Study of the Opsin Shift of Bacteriorhodopsin: Insight from QM/MM Calculations with Electronic Polarization Effects of the Protein Environment

Hirohiko Houjou, Yoshio Inoue, and Minoru Sakurai*

Department of Biomolecular Engineering, Tokyo Institute of Technology, Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan

Received: September 14, 2000

We develop a hybrid quantum mechanical/molecular mechanical-configuration interaction (QM/MM-CI) method for calculating the absorption maxima of photoreceptor proteins such as bacteriorhodopsin. A unique point of our method, discriminating it from usual QM/MM methods, is that the ground-state electronic structure of the whole protein is first evaluated by a linear scaling-molecular orbital calculation. The resultant electronic distribution is utilized to construct a modified Fock matrix for subsequent CI calculation. In the excitation energy calculation, only the chromophore located at the photoactive center of a protein is treated quantum mechanically and the surrounding environment is approximated by classical electrostatics. Another feature of the method is that the classical region is instantaneously polarized in response to the excitation of the chromophore. This corresponds to the incorporation of electronic polarization effects of the protein part. To allow the polarization of amino acid residues, each bond of them is approximated by a cylindrical dielectric with a given polarizability. The polarization in the classical part is determined self-consistently. Here, the above method is applied to the wild type of bacteriorhodopsin (bR_{568}) and its mutants. It is revealed that their absorption maxima are not reproduced without taking into account the effect of electronic polarization of the protein part. In particular, the polarization of Trp86, Trp182, and Tyr185 plays a predominant role in causing a bathochromic shift in the absorption band of bR_{568}.

Introduction

Bacteriorhodopsin (bR) is a retinal protein that functions as a light-driven proton pump in the purple membrane of *Halobacterium salinarum*.[1] Illumination of the light-adapted state (bR_{568}) initiates a sequential photoreaction cycle consisting of spectroscopically distinct intermediates K, L, M, N, and O.[2] The chromophore *all-trans* retinal is bound to Lys216 via protonated Schiff base linkage. The absorption maximum (568 nm) in the light-adapted state shifts to red by ∼130 nm with respect to that (∼440 nm) for protonated retinylidene Schiff base (PRSB) in methanol.[1] Such a protein-induced bathochromic shift is known as opsin shift. The elucidation of its mechanism has been an important subject in photochemistry of retinal proteins for these decades.

The absorption maximum of the chromophore corresponds to its lowest $\pi-\pi^*$ excitation energy. In the ground state, a positive charge is localized mainly on the Schiff base nitrogen, and upon excitation, it shifts toward the ionone ring.[3] Accordingly, any perturbations which stabilize this charge delocalization lead to a smaller energy gap between the ground and excited states, resulting in a red shift of the absorption maximum. On the basis of this background, several mechanisms for the opsin shift of bR were proposed as follows: (1) elongation of the π-conjugated system due to the ring/chain coplanarization,[4−7] (2) π-electron delocalization due to weakening of the interaction between PRSB and its counterion,[8,9] and (3) interaction of the chromophore with polar or polarizable residues in the protein.[10] Among them, the third mechanism must be investigated in more detail for the construction of a reliable molecular model of the opsin shift.

Several model studies have pointed out the importance of polarizable nature of the protein matrix. Polarizable amino acids could stabilize the excited state of the chromophore by compensatory electronic movement, if they are located near the chromophore.[10,11] This model is supported by the fact that the absorption maximum of PRSB in ethanol shows a red shift by addition of phenol, indole, etc.[12] In proteins, aromatic residues (phenylalanine, tyrosine, and tryptophan) would play a role similar to such polarizable solvents.

In previous studies,[13,14] we have developed a self-consistent reaction field (SCRF) theory based on a polarizable continuum model (PCM), where the protein part of bR was treated as a continuous and homogeneous dielectric medium. In that theory, the effect of electronic polarization of the protein is represented as a function of refractive index. It was revealed that the absorption maximum of PRSB significantly red shifts with an increase in the refractive index of the surrounding medium, consistent with corresponding experimental data. In addition, we identified seven aromatic residues in the chromophore-binding pocket: Trp86, Trp138, Trp182, Trp189, Tyr83, Tyr185, and Phe208. According to the calculation, the opsin shift of bR_{568} and M_{412} both are quantitatively reproduced if these residues cooperatively act as a medium with a refractive index of 1.51. However, the PCM approximation equalizes polarizability over the whole protein, and hence is not capable of appreciating the contribution of individual residues. To clarify the molecular detail of the opsin shift, it is necessary to develop a computational method that can explicitly take into account both three-dimensional structure of the protein and electronic polarization effects of all amino acid residues involved.

* Corresponding author. E-mail: msakurai@bio.titech.ac.jp. Fax: 81-45-924-5827.

10.1021/jp0032863 CCC: $20.00
Published on Web 01/05/2001

Hybrid quantum mechanics and molecular mechanics (QM/MM) methods allow for the calculation of large systems.[15] In those methods, the region of interest is calculated quantum chemically, and the remainder of the system is treated using molecular mechanics. For QM/MM methods to be applied to problems of molecular excitation, they must be formulated so as to allow electronic rearrangement in the MM region, caused by excitation of the QM region. In an early landmark study, Warshel and Levitt developed a QM/MM method capable of taking into account MM polarization.[16] Luzhkov and Warshel refined this method for calculation of solvatochromic shifts of molecules in solution.[17] Recently, a similar method was formulated by Thompson and Schenter who applied it to examine the effects of the protein environment on the low energy excited states of the primary electron donor of the photosynthetic reaction center of *Rhodopseudomonas viridis*.[18]

On the other hand, recently developed linear-scaling semiempirical molecular orbital (LSMO) methods enabled us to obtain the electronic structure of the whole protein without introducing molecular mechanics approximation.[19−22] LSMO has been already applied to the various problems of protein chemistry.[23−27] Once the ground-state wave function is obtained, the excitation energy may be evaluated by configuration interaction (CI) calculation. However, the direct use of the CI calculation for the whole protein is actually difficult, because the CPU time increases in proportion to more than the square of the number of atoms. At present, the application of LSMO is limited to the ground state of a protein.

In this study, we develop a novel hybrid QM/MM method capable of calculating the absorption spectra of large molecules under the approximation nearly equivalent to full-atomic representation with the aid of LSMO calculation. Our method has some unique features that discriminate it from the conventional QM/MM methods. First, LSMO is utilized to obtain the electronic distribution of the whole protein in the ground state. The LSMO which we adopted here is MOZYME developed by Stewart.[19] Second, in the present QM/MM method, each bond constituting a protein is approximated as a cylindrical stick made of polarizable dielectric. Namely, the protein part is regarded as a mosaic composed of a variety of cylindrical-shaped dielectrics. Third, excitation energy is calculated based on a modified CI method, coupled with "polarizable mosaic model" approximation, which enables us to evaluate the electronic rearrangement of proteins in the excited state. It is shown that the present method well reproduces the opsin shift of bR. From analysis of the electrostatic potential map for the ground and excited states, we clarify the contributions from the fixed charges and electronic polarization of the protein to the opsin shift. Furthermore, the calculations for several site-directed mutants will reveal the role of each residue in the wavelength regulation of bR. On the basis of these results, we propose a model of the chromophore−protein interactions responsible for the opsin shift.

Theory

Expression of Excitation Energy for a Large Molecular System. First, we describe a theory for evaluating the excitation energy of a protein system composed of a chromophore and an apoprotein. In the present formulation, the system is partitioned into two regions: (I) a region mainly composed of the chromophore, and (II) the remainder of the protein. In what follows, it is assumed that during excitation, the geometry of the protein is unchanged, and that the region II itself is inactive to excitation light. Then, Hamiltonian of the system can be divided into three terms as follows:

$$H = H^{\mathrm{I}} + H^{\mathrm{II}} + V \tag{1}$$

where H^{I} and H^{II} are the Hamiltonian operators for the regions I and II, respectively, and V indicates the interaction between the regions I and II.

The wave function Ψ_n for the whole system is obtained by solving the Schrödinger equation given by

$$H|\Psi_n\rangle = E_n|\Psi_n\rangle \tag{2}$$

If the electron exchange between the regions I and II is negligible, Ψ_n is represented by the product of Ψ_n^{I} and Ψ_n^{II}, which are the wave functions for the regions I and II, respectively. As a consequence, the total energy E_n is described as follows:

$$E_n = \langle\Psi_n^{\mathrm{I}}|H^{\mathrm{I}} + V[\Psi_n^{\mathrm{II}}]|\Psi_n^{\mathrm{I}}\rangle + \langle\Psi_n^{\mathrm{II}}|H^{\mathrm{II}}|\Psi_n^{\mathrm{II}}\rangle \tag{3}$$

where $V[\Psi_n^{\mathrm{II}}]$ is the coulomb potential generated from the charge distribution in the regions II, namely

$$V[\Psi_n^{\mathrm{II}}] = \left\langle \Psi_n^{\mathrm{II}}(\mathbf{r}')\left|\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right|\Psi_n^{\mathrm{II}}(\mathbf{r}')\right\rangle \tag{4}$$

If we focus on the ground state, the energy E_0 is given by

$$\begin{aligned} E_0 &= \langle\Psi_0^{\mathrm{I}}|H^{\mathrm{I}} + V[\Psi_0^{\mathrm{II}}]|\Psi_0^{\mathrm{I}}\rangle + \langle\Psi_0^{\mathrm{II}}|H^{\mathrm{II}}|\Psi_0^{\mathrm{II}}\rangle \\ &= \langle\Psi_0^{\mathrm{I}}|H^{\mathrm{I}}|\Psi_0^{\mathrm{I}}\rangle + \langle\Psi_0^{\mathrm{I}}|V[\Psi_0^{\mathrm{II}}]|\Psi_0^{\mathrm{I}}\rangle + \langle\Psi_0^{\mathrm{II}}|H^{\mathrm{II}}|\Psi_0^{\mathrm{II}}\rangle \\ &= E_0^{\mathrm{I}} + \langle\Psi_0^{\mathrm{I}}|V[\Psi_0^{\mathrm{II}}]|\Psi_0^{\mathrm{I}}\rangle + E_0^{\mathrm{II}} \end{aligned} \tag{5}$$

Similarly, for the ith excited state, the energy E_i is given by

$$E_i = E_i^{\mathrm{I}} + \langle\Psi_i^{\mathrm{I}}|V[\Psi_i^{\mathrm{II}}]|\Psi_i^{\mathrm{I}}\rangle + E_i^{\mathrm{II}} \tag{6}$$

Consequently, the excitation energy ΔE_i ($= E_i - E_0$) is represented as follows:

$$\begin{aligned} \Delta E_i = E_i^{\mathrm{I}} - \Delta E_0^{\mathrm{I}} + \langle\Psi_i^{\mathrm{I}}|V[\Psi_i^{\mathrm{II}}]|\Psi_i^{\mathrm{I}}\rangle - \\ \langle\Psi_0^{\mathrm{I}}|V[\Psi_0^{\mathrm{II}}]|\Psi_0^{\mathrm{I}}\rangle + E_i^{\mathrm{II}} - E_0^{\mathrm{II}} \end{aligned} \tag{7}$$

Hereafter, the interactions between the regions I and II (the third and forth terms in eq 7), and the interactions inside the region II (the fifth and sixth terms in eq 7) are treated by quasiclassical electrostatics. In the ground state, the ath atom in the region I is regarded as a point charge with a value of Q_a^{I}, while the mth atom in the region II as a point charge with a value of Q_m^{II}. The atomic charges in the regions I and II are collected in column vectors $\mathbf{Q}^{\mathrm{I}}$ and $\mathbf{Q}^{\mathrm{II}}$. For example, Q_a^{I} is defined as

$$(\mathbf{Q}^{\mathrm{I}})_a = Q_a^{\mathrm{I}} \tag{8}$$

Upon excitation, the values of Q_a^{I} and Q_m^{II} both are changed. If we express the induced charges for the excited state Ψ_i as $q_a^{\mathrm{I},i}$ and $q_m^{\mathrm{II},i}$, the atomic charge distributions are given by $\mathbf{Q}^{\mathrm{I}}+\mathbf{q}^{\mathrm{I},i}$ and $\mathbf{Q}^{\mathrm{II}}+\mathbf{q}^{\mathrm{II},i}$ for the regions I and II, respectively. Unless otherwise noted, the superscript i for $\mathbf{q}^{\mathrm{I},i}$ and $\mathbf{q}^{\mathrm{II},i}$ will be omitted in the subsequent derivation.

To describe the interaction between atoms a and m (a and m belong to the regions I and II, respectively), we introduce a

new matrix **G**, whose (a, m) component is defined as follows:

$$G_{am} = \frac{1}{|\mathbf{r}_a - \mathbf{r}_m|} \tag{9}$$

where $\mathbf{r}_a$ and $\mathbf{r}_m$ are the position vectors of atoms a and m, respectively. Then, using eqs 4, 8, and 9, we can obtain the following expressions:

$$\langle \Psi_0^{\mathrm{I}} | V[\Psi_0^{\mathrm{II}}] | \Psi_0^{\mathrm{I}} \rangle = \mathbf{Q}^{\mathrm{I}}\mathbf{G}\mathbf{Q}^{\mathrm{II}} \tag{10}$$

$$\langle \Psi_i^{\mathrm{I}} | V[\Psi_i^{\mathrm{II}}] | \Psi_i^{\mathrm{I}} \rangle = (\mathbf{Q}^{\mathrm{I}} + \mathbf{q}^{\mathrm{I}})\mathbf{G}(\mathbf{Q}^{\mathrm{II}} + \mathbf{q}^{\mathrm{II}}) \tag{11}$$

Similarly, to describe the atom−atom pair interaction inside the region II (both m and n belong to the region II), we introduce a matrix **R**, whose (m, n) component is defined as follows:

$$R_{mn} = \frac{1}{|\mathbf{r}_m - \mathbf{r}_n|} \qquad (m \neq n) \tag{12a}$$

$$R_{mm} = 0 \tag{12b}$$

Then, the electrostatic energies of the region II in the ground state and excited state are approximately represented by

$$E_0^{\mathrm{II}} = \frac{1}{2}\mathbf{Q}^{\mathrm{II}}\mathbf{R}\mathbf{Q}^{\mathrm{II}} \tag{13}$$

$$E_i^{\mathrm{I}} = \frac{1}{2}(\mathbf{Q}^{\mathrm{II}} + \mathbf{q}^{\mathrm{II}})\mathbf{R}(\mathbf{Q}^{\mathrm{II}} + \mathbf{q}^{\mathrm{II}}) \tag{14}$$

Using eqs 10, 11, 13, and 14, eq 7 is rewritten as follows:

$$\Delta E_i = E_i^{\mathrm{I}} - E_0^{\mathrm{I}} + (\mathbf{Q}^{\mathrm{I}} + \mathbf{q}^{\mathrm{I}})\mathbf{G}\mathbf{q}^{\mathrm{II}} + \mathbf{q}^{\mathrm{I}}\mathbf{G}\mathbf{Q}^{\mathrm{II}} + \frac{1}{2}\mathbf{q}^{\mathrm{II}}\mathbf{R}\mathbf{q}^{\mathrm{II}} + \mathbf{q}^{\mathrm{II}}\mathbf{R}\mathbf{Q}^{\mathrm{II}} \tag{15}$$

This is the exact expression for the excitation energy on the assumption that the wave function of the protein can be described by the product of Ψ_n^{I} and Ψ_n^{II}.

In what follows, we explore an approximate expression of ΔE_i in the framework of the configuration interaction (CI) method. The column vector $\mathbf{q}^{\mathrm{I}}$ collects the excitation-induced atomic charges in region I. Of course, these charges must be obtained under the condition that the perturbation from the region II is present. As understood from eq 3, the wave function Ψ_n^{I} is obtained from the following Schrödinger equation:

$$(H^{\mathrm{I}} + V[\Psi_0^{\mathrm{II}}])|\Psi_n^{\mathrm{I}}\rangle = E_n'|\Psi_n^{\mathrm{I}}\rangle \tag{16}$$

where the electronic distribution of the region II is frozen as it is in the ground state. Then, if $\mathbf{Q}^{\mathrm{II}}$ is given in advance, the Fock matrix is modified in the following way:

$$F_{\mu\mu} = F_{\mu\mu}^0 + (\mathbf{G}\mathbf{Q}^{\mathrm{II}})_l \qquad (\phi_\mu \in \text{atom } l) \tag{17}$$

where ϕ_μ is the μth atomic orbital (AO), which belongs to the atom l. Under this condition, SCF calculation gives the ground-state energy written by

$$E_0' = E_0^{\mathrm{I}} + \mathbf{Q}^{\mathrm{I}}\mathbf{G}\mathbf{Q}^{\mathrm{II}} \tag{18}$$

where $\mathbf{Q}^{\mathrm{I}}$ can be obtained from the ground-state wave function Ψ_0^{I}. Because the charges of the region II are fixed, any excited states of the region I will receive the electrostatic potential originating from $\mathbf{Q}^{\mathrm{II}}$. After some mathematical manipulations according to the standard CI formalism, the ith diagonal term ($\Delta E_i'$) of the CI matrix is given by

$$\Delta E_i' = E_i^{\mathrm{I}} - E_0^{\mathrm{I}} + \mathbf{q}^{\mathrm{I},i}\mathbf{G}\mathbf{Q}^{\mathrm{II}} \tag{19}$$

where $\mathbf{q}^{\mathrm{I},i}$ is the atomic charge alteration for the ith excited-state configuration. After diagonalization of the CI matrix, a set of excitation energies close to that of $\Delta E_i'$ will be obtained. Comparing eq 19 with eq 15, it is found that such simple CI treatment does not give the correct excitation energy. The excitation energy will be obtained by correcting $\Delta E_i'$ for the charge alteration in the region II, induced by excitation of the region I.

$$\Delta E_i = \Delta E_i' + (\mathbf{Q}^{\mathrm{I}} + \mathbf{q}^{\mathrm{I},i})\mathbf{G}\mathbf{q}^{\mathrm{II},i} + \frac{1}{2}\mathbf{q}^{\mathrm{II},i}\mathbf{R}\mathbf{q}^{\mathrm{II},i} + \mathbf{q}^{\mathrm{II},i}\mathbf{R}\mathbf{Q}^{\mathrm{II}} \tag{20}$$

Therefore, the approximate value of ΔE_i can be obtained by only once diagonalizing the CI matrix, whose diagonal elements ($\Delta E_i'$) are modified according to eq 20. This modification of the CI matrix corresponds to the replacement of $V[\Psi_0^{\mathrm{II}}]$ in eq 16 by $V[\Psi_n^{\mathrm{II}}]$.

Charge Distribution of the Region II in the Ground State. As shown in eqs 17−20, the values of $\mathbf{Q}^{\mathrm{II}}$ must be determined before the evaluation of ΔE_i. In standard QM/MM methods, the charges corresponding to $\mathbf{Q}^{\mathrm{II}}$ are taken from molecular mechanics force field parameters or determined from results of molecular orbital calculations for small molecular fragments. In those methods, electron rearrangement is not allowed after the initial construction of a molecule. This may be a poor approximation for the charge distribution in a protein, because the electric fields from a charged chromophore or dissociated side chains significantly polarize the surrounding protein matrix.[25] Therefore, for the ground-state we performed full-atomic quantum chemical calculations to obtain the self-consistent charge distribution over the whole protein. The charge distribution in the excited state was obtained using a method described in the following section.

Polarizable Mosaic Model. At this stage, the value of q_m^{II} in eq 18 is still unknown. The vector $\mathbf{q}^{\mathrm{II}}$ represents the atomic charge rearrangements in region II, induced by excitation of the region I. It is thus necessary to explore a method of calculating $\mathbf{q}^{\mathrm{II}}$ from the information about the known quantity such as $\mathbf{q}^{\mathrm{I}}$.

We propose a new model, hereafter called "polarizable mosaic model (PMM)", that allows us to explicitly treat the electronic polarization of a protein matrix. In PMM, every covalent bond in the region II is replaced by a cylindrical stick made of dielectric, whose permittivity depends on the nature of an original bond. As a consequence, a protein molecule is represented by a *mosaic* consisting of cylinders with various polarizabilities. For simplicity, first a diatomic molecule is taken as an example. Atoms m and n are distant by d_{mn} and have charges of Q_m and Q_n, respectively. The cylindrical dielectric connecting these atoms has a polarizability of α_{mn}. When an electrostatic field **F** is applied to the dielectric, partial charges q_m and q_n ($= -q_m$) are induced on both edges of the cylinder, resulting in an induced dipole moment μ_{mn}. The magnitude of μ_{mn} is in proportion to the component of **F**, F_{mn}, along the bond direction.

$$\mu_{mn} = q_m d_{mn} = \alpha_{mn} F_{mn} \tag{21}$$

In first-order approximation, F_{mn} could be replaced by the

difference in electrostatic potential between the atoms *m* and *n*.

$$F_{mn} = -\frac{\phi_n}{\phi_m} d_{mn} \tag{22}$$

where ϕ_m is the electrostatic potential at the center of the atom *m*. Then, from eqs 21and 22, the following relationship is obtained:

$$q_m^{\mathrm{II}} = \frac{\alpha_{mn}}{d_{mn}^2}(\phi_n - \phi_m) \tag{23}$$

This expression is easily extended to the case of a multivalent atom.

$$q_m^{\mathrm{II}} = \sum_n \frac{\alpha_{mn}}{d_{mn}^2}(\phi_n - \phi_m) \tag{24}$$

where *n* runs over all of the atoms except for *m*, and α_{mn} ($m \neq n$) has a nonzero value only when the atom *n* binds to the atom *m*, otherwise $\alpha_{mn} = 0$. Here we define a matrix **A** and a column vector **F**, whose components are given by

$$A_{mn} = \frac{\alpha_{mn}}{d_{mn}^2} \qquad (m \neq n) \tag{25a}$$

$$A_{mm} = -\sum_n \frac{\alpha_{mn}}{d_{mn}^2} \tag{25b}$$

$$(\Phi)_n = \phi_n \tag{26}$$

Then eq 21 can be rewritten in matrix form as follows:

$$\mathbf{q}^{\mathrm{II}} = \mathbf{A}\Phi \tag{27}$$

Because $\mathbf{q}^{\mathrm{II}}$ is induced in response to $\mathbf{q}^{\mathrm{I}}$, the main source of the electrostatic potential in eq 27 is the induced charges $\mathbf{q}^{\mathrm{I}}$. In addition, $\mathbf{q}^{\mathrm{II}}$ must be corrected for the mutual polarization among the $\mathbf{q}^{\mathrm{II}}$ charges themselves. Namely, ϕ_m is determined by the induced charges and coordinates of all the particles other than the atom *m* itself

$$\phi_m = \sum_a G_{am} q_a^{\mathrm{I}} + \sum_n R_{mn} q_n^{\mathrm{II}} \tag{28}$$

Eq 28 is also rewritten in matrix form.

$$\Phi = \mathbf{G}^{\mathrm{t}}\mathbf{q}^{\mathrm{I}} + \mathbf{R}\mathbf{q}^{\mathrm{II}} \tag{29}$$

By substituting eq 29 into eq 27, we obtain the following matrix equation:

$$\mathbf{q}^{\mathrm{II}} = \mathbf{A}(\mathbf{G}^{\mathrm{t}}\mathbf{q}^{\mathrm{I}} + \mathbf{R}\mathbf{q}^{\mathrm{II}}) \tag{30}$$

Consequently, the induced charge $\mathbf{q}^{\mathrm{II}}$ is given by

$$\mathbf{q}^{\mathrm{II}} = (\mathbf{I} - \mathbf{A}\mathbf{R})^{-1}\mathbf{A}\mathbf{G}^{\mathrm{t}}\mathbf{q}^{\mathrm{I}} \tag{31}$$

where **I** is the unit matrix. Therefore, the charge alteration in region II is uniquely determined from the charge alteration $\mathbf{q}^{\mathrm{I}}$, generated by excitation of region I. Thus, we can obtain the vector $\mathbf{q}^{\mathrm{II}}$ necessary for evaluation of the excitation energy (eq 20).

Parametrization of Bond Polarizability. To estimate the polarizability of an arbitrary bonding pair of atoms, we use an empirical rule that is satisfied for the refractive indices of molecules in the condensed phase. The refractive index *n* and isotropic polarizability a_{iso} of a molecule are related to each other according to the Lorentz−Lorentz equation

$$\frac{n^2 - 1}{n^2 + 2}\frac{M}{\rho} = \frac{4\pi N_a \alpha_{iso}}{3} = P_M \tag{32}$$

where *M* and ρ are molecular weight and the density in the liquid state, respectively, and N_a is Avogadro's number. P_M is molar refraction, which can be decomposed into atomic contributions (P_n) as follows:

$$P_M = \sum_n P_n \tag{33}$$

Here we explore a way of decomposing the molar refraction into the contributions of individual bonds of a molecule. For an atomic pair (*m*,*n*), we newly define "molar bond refraction" (P_{m-n}) as follows:

$$P_{m-n} = \frac{P_m}{\upsilon_m} + \frac{P_n}{\upsilon_n} + P_\pi \tag{34}$$

where υ_m is the valence number counted only for σ-bonds of the atom *m*. The third term in eq 34 is added only when the atomic pair has a π-bond. In a π-conjugated system, the contribution of any π-bond is assumed to be equally shared to all of the σ-bonds taking part in the conjugation. Then, P_π is the net π-bond contribution per each bond.

For example, the molar bond refraction for indole is obtained as described below. According to literature data,[28] the contributions of H, C, and N (aromatic secondary amine) to the molar refraction are given by 1.100, 2.418, and 3.59 $\mathrm{cm^3\ mol^{-1}}$, respectively. In addition, the contribution of one double bond is 1.733 $\mathrm{cm^3\ mol^{-1}}$.[28] Then, the molar refraction P_{M} is calculated to be 37.57 $\mathrm{cm^3\ mol^{-1}}$. According to eq 34, $P_{\mathrm{C-H}}$ and $P_{\mathrm{N-H}}$ are given by 1.906 and 2.30 $\mathrm{cm^3\ mol^{-1}}$, respectively. For indole, it can be supposed that four π-bonds are shared by eight C−C bonds and two C−N bonds. Thus, the π-bond contribution assigned to each bond is $1.733 \times 4/10 = 0.693$ $\mathrm{cm^3\ mol^{-1}}$. As a result, $P_{\mathrm{C-C}}$ and $P_{\mathrm{C-N}}$ are given by 2.305 and 2.70 $\mathrm{cm^3\ mol^{-1}}$, respectively. The sum of the bond contributions is equal to 37.57 $\mathrm{cm^3\ mol^{-1}}$, in agreement with the value of P_{M}. Table 1 summarizes the P_{m-n} values for all the types of bonds that constitute a protein.

On going back to eq 32, we can obtain the isotropic polarizability $a_{mn,\mathrm{iso}}$ for each *m*−*n* bond.

$$\alpha_{mn,iso} = \frac{3P_{m-n}}{4\pi N_a} \tag{35}$$

On the assumption that every bond has a cylindrical shape, the isotropic polarizability is given by

$$\alpha_{iso} = \frac{\alpha_{\parallel} + 2\alpha_{\perp}}{3} \tag{36}$$

where $\alpha_{\parallel}$ and $\alpha_{\perp}$ are the polarizabilities along the directions parallel and perpendicular to the bond, respectively. In general, the value of $\alpha_{\perp}$ is fully smaller than that of $\alpha_{\parallel}$. Here, it is assumed that the value of $\alpha_{\perp}$ equals 0. Thus, $\alpha_{\parallel} = 3\alpha_{\mathrm{iso}}$.

TABLE 1: List of Molar Bonding Refraction (P_M/cm^3 mol^{-1})

	H	C(sp^3)	C(sp^2)	C(benzene)	C(indole)	N(indole)	S
H	2.20	1.70	1.91	1.91	1.91	2.30	4.95
C (sp^3)	1.70	1.21	1.41	1.41	1.41		4.59
C (sp^2)	1.91	1.41	1.61	1.62	1.62		4.79
N (prim.amine)	1.92	1.42	1.62				
N (prim.amide)	1.98		1.69				
N (sec.amine)	1.86	1.36	1.56				
N (tert.amide)		1.51	1.71				
O (alcohol)	1.86	1.37	1.57	1.57			
O (ester)	1.92	1.42	1.63				
S	4.95	4.59	4.79				8.11
Multiple Bond							
C (sp^2)			3.34				
C (benzene)				2.48			
C (indole)					2.31	2.70	
N (imide)			4.43				
O (carbonyl)			4.75				
O (carboxyl)			3.19				

Consequently, the following expression is obtained:

$$\alpha_{mn,\mathrm{II}} = \frac{9P_{m-n}}{4\pi N_a} \tag{37}$$

Bond polarizabilities obtained from eq 37 are used to evaluate the electrostatic interactions among the cylindrical dipoles. In view of the physical meaning of eq 31, the **AR** part of the matrix $(\mathbf{I} - \mathbf{AR})^{-1}$ is responsible for the dipole−dipole interactions in region II. On the other hand, the **AG** part in eq 31 represents the charge-dipole interactions between regions I and II. If the dipole−dipole interaction is sufficiently weaker than the charge−dipole interactions, $(\mathbf{I} - \mathbf{AR})^{-1}$ is approximately equal to **I**. Then, $(\mathbf{I} - \mathbf{AR})^{-1}\mathbf{A}$ in eq 31 can be replaced by an effective polarizability matrix $\mathbf{A}_{\mathrm{eff}}$

$$\mathbf{q}^{\mathrm{II}} = \mathbf{A}_{\mathrm{eff}}\mathbf{G}^{\mathrm{t}}\mathbf{q}^{\mathrm{I}} \tag{38}$$

where the components of $\mathbf{A}_{\mathrm{eff}}$ are given by substituting $\alpha_{mn,\mathrm{II}}$ into a_{mn} in eqs 25a−c.

$$(\mathbf{A}_{\mathrm{eff}}) = \frac{\alpha_{mn,\mathrm{II}}}{d_{mn}^2} \qquad (m \neq n\text{:bonding}) \tag{39a}$$

$$(\mathbf{A}_{\mathrm{eff}})_{\mathrm{mn}} = 0 \qquad (m \neq n\text{:nonbonding}) \tag{39b}$$

$$(\mathbf{A}_{\mathrm{eff}})_{mm} = -\sum_n \frac{\alpha_{mn,\mathrm{II}}}{d_{mn}^2} \tag{39c}$$

In this study, we use eq 38 for evaluation of $\mathbf{q}^{\mathrm{II}}$, instead of eq 31.

Calculation

The computational procedure is shown in Figure 1. First, by using the MOPC2000 program, the full atom MO calculation was performed for a given geometry of bR (see below). Next, the coordinate of the protein was divided into *all-trans*-retinylidene ethylammonium ion (region I) and bacterioopsin (apoprotein: region II) in which the side chain of Lys216 was removed. At this stage, all of the atomic charges in the region II were stored in vector $\mathbf{Q}^{\mathrm{II}}$. Subsequently, Hartree−Fock calculation of region I was performed, taking into account the electrostatic interaction with the external charges ($\mathbf{Q}^{\mathrm{II}}$). For this calculation we used an INDO/S program, with Fock matrix elements modified so as to include the electrostatic potential from the external charges (see eq 17). In addition, the diagonal elements of the CI matrix were also modified according to eq 20. The elements of $\mathbf{A}_{\mathrm{eff}}$ were calculated by using the eqs 34, 35, and 39, after the types of all the atoms were specified according to Table 1. Finally, the modified CI matrix was diagonalized, providing a set of the excitation energy including the effects of static and induced charges of the protein matrix.

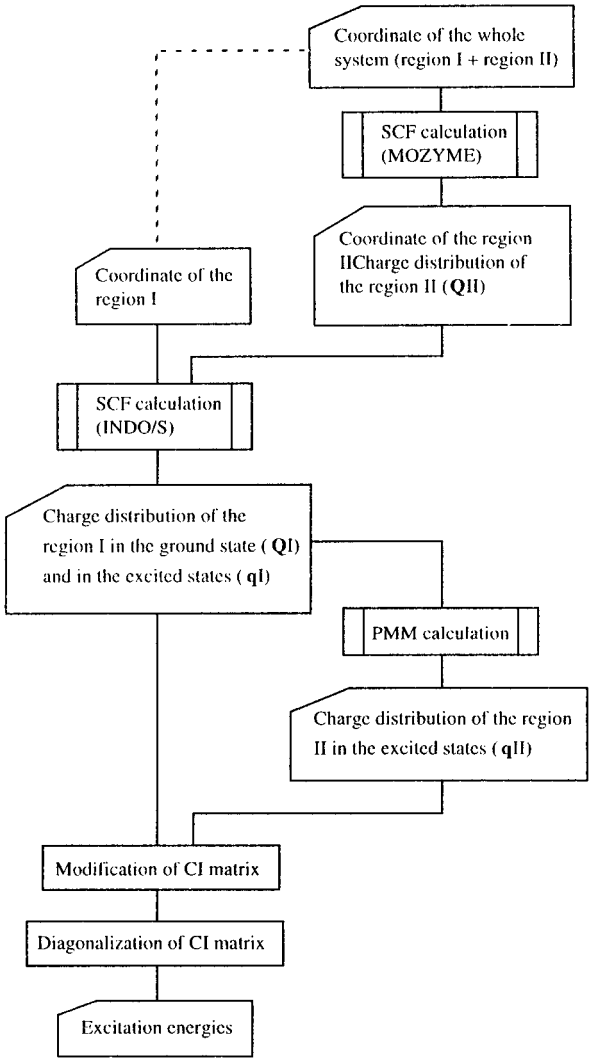

Figure 1. Flowchart of the whole computational procedure of the present QM/polarizable-MM model.

The geometry of bR was taken from the experimental structure registered as 2BRD in the protein data bank.[30] Hydrogen atoms were added to the skeleton by using the LEaP module of the AMBER4.1 program,[31] and then their positions were optimized by the MOPAC2000 program. The resulting structure is hereafter abbreviated as "WT". Mutants of bR studied are W86A, W138A, W182A, W189A, Y83A, Y185A, F208A, W86A/W182, W86A/Y185A, and W86A/W182A/Y185A. The atomic coordinates were obtained by optimizing only the geometry of the methyl group (Ala side chain) at the mutated position. In a similar way, the K216G mutant and the acidic form of bR (the carboxyl group of Asp85 is protonated: denoted as A-bR) were prepared. In all the MOZYME calculations, the AM1 Hamiltonian[32] was used and cutoff distance for short-range interactions (the keyword CUTOF2) was chosen to be 7.0 Å. Below that distance, the interaction between two atoms is represented by the exact AM1 approximation.

The electrostatic potentials, i.e., eqs 9 and 12a, were actually calculated using the Nishimoto−Mataga equation for two center Coulomb integral,[33] instead of the pure classical expression. Such a treatment was needed to smoothly extrapolate the results of region I (quantum mechanical region) into those of region II (classical region). In drawing electrostatic potential maps, a more simple expression, Qr^{-1} type potential, was used, where Q is a Mulliken charge obtained from MO calculations.

Results

Electrostatic Potential Map for the Chromophore-Binding Pocket of bR. Figures 2(a) and (b) show the molecular coordinate system used for analysis of electrostatic potential maps for bR. This coordinate system was defined as follows: the C10 atom was set at the origin; the C10−C11 bond was set on the *x*-axis; an *x*−*y* plane was taken so as to include the atoms C9, C10, and C11. As a result, this plane (Figure 2(a)) almost coincides with the conjugated plane of PRSB. Figures 3(a) and (b) show electrostatic potential maps drawn by using the atomic charges obtained from the MOZYME calculation for WT. The potential in these figures (given in a.u.) involves only the contributions from the atomic charges of the region II. The features of the map on the conjugated plane (Figure 3(a)) are summarized as follows: the potential value varies from −0.3 to 0.0 au; there are three apparent minima around (4.0, −7.0), (0.0, −4.0), and (3.0, 5.0); the region just including the chromophore has a relatively small absolute value of potential; the potential gradient is relatively small in that region, especially around the ionone ring. The deepest minimum around (4.0, −7.0) originates from the negative charge of ionized Asp85, which acts as the counterion of the chromophore. The minima around (0.0, −4.0) and (3.0, 5.0) originate from the presence of Trp86 and Trp182, respectively, which are arranged so as to surround the chromophore from its lateral sides. It is worth noting that the potential value gradually decreases on going from the ionone ring to the Schiff base linkage. The resulting potential gradient would localize the positive charge of the chromophore toward the Schiff base moiety. This is inconsistent with the well-known external charge model,[34] which states that the electrostatic effect of the protein would cause a delocalization of the positive charge toward the ionone ring.

Validity of the Polarizable Mosaic Model. To assess the reliability of the PMM approximation, it is sufficient to observe the response of a simple "mosaic" system to some fixed charges which correspond to $\mathbf{q}^{\mathrm{I}}$ in the region I. First, we examine the charge distribution of a hypothetical system in Figure 4, where a unit positive charge (regarded as $\mathbf{q}^{\mathrm{I}}$) induces the electronic

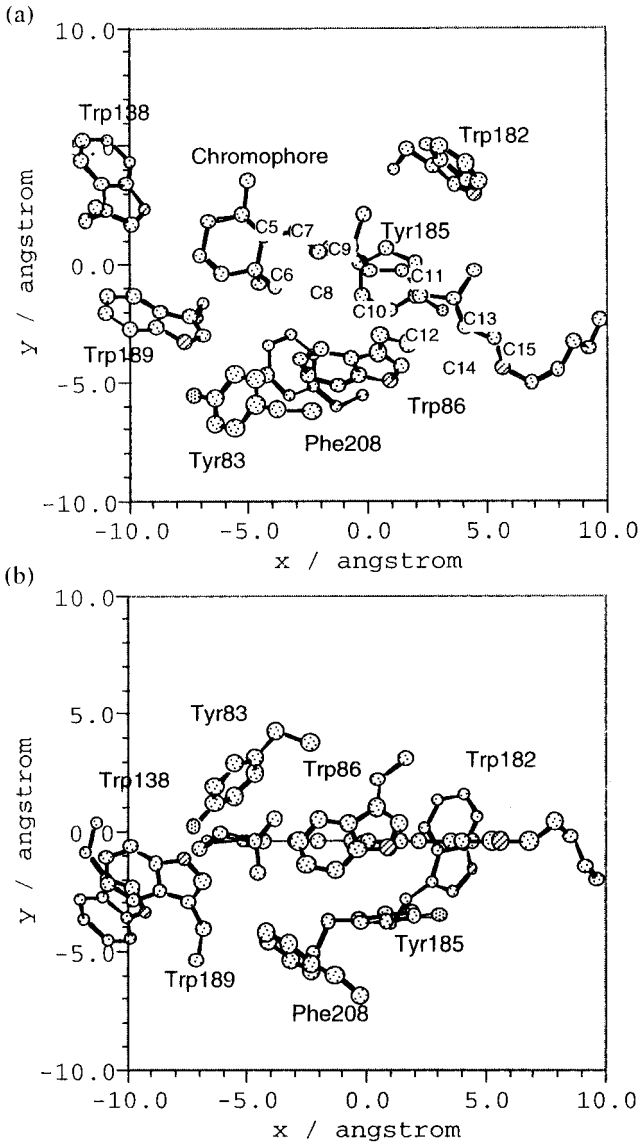

Figure 2. Molecular coordinate systems used for analysis of electrostatic potential maps. The C10 atom was set at the origin; the C10−C11 bond was set on the *x*-axis; an *x*−*y* plane is taken so as to include the atoms C9, C10, and C11. (a) and (b) show the *x*−*y* and *x*−*z* planes, respectively. The aromatic residues located in the chromophore-binding pocket are also shown.

polarization of an indole molecule, which is approximated by PMM. The conjugated plane of the indole molecule was placed on an *x*−*y* plane. The unit positive charge was placed at the coordinate (5.0, 0.0, 0.0) or (0.0, 5.0, 0.0). These two cases are hereafter called cases 1 and 2, respectively.

The unit charge serves as $\mathbf{q}^{\mathrm{I}}$ in eq 38. After evaluation of the matrices $\mathbf{A}_{\mathrm{eff}}$ and $\mathbf{G}$, we obtained the charge distribution corresponding to $\mathbf{q}^{\mathrm{II}}$. Figures 5(a) and (b) show maps of electrostatic potential (given in 10^{-3} a.u.) generated from $\mathbf{q}^{\mathrm{II}}$ for the cases 1 and 2, respectively. In these figures, the maps are drawn on the plane of $z = 5.0$Å, and the position of the point charge is indicated as "*". These maps clearly show that the electrons of the indole molecule were withdrawn toward the positive charge.

These results were compared with those from explicit quantum chemical calculations for the same geometrical configurations as in Figure 4. The AM1 calculation was performed for the whole system and an isolated indole molecule, which was the reference to measure the charge rearrangement $\mathbf{q}^{\mathrm{II}}$

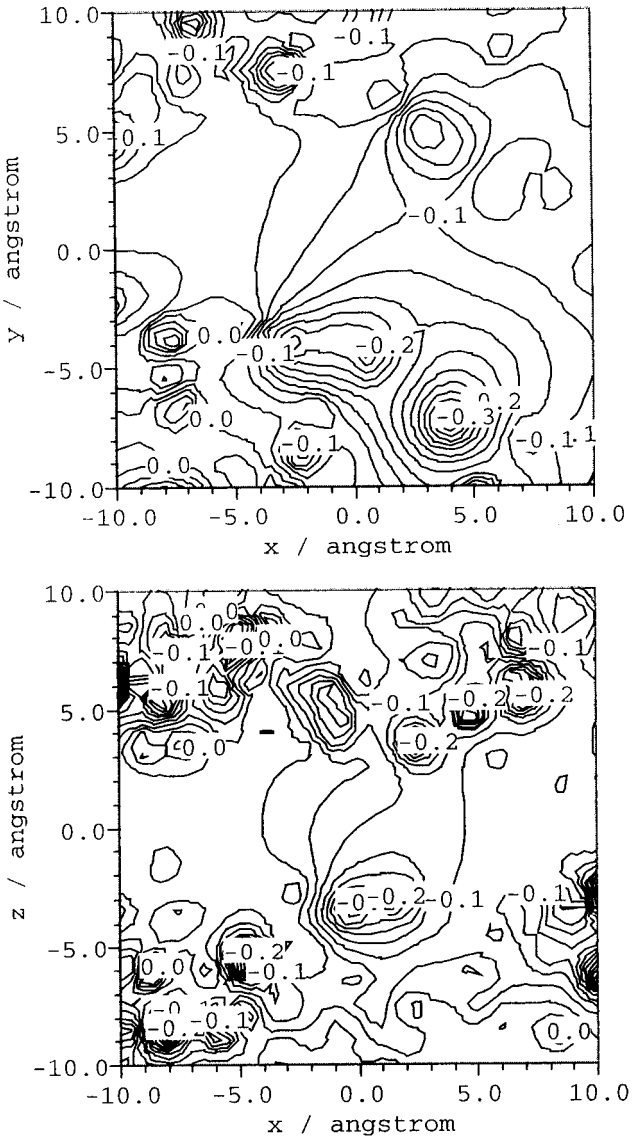

Figure 3. Electrostatic potential maps of the chromophore-binding pocket in bR in the ground state. These were drawn on the plane shown in Figures 2(a) and 2(b).

induced by the point charge. Namely, the induced charge on each atom was obtained by subtracting the atomic charge of the isolated indole molecule from that of the indole−charge complex, and was used to draw an electrostatic potential map. Figures 5(c) and (d) show the results for the cases 1 and 2, respectively. As can be seen, Figures 5(c) and (d) are quite similar to Figures 5(a) and (b), respectively. This indicates that PMM is sufficiently accurate for evaluating the electronic polarization in a molecule, induced by an external electrostatic field.

Next, we show results for bR, where the positive charges of the chromophore (region I) induce an electronic polarization of the protein. The atomic charges of the chromophore were regarded as $\mathbf{q}^{\mathrm{I}}$ in eq 38. These were obtained from the full-atomic calculation for WT. The matrix $\mathbf{A}_{\mathrm{eff}}$ was again constructed by applying the PMM approximation to the apoprotein of WT. Figure 6(a) shows an electrostatic potential map drawn by using the resulting induced charges ($\mathbf{q}^{\mathrm{II}}$). The map shows a broad minimum around (5.0, −5.0), where the protonated Schiff base linkage of the chromophore is placed. Clearly, this minimum arises from the positive charge of the chromophore.

Figure 6(a) was compared with the corresponding map

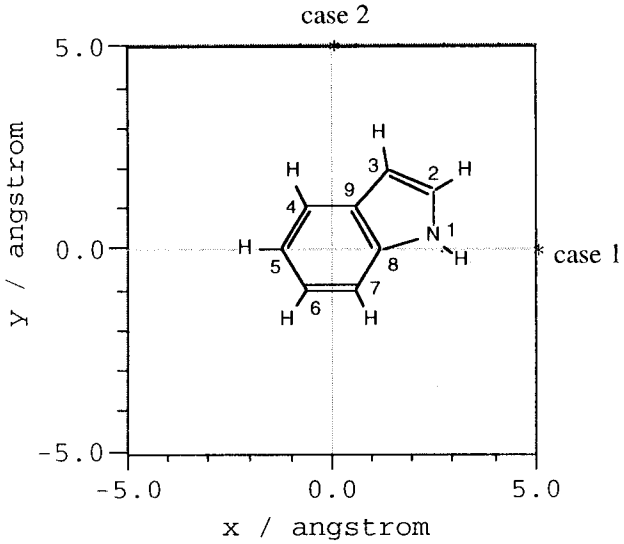

Figure 4. Molecular coordinate system for analysis of electrostatic potentials of indole-point charge systems. The midpoints of the C5 and C8 atoms were set at the origin; C5 and C8 atoms were set on the *x*-axis; an *x*−*y* plane is taken so as to include the atoms C4, C5, and C8. All of the conjugated carbon and nitrogen atoms were placed on the plane of $z = 0.0$ Å. The position of the unit charges in cases 1 and 2 are indicated by an asterisk (*).

obtained from the full-atomic quantum chemical calculation using MOZYME. The binding of the chromophore induces $\mathbf{q}^{\mathrm{II}}$ charges, which can be obtained by subtracting the charge distribution of K216G mutant from that of WT. Figure 6(b) shows an electrostatic potential map drawn by using the induced charges in region II. This figure is quite similar to Figure 6(a) in the landscape of the potential surface. However, a deviation of about 0.01 au is found for the depth of the potential well between Figures 6(a) and (b). At any positions on the *x*−*y* plane, the deviation is not more than 10% of the potential value observed for WT (see Figure 3). These results greatly encourage us to apply the PMM approximation to large molecules.

The potential value shown in Figure 6(a) varies from −0.06 to 0.00 au. The change of 0.06 au nearly amounts to 30% of the original potential observed for WT (see Figure 3). Therefore, the effect of electronic polarization of the protein is not negligible. This clearly shows the necessity of explicitly evaluating the electronic distribution in the ground state.

The Excitation Energy of bR and Its Mutants. In the present formalism, two types of approximation are possible for evaluating the excitation energy of a protein: (A) the atomic charges of an apoprotein are fixed during excitation (i.e., $\Delta E_i'$ eq 19), and (B) they are altered according to eq 34 (i.e., ΔE_i eq 20). The calculated absorption maxima based on these approximations are summarized in Table 2. The values from the approximations (A) and (B) were 465.3 and 534.9 nm, respectively. The latter is 69.6 nm longer than the former, which indicates that the electronic polarization of the apoprotein opsin induces a significant amount of red shift. Clearly, the calculation with the approximation (B) well reproduces the experimental absorption maximum (568 nm) of bR. Table 2 also shows the absorption maximum of the chromophore−Asp85 complex, where only the effect of counterion is taken into account. It is found that the counterion brings about a blue shift.

The vector $\mathbf{q}^{\mathrm{I}}$ indicates the charge rearrangements in the chromophore on going from the ground to the lowest $\pi-\pi^*$ excited state, and is obtained by subtracting the atomic charges for the ground-state configuration from those for the excited-state configuration. The vector $\mathbf{q}^{\mathrm{II}}$ collects the charges induced

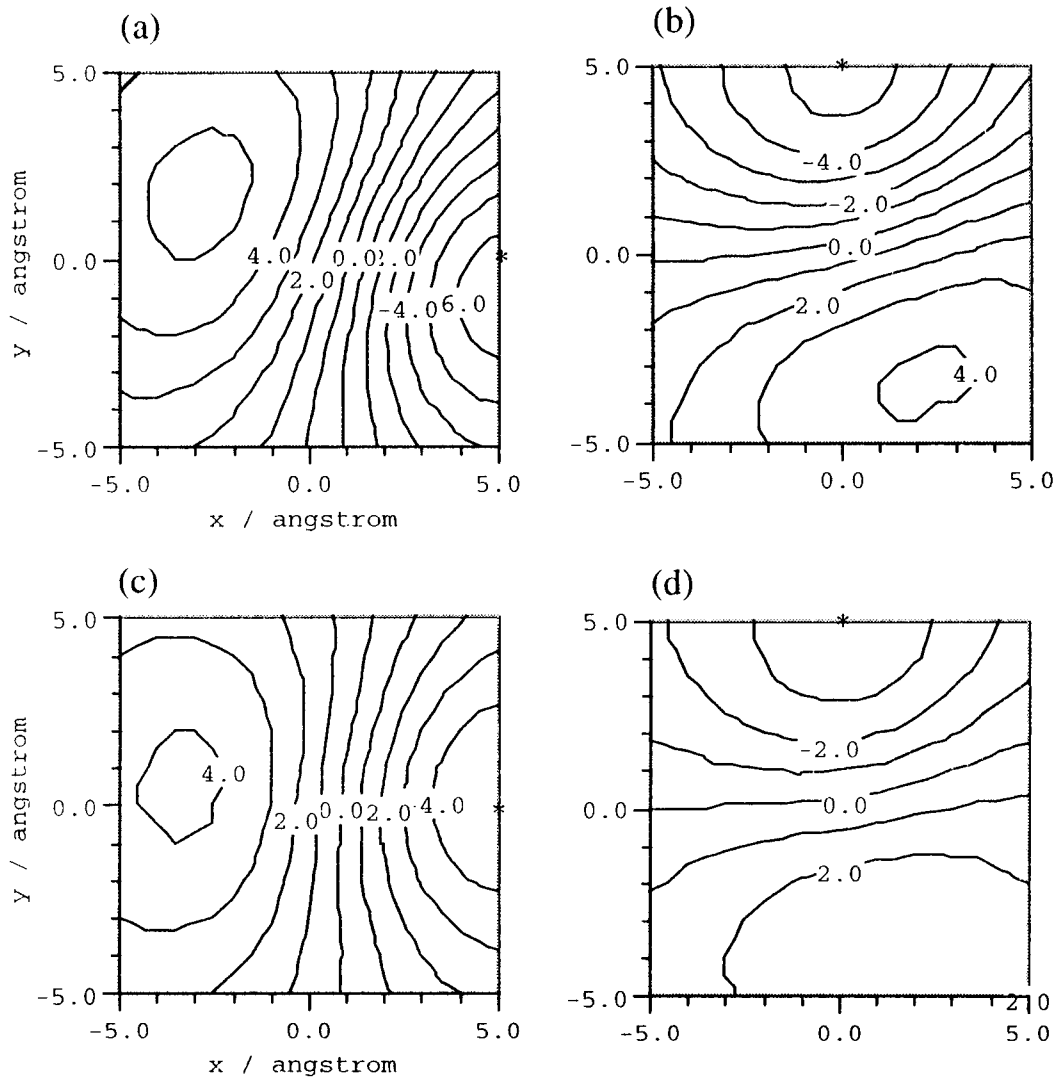

Figure 5. Electrostatic potentials generated from the charges induced on the indole molecule in the systems of Figure 4. The position of the unit charge is indicated by an asterisk (*). The induced charges, corresponding to $\mathbf{q}^{II}$, in (a) and (b) were calculated by the PMM approximation, and those in (c) and (d) were obtained by AM1 calculations. All potential maps were drawn on the plane of $z = 5.0$ Å.

TABLE 2: Comparison between the Calculated and Experimental Absorption Maxima of PRSB and bR

	λ_{max}/nm	
system	calcd.	exptl.
PRSB	524.9[a]	
PRSB + counterion	474.4[b]	465[c]
BR	465.3[d]	
	534.9[e]	568[f]

[a] Result for protonated all-*trans*-6s-trans-retinal Schiff base in vacuo. [b] Result for (protonated all-*trans*-6s-trans-retinal Schiff base + deprotonated Asp85) in vacuo. [c] Result for protonated all-*trans*-6s-trans-locked-retinal Schiff base. The data was taken from ref 4. [d] The absorption maximum based on approximation (A). [e] The absorption maximum based on approximation (B). [f] Result for bR in the light-adapted state. The data were taken from ref 1.

in the protein matrix on excitation. An electrostatic potential map produced by using $\mathbf{q}^{II}$ is shown in Figure 7(a), where the $x-y$ plane is the same as shown in Figure 2(a). As can be seen from this figure, there are two sets of pairs of a maximum and a minimum: one is centered around (−1.0, −4.0) and the other around (3.0, 5.0). Such a pair-potential formation implies the generation of a dipole-like field. By comparison of Figure 7(a) with Figure 2(a), it is found that the C11−C15 atoms of PRSB exist in a region with positive potentials, while the C5−C10 atoms are in a region with negative potentials. Such a characteristic polarization in the protein matrix of course reflects a charge rearrangement in PRSB induced by its excitation. Figure 8 shows the value of $\mathbf{q}^{I}$ for each of the conjugated carbons of the chromophore. Upon excitation, the charges on C11, C13, C15 largely decrease, while those on C5−C8, C10 and C12 increase. Consequently, in the excited state, the chromophore polarizes the surrounding protein matrix, resulting in the generation of a dipole-like field antiparallel to the dipole moment of the chromophore itself. As a result, the excited-state of the chromophore is stabilized. This explains why the absorption maximum based on the approximation (B) largely red shifts relative to that based on the method (A).

By comparing Figure 7(a) with Figure 2(a), Trp86 and Trp182 are identified as the possible sources of the dipole-like potential in the protein. In addition, from the vertical cross-section of the map (see Figures 2(b) and 7(b)), Tyr185 also seems to contribute to the formation of the dipole-like potential. In a region within 5 Å distant from the chromophore, there are seven aromatic residues: Trp 86, Trp138, Trp182, Trp189, Tyr83, Tyr185, and Phe208. The side chains of these residues have relatively large electronic polarizabilities. We visualized the distribution of $\mathbf{q}^{II}$ over the whole protein by using color code (data not shown). As a result, it is revealed that the residues

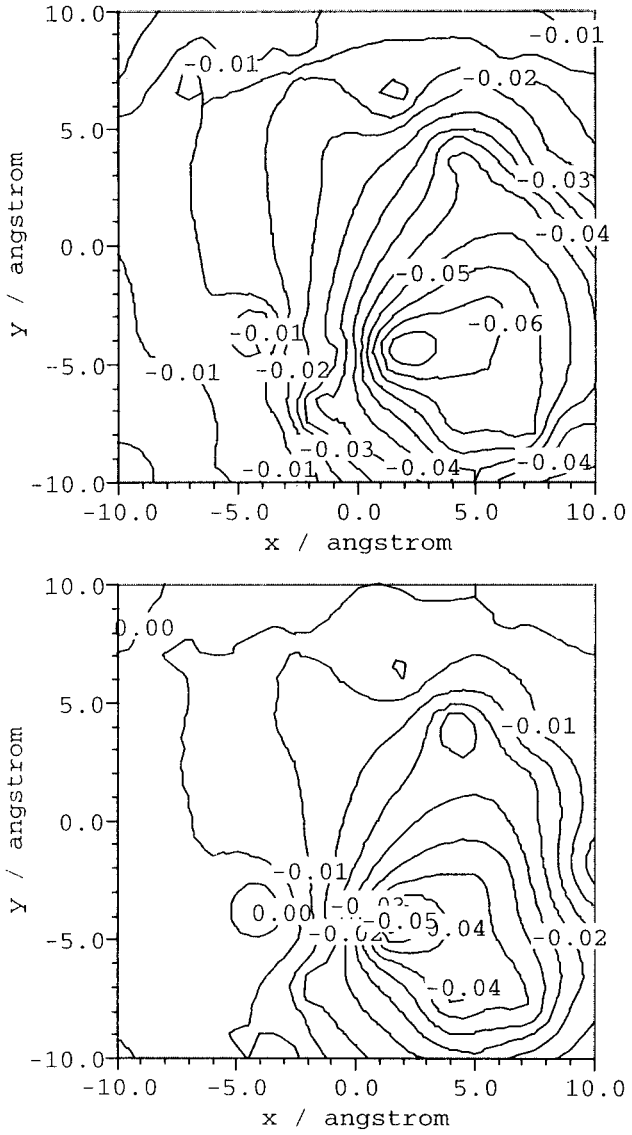

Figure 6. Electrostatic potentials generated from the charges ($\mathbf{q}^{II}$) induced in the region II of bR on binding of the chromophore. The induced charges in (a) were calculated by the PMM approximation, and those in (b) were obtained by AM1 calculations. These figures were drawn on the plane shown in Figure 2(a).

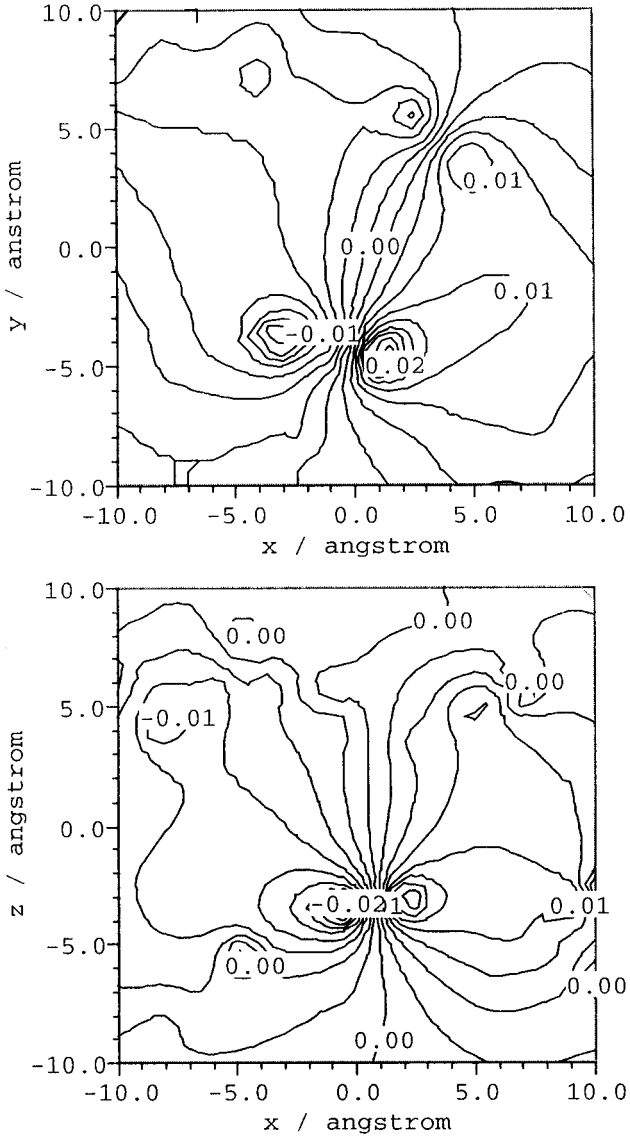

Figure 7. Electrostatic potentials generated from the charges ($\mathbf{q}^{II}$) induced by the $\pi-\pi^*$ excitation of the chromophore. These were drawn on the plane shown in Figures 2(a) and (b), respectively.

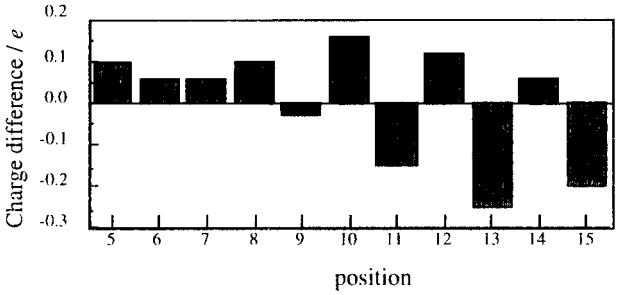

Figure 8. Change in atomic charges of the unsaturated carbons of the chromophore on going from the ground state to the $\pi-\pi^*$ excited state.

Trp86, Trp182, and Tyr185 were found to be remarkably polarized so as to make a significant contribution to the formation of the dipole-like field shown in Figure 7. The other aromatic residues exhibit no apparent polarization.

To appreciate the importance of the aromatic residues picked up above, we calculated the absorption maxima of a series of mutants: W86A, W138A, W182A, W189A, Y83A, Y185A, and F208A. The results are summarized in Table 3, where the second column contains the absorption maxima (λ^A) based on the approximation (A), and the fourth column does the absorption maxima (λ^B) based on the approximation (B). ΔE^A (ΔE^B) in this table also shows the change in $1/\lambda^A$ ($1/\lambda^B$) on going from WT to each mutant. A significant amount of blue shift was observed in λ^B values of W86A and Y185A relative to that of WT, while the λ^A values for these mutants, especially for Y185A, exhibit relatively small blue shifts. These results indicate that the electronic polarization effect of Trp86 and Tyr185 greatly contributes to the opsin shift of bR. In contrast to expectation from the potential map, W182A causes only a slight amount of blue shift (24.5 cm^{-1}).

Furthermore, we examined the effects of double and triple mutations (Table 3). The ΔE^B value for W86A/Y185A is 1826.2 cm^{-1}, which is significantly larger than the simple sum of the ΔE^B values for W86A and Y185A (982.2 + 509.8 = 1492.0 cm^{-1}). This suggests that additivity does not hold among the ΔE^B values. We attempted further analysis for the origin of the difference (1826.2−1492.0 = 334.2 cm^{-1}) in ΔE^B. The ΔE^B value for Y185A (509.8 cm^{-1}) corresponds to the shift induced by Tyr185 in the presence of Trp86. On the other hand, the

TABLE 3: Calculated Absorption Maxima of Several Mutants of BR

system	λ^A/nm	ΔE^{Aa}/cm^{-1}	λ^B/nm	E^{Bb}/cm^{-1}
WT	465.3	(0.0)	534.9	(0.0)
W86A	456.3	423.9	508.2	982.2
Y185A	464.0	60.2	520.7	509.8
W189A	462.6	125.4	530.0	172.8
W182A	464.1	55.6	534.2	24.5
Y83A	464.9	18.5	536.4	-52.3
W138A	465.8	-23.1	537.4	-87.0
F208A	465.9	-27.7	539.9	-173.1
W86A/Y185A	454.5	510.7	487.3	1826.2
W86A/W182A	455.4	467.2	501.2	1257.0
W86A/W182A/Y185A	453.1	578.7	482.7	2021.7

[a] Energy difference obtained by subtracting the inverse of λ^A of WT from that for each mutant. [b] Energy difference obtained by subtracting the inverse of λ^B of WT from that for each mutant.

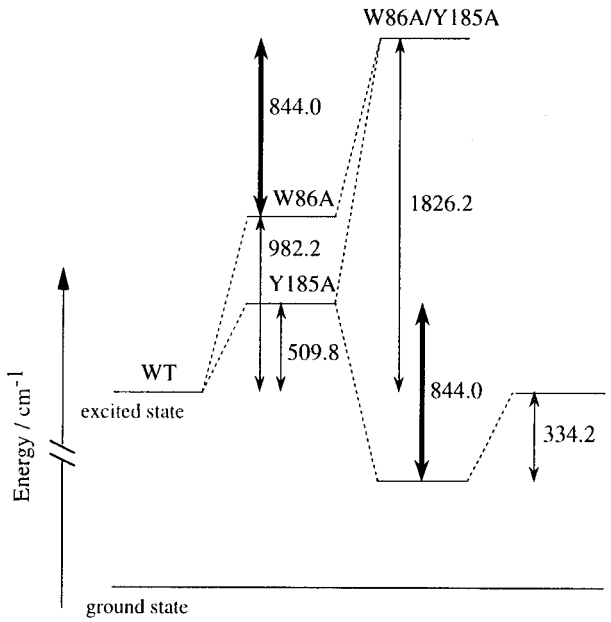

Figure 9. Energy diagram for WT, W86A, Y185A, and W86A/Y185A. The energy is given in cm^{-1}. The difference (844.0 cm^{-1}) in the excited-state energy between W86A/Y185A and W86A corresponds to the contribution of Tyr185 in the absence of Trp86. The subtraction of this value from the excited-state energy of T185A gives the excited-state energy lower than that of WT by 334.2 cm^{-1}, which corresponds to the cost of the mutual polarization of Trp86 and Tyr185.

difference in ΔE^B between W86A and W86A/Y185A is 844.0 cm^{-1}, a value corresponding to the contribution of Tyr185 in the absence of Trp86. Consequently, it costs 334.2 (= 844.0−509.8) cm^{-1} to simultaneously polarize Tyr185 and Trp86. This cost corresponds to the term $(1/2)\mathbf{q}^{II}\mathbf{R}\mathbf{q}^{II}$ in eq 20, meaning the work required for mutual polarization among the residues in the region II. Figure 9 illustrates the energy diagram for these mutants. The collapse of additivity arises from the fact that the excited-state energy is influenced by two opposing factors: stabilization due to the attraction between the chromophore ($\mathbf{q}^I$) and the induced bond dipoles ($\mathbf{q}^{II}$), and the work for mutual polarization among the induced bond dipoles.

On the basis of the above results, the origin of the small blue shift observed for W182A may be explained. By comparing the result for W86A/Y185A and that for W86A/W182A/Y185A, the shift induced by Trp182 in the absence of both Trp86 and Tyr185 is estimated to be 195.5 (= 2021.7−1826.2) cm^{-1}. On the other hand, the ΔE^B value for W86A/W182A/Y185A is 2021.7 cm^{-1}, and the simple sum of the ΔE^B values for W86A, W182A, and Y185A is 1516.5 cm^{-1}. Namely, the cost required for simultaneous polarization of Trp86, Trp182, and Tyr185 is 2021.7−1516.5 = 505.2 cm^{-1}, which is larger than the cost (334.2 cm^{-1}) of double polarization of Trp86 and Tyr185 by 171.0 (= 505.2−334.2) cm^{-1}. In other words, the removal of Trp182 reduces the cost of the mutual polarization by 171.0 cm^{-1}. Consequently, the blue shift of 195.5 cm^{-1}, corresponding to the separate contribution of Trp182, is nearly canceled out by the work of the mutual polarization (171.0 cm^{-1}) between Trp182 and the neighboring residues of Trp86 and Tyr185, resulting in a relatively small blue shift (195.5−171.0 = 24.5 cm^{-1}), in agreement with the difference in absorption maxima between WT and W182A.

TABLE 4: Experimental Absorption Maxima of Several Mutants of bR

	λ_{max}/nm		
	light-adated state	dark-adapted state	
WT	568	555	1
Y83F	540	533	36
W86F	547	534	37
W86F	539	529	35
W138C	547	543	35
W138F	561	551	35
W182F	491	477	35
Y185F	573	556	36
Y185F	563	548	39
Y185F		556	38
W189F	521	513	35

If the absorption maximum of the chromophore-Asp85 complex (474.4 nm; see Table 2) is taken as a reference, the opsin shift of WT is evaluated to be 2384.2 cm^{-1}. On the other hand, the opsin shift of the W86A/W182A/Y185A mutant is 362.5 is cm^{-1}, corresponding to only 15% of the opsin shift of WT. This indicates that the electronic polarization effect of Trp86, Trp182, and Tyr185 amounts to 85% of the total environmental effect of the protein. Thus, we can conclude that these residues dominate the opsin shift of bR. It should be noted that the λ^A values of all of the mutants studied above blue shifts relative to that of the chromophore-Asp85 complex. This indicates that the fixed charges of the protein never induce a red shift.

Discussion

Origin of the Opsin Shift. The present calculations demonstrated that the electronic polarization of the several aromatic residues is a main origin of the opsin shift. Inversely, the replacement of Trp86, Trp 182, and Tyr185 by alanine should induce a significant amount of blue shift. The importance of these residues in wavelength regulation has been pointed out from a variety of mutagenetic studies.[34−39] Table 4 summarizes the experimental absorption maxima of several mutants in which one of the aromatic residues of interest is replaced by Phe or Cys. The absorption maxima of these mutants is significantly blue shifted relative to that of the wild type. These results can be qualitatively interpreted. In W86F, for example, the blue shift is attributed to the decrease in polarizability at the substituted position, since the polarizability (10.4×10^{-24} cm^{-1}) of benzene is smaller than that (14.9×10^{-24} cm^{-1}) of indole. The blue shifts observed for most of the mutants are explained by the same reason. Exceptionally, Y185F exhibits relatively a small blue or red shift, probably because the polarizabilities of tyrosine and phenylalanine nearly equal to each other.

The site-directed mutagenetic studies also revealed that the substitution of ionizable residues exerts no apparent influence on the absorption maxima of bR.[40] Exceptionally, D85N, where

TABLE 5: Effect of the Protonation of Asp85 on the Absorption Maximum of bR

system	λ_{max}/nm	ΔE^a/cm^{-1}	ref
D85N	610	1212	41
D85N	594	771	40
D85N	590	656	39
acidic bR	605	1077	42
deionized bR	605	1077	42
O-intermedite	640	1981	2
A-bR	556.4	721.4	this work

[a] The amount of red shift measured from the absorption maximum of the wild-type bR.

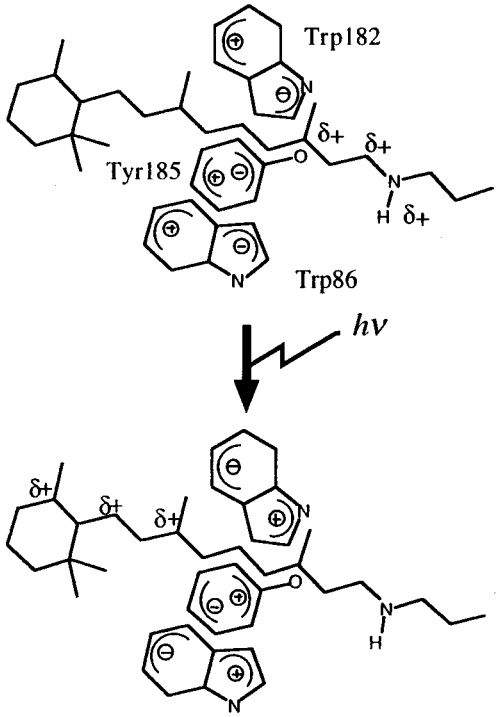

Figure 10. Schematic representation of the molecular model of the opsin shift. In the ground state, the positive charge of the chromophore is localized around the Schiff base nitrogen. Upon excitation, the positive charge moves toward the ionone ring, and the surrounding aromatic residues (Trp86, Trp182, and Tyr185) polarize so as to stabilize the excited state.

the negative charge of Asp85 is neutralized, shows the absorption maxima around 600 nm (Table 5). This large red shift is similar to the so-called purple-to-blue transition observed for acidic bR and deionized bR,[42] where the carboxyl group of Asp85 is protonated. In addition, the O intermediate, in which Asp85 is protonated, shows its absorption maxima at 640 nm.[1] These observations indicate that the neutralization of Asp85 induces a significant amount of red shift. In fact, the calculated absorption maximum of bR on protonation of Asp85 (denoted as A-bR) appears at 556.4 nm, which shifts to red by 721.4 cm^{-1} relative to that of WT (Table 5).

On the basis of evidence from the experiments and our calculations, we propose a new molecular mechanism for the opsin shift of bR: the electronic polarization effects of Trp86, Trp182, and Tyr185 cooperatively stabilize the excited state of the chromophore, resulting in a significant red shift. This mechanism is illustrated in Figure 10.

In previous studies,[13,14] we studied medium effects on the absorption maxima of retinal and its Schiff bases using a self-consistent reaction field (SCRF) calculation combined with the CI method. The SCRF calculation is based on a polarizable continuum model, where the protein matrix is approximated as a dielectric. From regression analysis for experimental and calculated data, it was clarified that the solvatochromic shifts of these compounds are well-defined functions of physical parameters (i.e., dielectric constant (ϵ) and refractive index (n)) of the surrounding medium. In particular, the absorption maximum of PRSB significantly shifts to red with an increase in the refractive index of the medium. On the other hand, on the basis of the molecular structure of bR, the effective refractive index of the chromophore-binding pocket was estimated to be 1.51. On the assumption of this value for the refractive index and 4.0 for dielectric constant, the SCRF calculations quantitatively reproduced both opsin shifts of bR_{568} and M_{412}. From decomposition analysis of the calculated opsin shift (4700 cm^{-1}) of bR_{568}, the polarizable medium effects of the protein matrix were shown to be 1000 cm^{-1}. Table 6 compares the results from the polarizable continuum model (PCM) with those from the present calculations (PMM). In PMM, the approximation (A) takes into account only the fixed charge effect of the protein. This is similar to the situation that refractive index equals 1.0. On the other hand, the approximation (B) corresponds to a situation with refractive index of $n > 1$. In fact, the results from PMM are in excellent agreement with those from PCM, where refractive index is assumed to be 1.51. This indicates that the residues around the chromophore act as a dielectric medium with a high refractive index.

TABLE 6: Comparison of the Absorption Maxima of PRSB Based on the Different Approximations for the Protein Environment of bR

approximation	λ_{max}/nm
PCM[a]	
in hexane (ϵ, n) = (1.89, 1.375)	432.9
in bR (ϵ, n) = (4.00, 1.00)	477.1
in bR (ϵ, n) = (4.00, 1.51)	543.5
PMM	
approximation (A)	465.3
approximation (B)	534.9

[a] Data were taken from Ref 14.

Comments on the Present QM/Polarizable-MM Model. In the present QM/MM model, the polarization of the MM part was described by using polarizable bond dipoles placed along the covalent bonds of amino acid residues. This point discriminates the present model from other QM/polarizable-MM models,[16−18,43] where the MM polarization is described by using atom-centered polarizable dipoles. According to the present formalism, the polarization of the bond dipole in the *i*th excited state is replaced by generation of the induced charges $\mathbf{q}^{II,i}$ on the two atoms constituting that bond. Such a treatment is similar to that developed in the MM3 molecular mechanics.[44−46] As a result, the interaction energy between regions I and II is expressed by charge−charge interactions, which can be exactly evaluated in the same way as the calculation of one-electron integrals in usual MO programs.[47]

In the first step of the present QM/MM procedure (see Figure 1), the ground-state wave function of the whole protein is obtained by MOZYME calculation. This gives advantages over conventional QM/MM methods. First, one can directly obtain the self-consistent charge distribution ($\mathbf{Q}^{I} + \mathbf{Q}^{II}$) over regions I and II. Second, because the whole protein is once treated quantum mechanically, it is unnecessary to add nonbonded terms, such as Lennard-Jones 6−12 potentials, to the van der Waals interactions between QM/MM and MM/MM atom pair.

There may be a little room for improvement or refinement of the present methodology. First, as first order approximation,

we took into account only the effect of longitudinal polarizability in the calculation of induced bond dipoles (see eqs 36 and 37), and neglected those of the transverse and vertical components of bond polarizability. This approximation corresponds to the modified Smith-Eyling methods given by Allinger and Wuesthoff.[44] The formulation involving the effects of the transverse and vertical polarizabilities has been already given by Allinger's group.[45,46] In a future study, our method will be modified in line with that model. Second, here we adopted eq 38 instead of directly solving eq 31, partly because the evaluation of inverse of the **A** matrix was not easy for large size molecules. This means that the mutual polarization in region II is not explicitly taken into account in the step of evaluation of $\mathbf{q}^{\mathrm{II},i}$. We are now developing a modified method in which the dipole−dipole interactions among nongeminal bonds can be explicitly taken into account. As expected naturally, a preliminary study indicated that the mutual polarization played a role in slightly depressing the polarization effect: namely, the λ_{max} value obtained by approximation (B) in Table 2 is decreased by about 10 nm. The reason the contribution of the mutual polarization within region II is relatively small may be explained as follows. In the case of bR, the chromophore (PRSB) has the total charge of +1, and the positive charge largely migrates from the Schiff base side to the ionon ring on excitation (Figure 8). Consequently, the $\mathbf{q}^{\mathrm{II},i}$ is influenced more largely by an electric field change induced by such a proton delocalization (the charge distribution change $\mathbf{q}^{\mathrm{I},i}$ in region I) than by that from the mutual polarization within region II. In other words, in eq 31 the quantity **AR** is sufficiently smaller than **AG**, leading to eq 38 (see "Parametrization of Bond Polarizability"). Therefore, it can be safely said that the neglect of the mutual polarization in region II does not affect the main conclusion of this study described below. Finally, since the publication of 2BRD structure by Grigorief et al.,[30] there have been reported more refined structures of bR from several groups.[48−50] It is of interest to investigate to what extent the calculated absorption maximum of bR is affected by the difference in the initial structure used.[51] The details of these three refinements will be reported elsewhere.

Concluding Remarks

The PMM approximation presented here proved to be a rapid and reliable method for estimating the electronic polarization of large molecules such as protein. The PMM approximation, combined with CI calculation, is applicable to problems concerning the absorption maximum of a chromophore in a protein. The most valuable finding in this study is that the bathochromic shift observed for bR is dominated by several aromatic residues around the chromophore, i.e., Trp86, Trp182, and Tyr185. On the basis of these results, we propose a new molecular model for the opsin shift of bR: the electronic polarization of these aromatic residues stabilizes the excited state, resulting in a red shift. The present results also agreed with those from the calculations based on the continuum approximation. This ensures us of the validity of the conclusion in the previous study[14] that the aromatic residues inside the chromophore-binding pocket serve as a polarizable medium.

Acknowledgment. The authors thank the Computer Center, Institute for Molecular Science, Okazaki, Japan, for the use of the supercomputer system. And we thank the Computer Center, Tokyo Institute of Technology, for the use of the SGI origin 2000 system.

References and Notes

(1) See for reviews: (a) Birge, R. R. *Annu. Rev. Phys. Chem.* **1990**, *41*, 683. (b) Lanyi, J. K. *Biochim. Biophys. Acta* **1993**, *1183*, 241. (c) Khorana, H. G. *Ann. N. Y. Acad. Sci.* **1986**, *471*, 272. (d) Mathies, R. A.; Lin, S. W.; Ames, J. B.; Pollard, W. T. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 491.

(2) Lozier, R.; Bogomlni, R. A.; Stoeckenius, W. *Biophys. J.* **1975**, *15*, 955.

(3) Mathies, R.; Stryer, L. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2169.

(4) van der Steen, R.; Biesheuvel, P. L.; Mathies, M. A.; Lugtenburg, J. *J. Am. Chem. Soc.*, **1986**, *108*, 6410.

(5) Harbison, G. S.; Smith, S. O.; Pardoen, J. A.; Courtin, J. M. L.; Lugtenburg, J.; Herzfeld, J.; Mathies, R. A.; Griffin, R. G. *Biochemistry* **1985**, *24*, 6955.

(6) Harbison, G. S.; Mulder, P. P. J.; Pardoen, H.; Lugtenburg, J.; Herzfeld, J.; Griffin, R. G. *J. Am. Chem. Soc.* **1985**, *107*, 4809.

(7) Wada, M.; Sakurai, M.; Inoue, Y.; Tamura, Y.; Watanabe, Y. *J. Am. Chem. Soc.* **1994**, *116*, 1537.

(8) Bassov, T.; Sheves, M. *J. Am. Chem. Soc.* **1985**, *107*, 7524.

(9) (a) Blatz, P. E.; Johnson, R. H.; Mohler, J. H.; Al-Dilaimi, S. K.; Dewhurst, S.; Erickson, J. O. *Photochem. Photobiol.* **1971**, *13*, 237. (b) Blatz, P. E.; Mohler, J. H.; Navanglu, H. V. *Biochemistry* **1972**, *11*, 848. (c) Blatz, P. E.; Mohler, J. H. *Biochemistry* **1972**, *11*, 3240. (d) Blatz, P. E.; Mohler, J. H. *Biochemistry* **1975**, *14*, 2340.

(10) (a) Irving, C. S.; Byers, G. W.; Leermakers, P. A. *J. Am. Chem. Soc.* **1969**, *91*, 2141. (b) Irving, C. S.; Byers, G. W.; Leermakers, P. A. *Biochemistry* **1970**, *9*, 858.

(11) Suzuki, T.; Kito, Y. *Photochem. Photobiol.* **1972**, *15*, 275.

(12) (a) Kliger, D. S.; Milder, S. J.; Dratz, E. A. *Photochem. Photobiol*. **1977**, *25*, 277. (b) Milder, S. J.; Kliger, D. S. *Photochem. Photobiol*. **1977**, *25*, 287.

(13) Houjou, H.; Sakurai, M.; Inoue, Y. *J. Chem. Phys.* **1997**, *107*, 5652.

(14) Houjou, H.; Inoue, Y. Sakurai, M. *J. Am. Chem. Soc.* **1998**, *120*, 4459.

(15) See for reviews: (a) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298. (b) Monard, G.; Merz, K. M., Jr. *Acc. Chem. Res.* **1999**, *32*, 904.

(16) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

(17) Lutzhov, V.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4491.

(18) Thompson, M. A.; Schenter, G. K. *J. Phys. Chem*. **1995**, *99*, 6374.

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

(20) Lee, T.-S.; York, D. M.; Yang W. *J. Chem. Phys.* **1996**, *105*, 2744.

(21) Dixon, S. L.; Merz, K. M. Jr. *J. Chem. Phys.* **1997**, *107*, 879.

(22) Daniels, A. D.; Millam, J. M.; Scuseria, G. E. *J. Chem. Phys.* **1997**, *107*, 425.

(23) York, D. M.; Lee, T.-S.; Yang, W. *J. Am. Chem. Soc*. **1996**, *118*, 10940.

(24) Nadig, G.; Van Zant, L. C.; Dixon, S. L.; Merz, K. M., Jr. *J. Am. Chem. Soc*. **1998**, *120*, 5593.

(25) Houjou, H.; Koyama, K.; Wada, M.; Sameshima, K.; Inoue, Y.; Sakurai, M. *Chem. Phys. Lett*. **1998**, *294*, 162.

(26) van der Vaart, A.; Merz, K. M., Jr. *J. Am. Chem. Soc*. **1999**, *121*, 9182.

(27) Greatbanks, S. P.; Gready, J. E.; Limaye, A. C.; Rendell, A. P. *J. Comput. Chem*. **2000**, *21*, 788.

(28) *Lange's Handbook of Chemistry*, 11th ed.; Dean, J. A., Ed.; McGraw-Hill: New York, 1973.

(29) Stewart, J. J. P. *MOPAC2000*; Fujitsu Ltd.: Tkyo, Japan, 1999.

(30) Grigorieff, T.; Ceska, K.; Downing, K. H.; Baldwin, J. M.; Henderson, R. *J. Mol. Biol.* **1996**, *259*, 393.

(31) Pearlman, D. A.; Case, D. A.; Caldwel, J. W.; Ross, W. S.; Cheatham, T. E., III; Ferguson, D. M.; Seibel, G. L.; ChandraSingh, U.; Weiner, P. K.; Kollman, P. A. 1995 AMBER ver. 4.1, University of Calfornia, San Francisco.

(32) Dewer, M. J. S.; Zoebisch, E. G.; Hearly, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(33) Mataga, N.; Nishimoto, K. *Z. Physik. Chem.* **1957**, *13*, 140.

(34) Nakanishi, K. *Pure Appl. Chem.* **1991**, *63*, 161.

(35) Mogi, T.; Marti, T.; Khorana, H. G. *J. Biol. Chem.* **1989**, *264*, 14197.

(36) Mogi, T.; Stern, L. J.; Hackett, N. R.; Khorana, H. G. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 5595.

(37) Hatanaka, M.; Kashima, R.; Kandori, H.; Friedman, N.; Shees, M.; Needleman, R.; Lanyi, J. K.; Maeda, A. *Biochemistry* **1997**, *36*, 5493.

(38) Subramanium, S.; Marti, T.; Rŝsselet, S. J.; Rothchild, K. J.; Khorana, H. G. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2583.

(39) Duňach, M.; Marti, T.; Khorana, H. G.; Rothchild, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9873.

(40) Mogi, T.; Stern, L. J.; Marti, T.; Chao, B. H.; Khorana, H. G. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 4148.

(41) Logunov, S. L.; El-Sayed, M. A.; Song, L.; Lanyi, J. K. *J. Phys. Chem*. **1996**, *100*, 2391.

(42) Zhag, Y. N.; El-Sayed, M. A.; Bonet, M. L.; Lanyi, J. K.; Chang, M.; Ni, B.; Needleman, R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1445.
(43) Gao, J. *J. Comput. Chem*. **1997**, *18*, 1061.
(44) Allinger, N. L.; Wuesthoff, M. T. *Tetrahedron* **1977**, *33*, 3.
(45) Došen-Mićović, L.; Jeremiæ,D.; Allinger, N. L. *J. Am. Chem. Soc*. **1983**, *105*, 1716.
(46) Ma, B.; Lii, J.-H.; Allinger, N. L. *J. Comput. Chem*. **2000**, *21*, 813.
(47) If one adopts atom-centered polarizable dipole approximation, the Fock matrix elements involving dipole−electron interactions must be evaluated. In refs 18 and 43, these integrals were approximately evaluated after expansion of the dipole into a set of six point charges.
(48) Pebay-Peyroula, E.; Rummel, G.; Rosenbusch, J. P.; Landau, E. M. *Science* **1997**, *277*, 1676.
(49) Kimura, Y.; Vassylyev, D. G.; Miyazawa, A.; Kidera, A.; Matsushima, M.; Mitsuoka, K.; Murata, K.; Hirai, T.; Fujiyoshi, Y. *Nature* **1997**, *389*, 6647.
(50) Luecke, H.; Schobert, B.; Richter, H.-T.; Cartailler, J.-P.; Lanyi, J. K. *J. Mol. Biol*. **1999**, *291*, 899.
(51) In all the bR structures which we examined so far (refs 30 and 48−50), there are seven aromatic residues (W86A, W138A, W182A, W189A, Y83A, Y185A, and F208A) surrounding the chromophore and no significant difference in their side chain orientations. Therefore, the use of the structures other than 2BRD would not affect the main conclusion of this study, insisting the dominant role of these residues in causing the opsin shift of bR.