Theor Chem Acc (2003) 109:125–132
DOI 10.1007/s00214-002-0418-y

Theoretical
Chemistry Accounts

Regular article

Investigation of the $S_0 \rightarrow S_1$ excitation in bacteriorhodopsin with the ONIOM(MO:MM) hybrid method

Thom Vreven[1,2], **Keiji Morokuma**[1]

[1] Cherry L. Emerson Center for Scientific Computation and Department of Chemistry, Emory University, Atlanta, GA 30322, USA
[2] Gaussian Inc., 140 Washington Ave., North Haven, CT 06473, USA

Received: 13 October 2001 / Accepted: 6 September 2002 / Published online: 3 February 2003

Abstract. We have investigated the S_0 and S_1 electronic states in bacteriorhodopsin using a variety of QM/MM levels. The decomposition of the calculated excitation energies into electronic and electrostatic components shows that the interaction of the chromophore with the protein electric field increases the excitation energy, while polarization effects are negligible. Therefore, the experimentally observed reduction in excitation energy from solution phase to protein environment (the Opsin shift) does not come from the electrostatic interaction with the protein environment, but from either the interaction of the chromophore with the solvent or counter ion, or structural effects. Our high-level ONIOM(TD– B3LYP:Amber) calculation predicts the excitation energy within 8 kcal/mol from experiment, the discrepancy probably being caused by the neglect of polarization of the protein environment. In addition, we have shown that the level of optimization is extremely critical for the calculation of accurate excitation energies in bacteriorhodopsin.

Keywords: Opsin shift – Bacteriorhodopsin – ONIOM – QM/MM – Hybrid methods

Introduction

The protonated Schiff base of retinal (PSBR) is the chromophore in the bacteriorhodopsin protein (bR) [1, 2], which acts as a light-driven proton pump across the cell membrane in *Halobacterium halobium*. Upon absorption of a photon, PSBR photo-isomerizes from the *all–trans* to the *13–cis* isomer, which starts a series of proton transfer reactions that has as net result the migration of one proton from the cytoplasmic surface to the extracellular surface of the protein.

A reasonably accurate structure of bR was first determined by Henderson and co-workers [3], and to date at a highest resolution of 1.55 Å by Luecke and co-workers [4]. The protein is roughly shaped as a barrel, made up from 7 helices that are parallel with the lipids in the cell-membrane. The chromophore is bound covalently to Lys216 in the center of the barrel, and a hydrogen-bonded network consisting of functional groups and water molecules surrounds the Schiff base.

Intriguingly, the chromophore in bR absorbs at 568 nm [5], while in methanol a similar chromophore absorbs at 465 nm[6]. This phenomenon is known as the Opsin shift, and a number of explanations have been proposed to account for it. First, the solvent can affect the excitation. Second, the conformation of the chromophore can be constrained by the protein. Third, the interaction of the Schiff base with the counter ion can be different. Fourth, the charge distribution of the protein can interact with the chromophore. In the current paper we computationally explore this fourth possibility, using our ONIOM hybrid method, which can treat large systems at a very accurate level.

Several other computational studies on the excitation of PSBR in bacteriorhodopsin have been published recently. Houjou et al. presented semiempirical INDO/S excitation calculations of the chromophore, with the protein charge distribution calculated using the Mozyme method [7]. This study was later extended to include a polarizable protein environment [8], of which the importance was suggested from earlier polarizable continuum calculations [9]. For the explicit bR modeling, Houjou used the 2BRD structure by Henderson [10], keeping the non-H atoms frozen to the experimental values. Hayashi and Ohmine presented CASSCF and CASSCF–MRMP calculations, based on fully HF/MM optimized structures [11]. Both Hayashi and Houjou

Contribution to the Proceedings of the Symposium on Combined QM/MM Methods at the 222nd National Meeting of the American Chemical Society, 2001

Correspondence to: K. Morokuma
e-mail: morokuma@emory.edu

also investigated the quantum effect of the Asp85 counter-ion on the excitation.

Methods

ONIOM hybrid method

The ONIOM hybrid method [12, 13, 14, 15, 16, 17, 18, 19] in the Gaussian package [20] can combine any two or three quantum mechanical (QM) and molecular mechanical (MM) computational methods into one calculation. In the current work we deal with two-layer QM:MM combinations, for which the total energy of the system is obtained from three independent calculations:

$$E^{\mathrm{ONIOM}} = E^{\mathrm{QM,model}} + E^{\mathrm{MM,real}} - E^{\mathrm{MM,model}} \tag{1}$$

Real denotes the full system, which only needs to be calculated at the MM level, and *model* denotes the part of the system that needs to be calculated at both the QM and MM level. The method can be regarded as an extrapolation scheme. Starting from the MM energy for the model system, the extrapolation to the QM level ($E^{\mathrm{QM,model}} - E^{\mathrm{MM,model}}$) and to the full system ($E^{\mathrm{MM,real}} - E^{\mathrm{MM,model}}$) are assumed additive to give an estimate for the target, $E^{\mathrm{QM,real}}$. Many other QM/MM formalisms have been proposed in recent years, but most are presented as summation schemes instead of the ONIOM extrapolation scheme [21, 22].

$$E^{\mathrm{QM/MM}} = E^{\mathrm{QM,model}} + E^{\mathrm{MM,MM}} + E^{\mathrm{QM-MM}} \tag{2}$$

$E^{\mathrm{MM,MM}}$ stands for the MM energy of the MM layer alone, while $E^{\mathrm{QM-MM}}$ describes the interaction between the QM layer and the MM layer. In its simplest form, $E^{\mathrm{QM-MM}}$ contains the classically (using MM terms) evaluated electrostatic and van der Waals non-bonded interactions, and the bonded MM terms in the interface region. The "QM/MM expression" (Eq. 2) and the "ONIOM expression" (Eq. 1) are conceptually the same; the ONIOM ($E^{\mathrm{MM,real}}$ $E^{\mathrm{MM,model}}$) terms describe the energy of the MM layer plus the interaction between the two layers, just as ($E^{\mathrm{MM,MM}}$ + $E^{\mathrm{QM-MM}}$) in Eq. 2.

When the layers are not covalently bound, the model system is identical to the *high level layer*. In that case, the ONIOM and QM/MM formalism are equivalent. When covalent bonds do exist, the resulting dangling bonds can be saturated with *link atoms*, which are chosen so that they best mimic the substituents. This is the most straightforward approach, and used in ONIOM, although a number of hybrid methods employ alternative schemes that are based on 'frozen orbitals' or other ways to terminate the boundary [23, 24, 25]. In ONIOM, hydrogen link atoms usually yield good results when carbon–carbon bonds are broken. The C^{model}–H^{link} bonds are then assigned the same angular and dihedral values as the C^{model}–C^{real} bonds in the real system, while the C^{model}–H^{link} bond lengths are obtained by scaling the C^{model}–C^{real} bond length. This ensures that the number of degrees of freedom remains as 3N–6, so that any method for the investigation of potential energy surfaces available for conventional methods can be used for ONIOM as well.

The geometrical derivatives of the ONIOM energy can be obtained in a similar fashion as the energy. If link atoms are present, the Jacobian $\boldsymbol{J}$ must be used to convert the coordinate system for the model system to that for the real system. The gradient can be written as

$$\frac{\partial E^{\mathrm{ONIOM}}}{\partial \lambda} = \frac{\partial E^{\mathrm{QM,model}}}{\partial \lambda} \cdot \mathbf{J} + \frac{\partial E^{\mathrm{MM,real}}}{\partial \lambda} - \frac{\partial E^{\mathrm{MM,model}}}{\partial \lambda} \cdot \mathbf{J} \tag{3}$$

From Eq. 1 it follows that the interaction between the QM and the MM layer is included at the MM level, via the $E^{\mathrm{MM,real}}$ term. This type of embedding is referred to as mechanical embedding (ME). In fact, in the ONIOM scheme, the interaction between any two layers is always included at the lowest of the two associated levels. However, in many QM/MM methods, the electrostatic interaction between the layers is moved from the MM level to the QM level, which is referred to as electronic embedding (EE) [26]. In that case, the QM calculation of the model system is calculated in the potential field generated by the MM region, while the $E^{\mathrm{QM-MM}}$ term no longer includes the classically evaluated electrostatic interaction between the two layers.

$$\hat{H}_V^{Model} = \hat{H}_0^{Model} - \sum_i \sum_N \frac{q_N}{r_{iN}} + \sum_J \sum_N \frac{Z_J q_N}{r_{JN}} \tag{4}$$

$$E^{\mathrm{QM/MM-EE}} = E_V^{\mathrm{QM,model}} + E^{\mathrm{MM,MM}} - E^{\mathrm{bonded,VDWQM-MM}} \tag{5}$$

Where qs are the MM partial atom charges, Zs the nuclear charges, subscript i indicates the electrons, subscript N the point charges, and subscript J the QM atoms. Electronic embedding allows the QM wave function to be polarized, and also provides an improved description of the electrostatic interaction between the layers because the QM charge distribution is no longer approximated by MM atomic partial charges. The way electronic embedding is incorporated differs between the various QM/MM formalisms, depending on the exact way the coupling between the layers is treated in the first place. Equations 4 and 5 are therefore only 'prototype expressions'; in most QM/MM schemes a fine balance is obtained by including and excluding certain interactions that involve atoms in the interface region.

Electronic embedding can also be incorporated in the ONIOM framework. Since we want to stay true to the extrapolation philosophy, we modify *both* the QM and MM model system calculations by adding the electrostatic interaction of the QM region with the MM region:

$$E^{\mathrm{ONIOM-EE}} = E_{\mathrm{V}}^{\mathrm{QM,model}} + E^{\mathrm{MM,real}} - E_{\mathrm{V}}^{\mathrm{MM,model}} \tag{6}$$

$$E_v^{\mathrm{MM,model}} = E_0^{MM,model} + \sum_J \sum_N \frac{q_J q_N}{r_{JN}} \tag{7}$$

The real system MM calculation remains the same as in Eq. 1, including the electrostatic interaction between the MM and QM layers. Although the electrostatic interaction between the two layers is evaluated three times, once in each term on the r.h.s. of Eq. 6 (the $E^{\mathrm{MM,real}}$ term is the same as in Eq. 1, and still includes the electrostatic interaction between the two layers at the MM level), it is more or less equivalent to the QM/MM−EE formalism in Eq. 5, since many electrostatic terms in the second and third term cancel. In fact, when there is no bonded interaction between the two layers, the electronic embedding versions of ONIOM and QM/MM are also exactly the same. It must be noted that in the ONIOM electronic embedding scheme also the link atoms 'see' the surrounding charges. In a number of QM/MM schemes the link atoms are treated in a different way to the remaining QM atoms, which requires modifications to the electronic structure theory. Since this would compromise the generality of the scheme, ONIOM treats all the atoms in the model system in the same way.

When there is covalent interaction between the two layers, the situation is slightly more complicated with electronic embedding schemes, because large MM charges close to the QM region can over-polarize the wave function. In addition, QM/MM schemes can suffer from over-counting or under-counting. This is the result of electrostatic interactions in MM methods being scaled when they are close together (based on bond distance), which in its most straightforward way cannot been done with electronic embedding because the charge distribution in the QM region is not atom-based. These problems have been solved in a variety of implicit and explicit ways. In link-atom QM/MM schemes, the over-polarization is usually taken care of by scaling the charges that are close to the QM region [21], while the over- or under-counting can be corrected by an extra classical term that uses localized charges in the QM region to balance the electrostatic interaction at the electronic level [11]. In the ONIOM formalism, however, this correction term is naturally included, since the scaling of the charges close by the QM region is done in both the model system terms in Eq. 6. The exact formalism and implications

of electronic embedding in ONIOM(MO:MM) and the comparison to other schemes is beyond the scope of the current paper, and will be discussed in detail elsewhere [18].

Model

We used the X-ray structure of 1.55 Å resolution by Luecke [4] as the basis of our calculations. This structure contains the chromophore, the protein (with 222 amino acid residues), 18 membrane lipids, and 23 water molecules. We removed the lipids from the X-ray structure, and added hydrogen atoms with the GaussView [27] graphical interface. All protonation states were standard, apart from Asp96, Asp115, and Glu204, which were protonated. [28, 29] A section of the loop between helices E and F is missing from the X-ray structure (this sequence has a zero net charge [3]). These two artificial and the two regular ends of the protein were capped with NH_3^+ and COO^-. Thus the "real" system in the present calculation contains 3624 atoms.

Calculations

A private development version of the Gaussian package [20] was used for ONIOM calculations. The ONIOM model system is the chromophore, consisting of 50 atoms (plus one link H atom), which was treated both by QM and MM, while the above-mentioned real system was treated by MM. As the QM, we used the Hartree Fock method with the 6–31G(d) basis set (abbreviated as HF) or the B3LYP hybrid density functional method with the same basis set (abbreviated as B3LYP), and as the MM we adopted the Amber [30] molecular mechanics force field. Full optimizations of the geometry of the entire real system were performed with the ONIOM method, without boundary conditions or inclusion of solvent. Optimization was carried out both with the electronic embedding (−EE) as well as the mechanical embedding (−ME), thus yielding a total of four different geometries. In the model system calculations with the electronic embedding, we set the MM charges to zero for atoms closer than four bonds from the QM region. For the protein, we used the standard Amber charges, while for the chromophore, Mulliken charges were obtained at the B3LYP/6–31G(d) level for the ground state[1]. The latter were fixed for all the calculations, although in reality the charge distribution of the S_0 and S_1 states are very different.

For the excited state calculations, we used CIS (configuration interaction with single excitations), TD–HF (time-dependent HF), and TD–B3LYP (time-dependent B3LYP) methods with the 6-31G(d) basis set at various optimized geometries of the ground state. Only for the CIS calculations were we able to decompose the excitation energies into polarization and electrostatic components [31], or the contributions from the individual residues. One may note that these calculations are not the first on excited states using the ONIOM method; previously we investigated the performance of QM:QM combinations for the photo-isomerization of PSBR [32], and the $S_0 \rightarrow T_1$ and T_2 excitations for a series of cyclic alkenes and enones [33].

Lastly, we give the details of the energy minimization procedures. A combination of optimization techniques was applied. Using a conjugate gradient optimizer in Cartesian coordinates, we first optimized the hydrogen atoms, then included the water molecules, followed by the inclusion of the amino acid side-chains, and finally full relaxation of the protein, keeping only the chromophore fixed. The last step took about 5000 steps, and the RMS force of the optimized structure was 9×10^{-6} a.u. This represents the relaxation of the protein environment; the deviations in the protein part of the system of the ONIOM optimized structures (see below) with this protein-only optimized structure are negligible. We used the micro-iterations technique for the optimization of the mechanical embedding structures: for each optimization step in the chromophore QM region, the protein MM region is fully relaxed. For the QM region we employ the regular Gaussian optimizer in redundant internal coordinates, while for the MM region we use the Cartesian coordinate optimizer as described above. For the electronic embedding optimization we could not employ the micro-iterations, and needed to use a very basic (no conjugated gradient) Cartesian optimizer.

[1] For the chromophore we used Mulliken charges obtained from a B3LYP/6–31G(d) gas phase optimized structure, averaging the charges of equivalent atoms, although they are not necessary the best choice in QM/MM (or MM) methods. However since we are mainly interested in the electronic embedding calculations, where virtually all terms involving these charges vanish, we have not attempted to derive a more appropriate set of partial charges.

Results

Structure

Before investigating the Opsin shift, we must validate the structural integrity of our computational model. The main points of concern in our calculations are the negligence of the environment beyond the protein (most importantly the lipid bilayer), and the fragment missing from the X-ray structure. In addition, since we only perform energy minimizations, and no molecular dynamics (MD) calculations, temperature or other dynamics effects are not taken into account. We assume that this does not affect the excitation energies much. In principle we could improve upon the neglect of environment by setting some boundary conditions, while the missing loop could be constructed based on other available experimental structural or sequence data. However, we preferred not to introduce new parameters, and below we show that the optimization did not distort the structure significantly.

At first, the ground state structure fully optimized by ONIOM was compared with the X-ray structure. The deviations of C–α coordinates for ONIOM(HF:Amber)–ME optimized structure from the X-ray results are shown in Fig. 1[2]. Only the deviations of C–α backbone atoms are used to avoid noise caused by the floppy side chains. From the figure we see that only five C–α atoms have moved more than 3 Å, which indicates that the overall structure has remained intact during the optimization. Even the loose ends that terminate the protein and cap the missing fragment have remained more or less in place. The stability of the structure is probably the result of the barrel conformation and the fact that we did not perform dynamics calculations. Despite using fairly tight criteria[3], geometry optimization procedures are not likely to result in major structural changes. The main reasons are that the individual forces that govern such events are small, and that the optimization nearly always converges to the closest minimum. Thus, although the approximations in

[2] The structures were overlapped by minimizing the RMS deviation (all atoms were included with equal weights), using the Tinker package: Tinker, version 3.1, Ponder JW (1996) Washington University School of Medicine, St. Louis, Missouri.

[3] The RMS forces of the optimized structures are 6×10^{-6}, 16×10^{-6}, 13×10^{-6}, and 17×10^{-6} for ONIOM(HF:Amber)–ME, ONIOM(HF:Amber)–EE, ONIOM(B3LYP:Amber)–ME, and ONIOM(B3LYP:Amber)–EE, respectively.

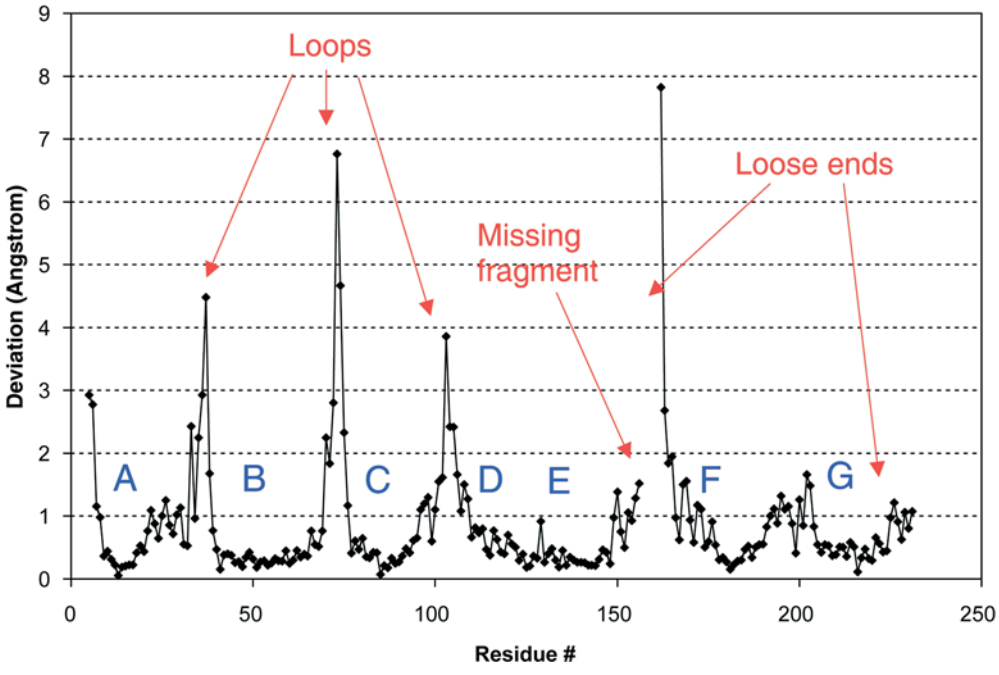

Fig. 1. Deviations of the ONIOM(HF:Amber)– ME optimized C–α coordinates from the X-ray structure

our model make it unsuitable for MD studies, they pose no problems for our current objectives. It must be noted that local relaxation does occur over large regions, which is important in this work because we compare subtle geometrical differences between computational levels. If we had kept the crystal structure rigid and only optimized the QM region, even a small bias of the experimental structure for any particular chromophore geometry might have affected the results.

Looking more closely at the figure, we see 7 valleys and 8 peaks, which correspond to the alpha-helices (labeled A to G) and the loops between them, respectively. The helices move less during the optimization than the loops, because they are better defined in the X-ray structure and are more tightly bound. Because the chromophore is located more or less in the center of the barrel, we need to pay attention mainly to the deviations of the helices. Since the deviation is nearly always less than 1 Å, it is likely that the active center and its environment are similar in the optimized and experimental structures.

In Fig. 2 and Table 1 we show selected geometrical parameters of the chromophore and several residues and water molecules that are close by or involved in the initial phase of the proton pump mechanism. The agreement between the experimental and the computed structures for the hydrogen-bonding network around the chromophore in Fig. 2 is excellent in general, regardless of the computational level employed. The only major difference between calculation and experiment is the hydrogen bond between Asp212 and W402, where different carboxylate oxygens are involved in the experimental and optimized structures. Our present results are very similar to the results of Hayashi et al, shown also in Fig. 2, although they also found with a different QM/MM partitioning that the Asp212–W402 bond is as in the X-ray structure, while the PSB–W402 bond is broken [11]. All our other calculated hydrogen bond lengths in the reaction center are within 0.2 Å of the experimental values. The differences are slightly larger for several of the other distances, specifically between Asp85 and the Schiff base. These are 3.59–3.66 Å in calculations vs 3.79 Å in experiment, and 3.91–3.93 vs 4.38 Å for the second oxygen atom.

Looking more closely at the chromophore dihedral (Fig. 2) and bond length (Table 1) parameters, we see that electronic embedding causes the single–double bond alternation to be more significant and the polyene to be slightly flatter. This corresponds to a decrease in the contribution of the resonance structure that has the singe/double bonds inverted and the positive charge migrated away from the Schiff base [34], which is most likely caused by the negative charges of Asp85 and Asp212. In the EE calculation, the negative charges of the nearby amino acid residues polarize the chromophore and increase the single–double bond alternation. One recognizes that this effect is much more pronounced in the ONIOM(HF:Amber) calculations than in the ONIOM(B3LYP:Amber); the differences in bond distances in the vicinity of the PSB N^+ center in Table 1 between −EE and −ME calculations with ONIOM (HF:Amber) are about twice as large as those with ONIOM(B3LYP:Amber), which is the result of the alternation being over-estimated at the HF level in the first place. From Table 1 it is also clear that the main geometrical effect of the protein environment on the chromophore can be found in the dihedral angles in the vicinity of the Schiff base. The chromophore in bR is significantly twisted, specifically about the $C_{14}=C_{13}$ bond, which isomerizes in the protein pump mechanism, while the polyene chain is virtually planar in the gas phase.

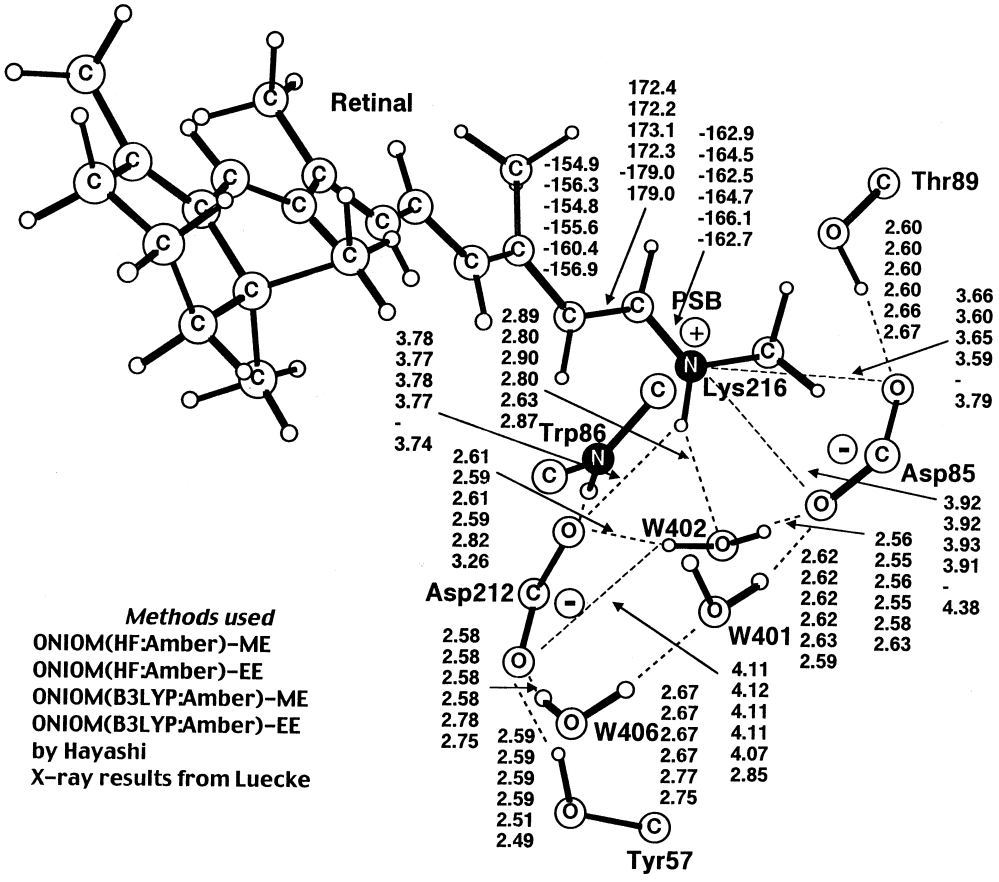

Fig. 2. Selected geometrical parameters of the chromophore binding site (distances in Å, dihedral angles in degrees). Numbers from *top to bottom*, for ONIOM(HF:Amber)–ME, ONIOM(HF:Amber)–EE, ONIOM(B3LYP:Amber)–ME, ONIOM(B3LYP:Amber)–EE, values reported by Hayashi (Asp85, Asp212, W401, W402, and W406 included in the QM region as well), and X-ray results from Luecke

Table 1. Optimized bond lengths (Å) and dihedral angles (degrees) along the chromophore polyene chain

	$N_{16}=C_{15}$	$C_{15}–C_{14}$	$C_{14}=C_{13}$	$C_{13}–C_{12}$	$C_{12}=C_{11}$	$C_{11}–C_{10}$	$C_{10}=C_9$	$C_9–C_8$	$C_8=C_7$	$C_7–C_6$	$C_6=C_5$
Structure					Bond lengths						
ONIOM(HF:Amber)–ME	1.32	1.37	1.40	1.40	1.37	1.41	1.36	1.45	1.34	1.47	1.35
ONIOM(HF:Amber)–EE	1.29	1.40	1.37	1.43	1.34	1.44	1.34	1.46	1.33	1.48	1.34
ONIOM(B3LYP:Amber)–ME	1.33	1.37	1.41	1.39	1.39	1.40	1.39	1.42	1.37	1.44	1.38
ONIOM(B3LYP:Amber)–EE	1.32	1.39	1.39	1.41	1.37	1.41	1.38	1.44	1.36	1.45	1.37
HF (gas phase)	1.32	1.37	1.41	1.40	1.38	1.41	1.37	1.45	1.35	1.47	1.35
B3LYP (gas phase)	1.34	1.38	1.41	1.40	1.39	1.40	1.40	1.43	1.38	1.44	1.38
Structure					Dihedral angles[a]						
ONIOM(HF:Amber)–ME	–162.89	172.40	–154.89	177.89	–174.77	–177.24	179.38	–174.58	177.76	172.58	–177.73
ONIOM(HF:Amber)–EE	–164.51	172.24	–156.34	176.35	–173.61	–176.70	179.20	–174.02	178.08	171.61	–178.27
ONIOM(B3LYP:Amber)–ME	–162.53	173.06	–154.77	177.28	–174.95	–176.94	178.83	–175.25	177.67	173.26	–177.35
ONIOM(B3LYP:Amber)–EE	–164.66	172.25	–155.56	176.32	–174.09	–176.76	179.00	–174.48	178.04	172.07	–177.99
HF (gas phase)	179.99	–179.98	179.96	–179.96	179.94	179.94	179.94	179.98	–178.57	166.02	–174.87
B3LYP (gas phase)	–179.98	179.86	–179.76	179.61	–179.61	179.20	–179.52	178.17	–177.95	169.79	–172.43

[a]H–N–C–C, N–C–C–C, and C–C–C–C dihedrals angles along the polyene chain

Excitation energies

In Table 2 we show the $S_0 \rightarrow S_1$ vertical excitation energies ($\triangle$E) calculated at the CIS, TD–HF, and TD–B3LYP levels with the 6–31G(d) basis set, both with and without electronic embedding, for the structures optimized with and without electronic embedding. $\triangle < Y^0|H^0|Y^0 >$ is the excitation energy without protein field and $\triangle < Y^V|H^V|Y^V >$ is the excitation energy in the protein field (H^V) with the wave function Y^V that is polarized by the protein field, i. e., with EE. The values of $\triangle < Y^V|H^V|Y^V >$ with the CIS and TD–HF methods overestimate the experimental value of 50.3 kcal/mol (568 nm, in bR) by 20–30 kcal/mol, while TD–B3LYP approaches it to within 10 kcal/mol.

It is interesting to note that for both TD–HF and CIS calculations, the excitation energy without protein field, $\triangle < Y^0|H^0|Y^0 >$, is larger with the electronic embedding ONIOM(HF:Amber)–EE structure than with the mechanical embedding ONIOM(HF:Amber)–ME structure

Table 2. Vertical excitationenergies $\triangle$E (kcal/mol) at the various computational levels. Y^0 and Y^V stand for unpolarized and polarized wave functions, respectively, and H^V and H^0 for the Hamiltonians,with and without inclusion of the electrostatic interaction with the proteinenvironment

Structure		TD–HF or TD–B3LYP[a]		CIS			
	Exp.[b]	$\triangle < Y^0\|H^0\|Y^0 >$	$\triangle < Y^V\|H^V\|Y^V >$	$\triangle < Y^0\|H^0\|Y^0 >$	$\triangle < Y^0\|H^V\|Y^0 >$	$\triangle < Y^V\|H^0\|Y^V >$	$\triangle < Y^V\|H^V\|Y^V >$
ONIOM(HF:Amber)–ME	50.3	68.34	70.97 (+2.63)	73.10	75.93	73.08	75.98 (+2.89)
ONIOM(HF:Amber)–EE		73.59	79.39 (+5.80)	77.87	84.20	78.22	83.95 (+6.08)
ONIOM(B3LYP:Amber)–ME		56.97	58.30 (+1.33)				
ONIOM(B3LYP:Amber)–EE		55.80	58.23 (+2.43)				
HF (gas phase)		68.97		73.66			
B3LYP (gas phase)		57.09					
In methanol	61.4						

[a]TD–HF forONIOM(HF:Amber) and HF structures, and TD–B3LYP forONIOM(B3LYP:Amber) and B3LYP structures
[b]From references [5] and [6]

(68.34 vs. 73.59 kcal/mol for TD–HF, and 73.10 vs. 77.87 kcal/mol for CIS). This can be related to the change in dihedral and bond length parameters of the polyene, discussed above. The inverted resonance structure resembles the S_1 electronic state, so that the electronic embedding structure will be favorable for the ground state, and unfavorable for the excited state, resulting in larger excitation energy. The pure gas phase HF optimized structure of the chromophore is very similar to that of ME (except for some dihedral angles; the gas phase optimized structures are nearly planar, with maximum dihedral angles in the polyene fragment of 178.6 and 178.0 degrees at the HF and B3LYP levels, respectively), and Table 2 shows that these two structures give virtually the same excitation energy. The large geometrical difference between the ME and EE optimized structures does not exist between the TD–B3LYP optimized geometries, as discussed above, and therefore the excitation energy at ONIOM(B3LYP:Amber)–EE structure is within 1.2 kcal/mol from that with – ME. Table 2 also shows that at a given structure, the excitation energy in the protein field (H^V) with the EE polarized wave function, $\triangle < Y^V|H^V|Y^V >$, is larger than the gas phase excitation energy, $\triangle < Y^0|H^0|Y^0 >$, by 1.3 to 6.1 kcal/mol, depending on the method of calculation and the geometry used. The excitation energy is *blue-shifted* going from gas phase to protein environment *at a given structure*. In the case of CIS calculations at the ONIOM(HF:Amber)–ME and –EE structures, one can analyze the difference by calculating intermediate energies. Starting from $\triangle < Y^0|H^0|Y^0 >$, one can turn on the protein field H^V but still use the gas phase wave function Y^0 to calculate $\triangle < Y^0|H^V|Y^0 >$. The difference, $\triangle < Y^0|H^V|Y^0 > - < Y^0|H^0|Y^0 >$, represents the electrostatic effects of the protein potential, and is +2.8 and +6.3 kcal/mol for the –ME and the –EE structure, respectively. Alternatively, one can calculate $\triangle < Y^V|H^0|Y^V >$ with the EE polarized wave function Y^V and the gas phase potential. The difference, $\triangle < Y^V|H^0|Y^V > - \triangle < Y^0|H^0|Y^0 >$, represents the effects of the wave function polarization on the gas phase excitation energy, and is –0.2 and +0.35 kcal/mol, respectively. When both Y^V and H^V are used, $\triangle < Y^V|H^V|Y^V >$ is the self-consistent excitation energy in the protein field (with the polarized wave function), and $\triangle < Y^V|H^V|Y^V > - \triangle < Y^0|H^0|Y^0 >$ is the overall shift of excitation energy at the given geometry due to the protein field. They are +2.9 and +6.1 kcal/mol for the –ME and the –EE structure, respectively. These results clearly show that the effect of wave function polarization by the protein field on the excitation energy is very small, and the shift of the excitation energy, $\triangle < Y^V|H^V|Y^V > - \triangle < Y^0|H^0|Y^0 >$, can be calculated essentially with the protein potential, $H^V - H^0$, and the unpolarized wave function Y^0 as $\triangle < Y^0|H^V - H^0|Y^0 >$. Now let us discuss the origin of the blue shift by the protein field. Upon excitation, the positive charge of the chromophore moves away from the Schiff base, which reduces the electrostatic interaction with the negatively charged Asp212 and Asp85, thus increasing $\triangle$E. The increase in $\triangle$E is larger for the EE optimized structure, in which water molecules (especially W402) and Asp residues (especially Asp85) are slightly closer to the Schiff base. The latter would be expected based on the polarization of the chromophore in the electronic embedding optimization, which increases the electrostatic interaction with the negative groups and thus decreases the distance. Finally, in Fig. 3 we show the contributions to $\triangle < Y^0|H^V - H^0|Y^0 >$, the electrostatic effect of the protein environment on the excitation energy, of all the protein residues within a certain distance from the chromophore. Note that we use the shortest distance between the residue and the chromophore, and not the distance between the center of the residue and the center of the chromophore. Contributions from negatively charged, positively charged and neutral residues are shown separately, as well as all the residues together. The effect from the charged groups is large and erratic close to the chromophore. However, beyond 15 Å from the chromophore, all the contributions from the negatively charged residues increase the excitation energy, while all the positively charged groups reduce the excitation energy. This is simply the result of the location of the chromophore in the protein. The protonated Schiff

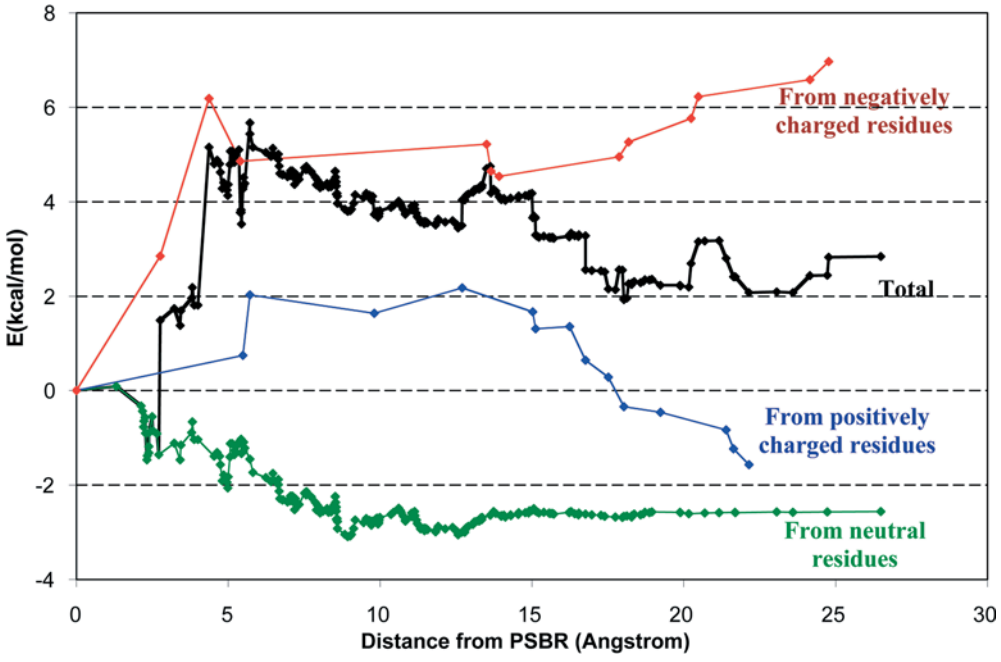

Fig. 3. Electrostatic effects of the protein environment on the excitation, as integrated functions of the distance from the chromophore, total and separately from negatively charged, positively charged, and neutral protein residues. For this figure we used the ONIOM(HF:Amber)–ME optimized structure, and the unpolarized wave function

base is more or less placed in the center of bR, and thus the polyene fragment extends away from the center, towards the surface of the protein. Because of the size of the system, the region that is further than 15 Å away from the chromophore cannot be located on the polyene side of the protein (with respect to the Schiff base) but must be at the opposite side. Now the positive charge of the Schiff base migrates along the polyene chain upon excitation, thus always away from the >15 Å region, and the potential at *each* residue in the >15 Å region is lowered. The amount the potential is lowered with depends on the exact location of the residue, but the contributions to the excitation energy will all be of the same sign for residues with the same charge. Furthermore, when we look at the effect from only the neutral residues, the curve is quite smooth and goes steadily down, up to about 10 Å. If the orientation of the groups were random, the curve should remain at zero. There is a net decrease in the excitation energy by about 2 kcal/mol, suggesting that there is a specific orientation of the (neutral) residues. This could be provided by the direction of the helices, but in the present system this effect should more or less cancel since the helices are parallel and in different directions. A second explanation could be the orientation of the residues to the chromophore charge distribution. Some detailed analysis of individual contributions may be useful in the future.

Discussion

Our results show that *the effect of the charge distribution of the protein environment at a given ground state geometry causes a blue-shift,* in accordance with the findings by Houjou [7] and Hayashi [11]. Our most important finding is that at the ONIOM(HF:Amber) optimized geometries, the shift is large (5.8–6.1 kcal/mol with – EE), and this is caused by a large change in the geometry of the chromophore when the effect of protein is taken into account. This shows that the level of optimization is critical, which has not been investigated previously. However, the blue shift is reduced to 2.4 kcal/mol at the ONIOM(B3LYP:Amber)–EE optimized geometry. Considering the fact that HF tends to overemphasize changes in bond alternation and that B3LYP results are in general more reliable than the HF results, with the lack of better calculation, we are inclined to side with the B3LYP results. We believe our ONIOM (TD–B3LYP:Amber) value of 58.2 kcal/mol is a good prediction of the excitation energy within the limits of our model. For the same partitioning, Hayashi predicts 63.5 kcal/mol at the CASSCF–MRMP level [11], where the difference may be caused by the HF(3–21G)/Amber level used for the optimization. Houjou predicts 62.1 kcal/mol at a semi-empirical MO level without optimization of non-H atoms [7]. However, in this paper, we have not investigated the electronic effect of the counter ion Asp85. Hayashi predicts an increase of 13 kcal/mol at the CASSCF level, which is very much in contrast with the prediction of less than 1 kcal/mol by Houjou. One likely reason is that Hayashi's calculation also includes the three water molecules in the reaction center, while these do not exist in Houjou's structure. It is not clear how the counter ions would affect our work. The main reason, however, for the discrepancy of our prediction and the experimental value is the neglect of the polarization of the protein environment, which has just recently been investigated in detail by Houjou [8]. A decrease of 8.0 kcal/mol was predicted at their semi-empirical level for the model similar to ours. Adjusting our prediction with this value yields 50.2 kcal/mol, which is extremely close to the experimental value of 50.3 kcal/mol.

It is clear that the protein electrostatic field alone cannot produce the Opsin shift, although we do not have a direct comparison of calculated excitation energies in both solution and protein. The gas phase results show similar excitation energies as in bR, thus the most likely origin of the Opsin shift is the effect of the solvent or the counter ion.

Conclusions

We have shown that the level of optimization is critical in the calculation of the excitation energy in bR. Especially at the HF level, where the geometry of the chromophore depends much on the protein field, large differences in the excitation energy are observed for the different optimization levels. However, for all the structures, we find the protein electric field *to increase the excitation energy with respect to the in vacuo excitation energy of the chromophore in the protein*, while the Opsin shift corresponds *to a decrease of the excitation energy from that in methanol solution*. Our level of theory is high, but since the model is limited and polarization effects of the protein environment have not been taken into account, our results still differ 8 kcal/mol from the experimental value. We believe that the limitations of our model are understood, and that future studies at this level can fully account for the excitation in bR. In addition, the protein environment seems to be constructed in a specific way that lowers the excitation energy. This will be subject to further investigation.

Acknowledgements. The authors are grateful to Profs. H. Luecke and J. K. Lanyi for information on the bR X-ray structure. The present research is in part supported by a grant (CHE96–27775) from the National Science Foundation and by a grant from Gaussian Inc. Acknowledgement is made to the Cherry L. Emerson Center of Emory University for the use of its resources, which is in part supported by a National Science Foundation grant (CHE–0079627) and an IBM Shared University Research Award.

References

1. Needleman R (1995) In: Horspool WM, Song P-S (eds) CRC handbook of organic photochemistry and photobiology. CRC Press, Boca Raton, FL p 1508
2. Rothschild KJ, Sonar S In: Horspool WM, Song P-S (eds) CRC handbook of organic photochemistry and photobiology. CRC Press, Boca Raton, FL p 1521
3. Henderson R, Baldwin JM, Ceska TA, Zemlin F, Beckmann E, Downing KH (1990) J Mol Biol 213:899
4. Luecke H, Schobert B, Richter HT, Cartailler JP, Lanyi JK (1999) J Mol Biol 291:899
5. Birge RR, Zhang CF (1990) J Chem Phys 92:7178
6. Steen R van der, Biesheuvel PL, Mathies RA, Lugtenburg J (1986) J Am Chem Soc 108:6410
7. Houjou H, Koyama K, Wada M, Sameshima K, Inoue Y, Sakurai M (1998) Chem Phys Letters 294:162
8. Houjou H, Inoue Y, Sakurai M (2001) J Phys Chem B 105:867
9. Houjou H, Inoue Y, Sakurai M (1998) J Am Chem Soc 120:4459
10. Grigorieff N, Ceska TA, Downing KH, Baldwin JM, Henderson R (1996) J Mol Biol 259:393
11. Hayashi S, Ohmine I (2000) J Phys Chem B 104:10678
12. Maseras F, Morokuma K (1995) J Comp Chem 16:1170
13. Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K (1996) J Phys Chem 100:19357
14. Humbel S, Sieber S, Morokuma K (1996) J Chem Phys 16:1959
15. Dapprich S, Komáromi K, Byun KS, Morokuma K, Frisch MJ (1999) J Mol Str (Theochem) 461–462:1
16. Vreven T, Morokuma K (2000) J Comp Chem 16:1419
17. Vreven T, Mennucci B, da Silva CO, Morokuma K, Tomasi J (2001) J Chem Phys 115:62
18. Vreven T, Komáromi I, Dapprich S, Byun KS, Montgomery Jr. JA Morokuma K, Frisch MJ (2003) (in preparation)
19. Vreven T, Morokuma K, Farkas Ö, Schlegel HB, Frisch MJ (2002) (submitted)
20. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery Jr. JA, Stratmann RA, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Mennucci B, Cossi M, Adamo C, Jaramillo J, Cammi R, Pomelli C, Ochterski J, Petersson GA, Ayala PY, Morokuma K, Salvador P, Dannenberg JJ, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al–Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head–Gordon M, Replogle ES, Pople JA (2001) Gaussian 01, Development Version (Revision A.01). Gaussian Inc., Pittsburgh
21. Singh UC, Kollman PA (1986) J Comp Chem 7:718
22. Field MJ, Bash PA, Karplus M (1990) J Comp Chem 11:700
23. Théry V, Rinaldi D, Rivail JL, Maigret B, Ferenczy JJ (1994) J Comp Chem 15:269
24. Gao J, Amara P, Alhambra C, Field MJ (1998) J Phys Chem A 102:4714
25. Reuter N, Dejaegere A, Maigret B, Karplus M (2000) J Phys Chem A 104:1720
26. Bakowies D, Thiel W (1996) J Phys Chem 100:10580
27. http://www.gaussian.com/gvbroc.html. Cited 13 Oct 2001
28. Sasaki J, Lanyi JK, Needleman R, Yoshizawa T, Maeda A (1994) Biochemistry 33:3178
29. Brown LS, Sasaki J, Kandori H, Maeda A, Needleman R, Lanyi JK (1995) J Biol Chem 270:27122
30. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117–5179
31. Gao J (1996) In: Lipkowitz KB, Boyd DB Reviews in computational chemistry, vol 7.VCH, New York, p 119
32. Vreven T, Morokuma K (2000) J Chem Phys 113:2969
33. Froese RDJ, Morokuma K (1996) Chem Phys Lett 263:393
34. Froese RDJ, Komaromi I, Byun KS, Morokuma K (1997) Chem Phys Lett 272:335