of pG,[10] the active site including the chromophore is protected from the surrounding solvent and it is embedded in a hydrophobic region of the protein interior. Thus, the QM/MM result for pG is consistent with the result of the continuum model for a low dielectric medium. On the other hand, the X-ray study indicated that the structural change on going from pG to pB is confined to the active site and its neighborhood. This fact does not allow us to intuitively interpret the drastic ΔpK_a change occurring in pB.[15]

To address this issue, we analyze the ΔpK_a from the following two contributions: electrostatic and polarizable contributions of the protein environment. The second column of Table 3 indicates the energies of region I, which receives the constant electric field coming from **Q**(0) irrespective of its protonation states, A or B. Thus, the ΔpK_a values obtained from these energy data, that is, the third column in Table 4, do not involve the effect of electronic polarization of the protein environment that should receive different electric fields according to the protonation state of region I. On the other hand, such an effect is explicitly taken into account in the ΔpK_a values given in the fourth column of Table 4, because **Q**(A) or **Q**(B) was properly used for the energy calculation of region I (the third and fourth columns of Table 3). Thus, the difference in ΔpK_a between the second (in vacuo) and third columns of Table 4 corresponds to the pure electrostatic contribution of the protein environment, and that between the third and fourth columns corresponds just to the polarization contribution. In the pG state, on going from the **Q**(0) environment to the native protein environment, the ΔpK_a changes by -3.6 ($=2.9 - 6.5$), which is larger in magnitude than the change (-1.8) on going from in vacuo to the **Q**(0) environment. Consequently, the polarization effect of region II exerts a large influence on the determination of the final ΔpK_a ($=2.9$). In the pB state, the polarization contribution is -2.9 ($= -1.1 - 1.8$), which is close to that in the pG state. However, the final ΔpK_a ($= -1.1$) is mainly determined by the ΔpK_a change on going from in vacuo to the **Q**(0) environment. An important finding from Table 4 is that the ΔpK_a inversion between pG and pB is mainly ascribed to the largely different contributions of the **Q**(0) environments. As is evident from the definition of **Q**(0), its value is determined only by the property of region II, especially its conformation. Taken together, the difference in ΔpK_a between pG and pB can be ascribed to the change in electrostatic nature arising from the conformational difference, even if it is small in the crystalline state. Accordingly, the photocycle can occur even in the crystalline state.

In the actual photocycle of PYP, the proton transfer from Glu46 to the chromophore occurs between pR and pB, where pR is an intermediate formed within a few nanoseconds after absorption of a photon.[3,4,25,26] In this regard, it may be more interesting to study the ΔpK_a change on going from pR to pB. However, it has been reported that the crystal structure for pR,[27] based on time-resolved Laue crystallography, includes serious errors.[28−32] This forced us to abandon further research for pR. According to a variety of experimental and theoretical studies,[33−37] pB exhibits larger thermal fluctuations and/or conformational changes than pG. As a result, solvent accessible surface is increased with the formation of pB.[33] In fact, our recent molecular dynamics simulation demonstrated that a water molecule enters the active site and forms a hydrogen bond with Glu46.[37] These results imply that in pB the dielectric constant of the surrounding the chromophore increases. Thus, it is reasonable from Figures 3 and 4 that the proton transfer from Glu46 to the chromophore occurs in the pB of solution state.

To interpret the pK_a behavior of PYP under physiological conditions, it is required to obtain the solution conformations of the protein and to use a pK_a calculation method that can handle the protein embedded in the aqueous medium. Recently, we have developed a precise pK_a calculation method based on the ONIOM methodology with an integration of MOZYME and DFT methods.[38] The pK_a changes occurring in the photocycle of bacteriorhodopsin have been successfully reproduced by the use of this method.[39] In our future work, it will be applied to PYP.

In conclusion, the slight conformation change from pG to pB in the crystalline state causes a substantial change in the microenvironment surrounding the chromophore, resulting in the ΔpK_a inversion between the chromophore and Glu46.

Acknowledgment. The authors thank the Computer Center, Institute for Molecular Science, Okazaki, Japan, for the use of the supercomputer system. We also thank the Computer Center, Tokyo Institute of Technology, for the use of the SGI Origin 2000 system.

References and Notes

(1) Mayer, T. E. *Biochim. Biophys. Acta* **1985**, *806*, 175.
(2) Sprenger, W. W.; Hoff, W. D.; Armitage, J. P.; Hellingwerf, K. J. *J. Bacteriol.* **1993**, *175*, 3096.
(3) Mayer, T. E.; Yakali, E.; Casanovich, M. A.; Tollin, G. *Biochemistry* **1987**, *26*, 418.
(4) Hoff, W. D.; Van Stokkum, I. H. W.; Van Ramesdonk, H. J.; Van Brederode, M. E.; McRee, D. E.; Casanovich, M. A. *Biophys. J.* **1994**, *67*, 1691.
(5) Van Beeumen, J. J.; Devreese, B. V.; van Bun, S. M.; Hoff, W. D.; Hellingwerf, K. J.; Meyer, T. E.; McRee, D. E.; Cusanovich, M. A. *Protein Sci.* **1993**, *2*, 1114.
(6) Hoff, W. D.; Düx, P.; Hård, K.; Devreese, B.; Nugteren-Roodzant, I. M.; Crielaard, W.; Boelens, R.; Kaptein, R.; Van Beeumen, J.; Hellingwerf, K. J. *Biochemistry* **1994**, *33*, 13959.
(7) Kroon, A. R.; Hoff, W. D.; Fennema, H. P. M.; Gijzen, J.; Koomen, G.-J.; Verhoeven, J. W.; Crielaard, W.; Hellingwerf, K. J. *J. Biol. Chem.* **1996**, *271*, 31949.
(8) Yoda, M.; Houjou, H.; Inoue, Y.; Sakurai, M. *J. Phys. Chem. B* **2001**, *105*, 9887.
(9) Baca, M.; Borgstahl, G. E. O.; Boissinot, M.; Burke, P.; Williams, D. R.; Slater, K. A.; Getzoff, E. D. *Biochemistry* **1994**, *33*, 14369.
(10) Borgstahl, G. E. O.; Williams, D. R.; Getzoff, E. D. *Biochemistry* **1995**, *34*, 6278.
(11) For example, see: Simonson, T. *Int. J. Quantum Chem.* **1999**, *73*, 45.
(12) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 1029.
(13) Bashford, D.; Gerwert, K. *J. Mol. Biol.* **1992**, *224*, 473.
(14) Demchuk, E.; Genick, U. K.; Woo, T. T.; Getzoff, E. D.; Bashford, D. *Biochemistry* **2000**, *39*, 1100.

(15) Genick, U. K.; Borgstahl, G. E. O.; Ng, K.; Ren, Z.; Pradervand, C.; Burke, P. M.; Srajer, V.; Teng, T.-Y.; Schildkamp, W.; McRee, D. E.; Moffat, K.; Getzoff, E. D.*Science* **1997**, *275*, 1471.

(16) *Encyclopedic Dictionary of Chemistry*, 1st ed.; Ohki, M., Ohsawa, T., Tanaka, M., Chihara, H., Eds.; Tokyo Kagaku Dozin Co. Ltd.: Tokyo, Japan, 1989.

(17) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. *J. Phys. Chem.***1996**, *100*, 16098.

(18) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799.

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratman, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*; Gaussian Inc.: Pittsburgh, PA, 1998.

(20) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham III, T. E.; Ferguson, D. M.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER*, ver. 4.1; University of California: San Francisco, CA, 1995.

(21) Stewart, J. J. P. *Int. J. Quantum Chem.* **1996**, *58*, 133.

(22) Stewart, J. J. P. *MOPAC2000*; Fujitsu Ltd.: Tokyo, Japan, 1999.

(23) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(24) Schüürmann, G. *J. Chem. Phys.* **1998**, *109*, 9523.

(25) Ujj, L.; Devanathan, S.; Meyer, T. E.; Cusanovich, M. A.; Tollin, G.; Atkinson, G. H. *Biophys. J.* **1998**, *75*, 406.

(26) Ohishi, S.; Shimizu, N.; Mihara, K.; Imamoto, Y.; Kataoka, M. *Biochemistry* **2001**, *40*, 2854.

(27) Perman, B.; Srajer, V.; Ren, Z.; Teng, T.-Y.; Pradervand, C.; Ursby, T.; Bourgeois, D.; Schotte, F.; Wulff, M.; Kort, R.; Hellingwerf, K.; Moffat, K. *Science* **1998**, *279*, 1946.

(28) Brudler, R.; Rammelsberg, R.; Woo, T. T.; Getzoff, E. D.; Gerwert, K. *Nat. Struct. Biol.* **2001**, *8*, 265.

(29) Xie, A.; Kelemen, L.; Hendriks, J.; White, B. J.; Hellingwerf, K. J.; Hoff, W. D. *Biochemistry* **2001**, *40*, 1510.

(30) Imamoto, Y.; Shirahige, Y.; Tokunaga, F.; Kinoshita, T.; Yoshihara, K.; Kataoka, M. *Biochemistry* **2001**, *40*, 8997.

(31) Imamoto, Y.; Mihara, K.; Tokunaga, F.; Kataoka, M. *Biochemistry* **2001**, *40*, 14336.

(32) Unno, M.; Kumauchi, M.; Sasaki, J.; Tokunaga, F.; Yamauchi, S. *Biochemistry* **2002**, *41*, 5668.

(33) Rubinstenn, G.; Vuister, G. W.; Mulder, F. A. A.; Dux, P. E.; Boelens, R.; Hellingwerf, K. J.; Kaptein, R. *Nat. Struct. Biol*. **1998**, *5*, 568.

(34) van Brederode, M. E.; Hoff, W. D.; van Stokkun, I. H. M.; Groot, M.-L.; Kroon, A. R.; Hellingwerf, K. J. *Biophys. J*. **1996**, *71*, 365.

(35) Hoff, W. D.; Xie, A.; van Stokkum, I. H. M.; Tang, X.-J.; Gural, J.; Kroon, A. R.; Hellingwerf, K. J. *Biochemistry* **1999**, *38*, 1009.

(36) Kandori, H.; Iwata, T.; Hendriks, J.; Maeda, A.; Hellingwerf, J. K. *Biochemistry* **2000**, *39*, 7902.

(37) Shiozawa, M.; Yoda, M.; Kamiya, N.; Asakawa, N.;Higo, J.; Inoue, Y.; Sakurai, M. *J. Am Chem. Soc.* **2001**, *123*, 7445.

(38) Ohno, K.; Kamiya, N.; Asakawa, N.; Inoue, Y.; Sakurai, M. *Chem. Phys. Lett.* **2001**, *341*, 387.

(39) Nakajima, S.; Ohno, K.; Inoue, Y.; Sakurai, M. *J. Phys. Chem. B* **2003**, *107*, 2867.