Toward An Understanding Of The Retinal Chromophore In Rhodopsin Mimics

Mark M. Huntress
Samer Gozem
Konstantin R. Malley
Askat E. Jailaubekov
Chrysoula Vasileiou

See next page for additional authors

Follow this and additional works at: https://scholarworks.bgsu.edu/chem_pub

Part of the Chemistry Commons

Repository Citation
Huntress, Mark M.; Gozem, Samer; Malley, Konstantin R.; Jailaubekov, Askat E.; Vasileiou, Chrysoula; Vengris, Mikas; Geiger, James H.; Borhan, Babak; Schapiro, Igor; Larsen, Delmar S.; and Olivucci, Massimo, "Toward An Understanding Of The Retinal Chromophore In Rhodopsin Mimics" (2013). Chemistry Faculty Publications. 65.
https://scholarworks.bgsu.edu/chem_pub/65

This Article is brought to you for free and open access by the Chemistry at ScholarWorks@BGSU. It has been accepted for inclusion in Chemistry Faculty Publications by an authorized administrator of ScholarWorks@BGSU.
Toward an Understanding of the Retinal Chromophore in Rhodopsin Mimics

Mark M. Huntress,† Samer Gozem,‡ Konstantin R. Malley,‡ Askat E. Jailaubekov,‡ Chrysoula Vasileiou,§ Mikas Vengris,§ ‡ James H. Geiger,§ Babak Borhan,§ Igor Schapiro,† Delmar S. Larsen,*‡ and Massimo Olivucci*†

†Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43402, United States
‡Department of Chemistry, University of California Davis, One Shields Avenue, Davis, California 95616, United States
§Department of Chemistry, Michigan State University, Lansing, Michigan 48824, United States
¶Faculty of Physics, Vilnius University, Sauletekio 10 LT10223 Vilnius, Lithuania

ABSTRACT: Recently, a rhodopsin protein mimic was constructed by combining mutants of the cellular retinoic acid binding protein II (CRABPII) with an all-trans retinal chromophore. Here, we present a combined computational quantum mechanics/molecular mechanics (QM/MM) and experimental ultrafast kinetic study of CRABPII. We employ the QM/MM models to study the absorption (λ\text{max}), fluorescence (λ\text{fluorescence}), and reactivity of a CRABPII triple mutant incorporating the all-trans protonated chromophore (PSB-KLE-CRABPII). We also study the spectroscopy of the same mutant incorporating the unprotonated chromophore and of another double mutant incorporating the neutral unbound retinal molecule held inside the pocket. Finally, for PSB-KLE-CRABPII, stationary fluorescence spectroscopy and ultrafast transient absorption spectroscopy resolved two different evolving excited state populations which were computationally assigned to distinct locally excited and charge-transfer species. This last species is shown to evolve along reaction paths describing a facile isomerization of the biologically relevant 11-cis and 13-cis double bonds. This work represents a first exploratory attempt to model and study these artificial protein systems. It also indicates directions for improving the QM/MM models so that they could be more effectively used to assist the bottom-up design of genetically encodable probes and actuators employing the retinal chromophore.

1. INTRODUCTION

Rhodopsins comprise a family of light-absorbing proteins that have sensory and ion-pumping functions in a wide range of organisms.1,2 Rhodopsins feature an α-helical transmembrane structure (Figure 1A) hosting a retinal chromophore which remains isolated from the solvent in a protein pocket.3 The chromophore in such rhodopsins can normally be found in the 11-cis (in visual pigments), 13-cis (in microbial pigments), or all-trans conformation (in both visual and microbial pigments) and is covalently connected to a lysine residue via a protonated Schiff base linkage. The interaction of the chromophore with diverse protein environments induces a wide range of absorption maxima (λ\text{max}) ranging from 350 to 630 nm.4,5 A significant amount of this λ\text{max} determination (i.e., color tuning) results from changes in the chromophore binding pocket (i.e., the environment around the chromophore), although more extreme shifts can be induced by manipulating the protonation of the Schiff base linkage,6 or by chemical modifications of the chromophore7,8 (for instance, one which causes it to be held in an unbound state as a neutral retinal molecule).9

Received: June 16, 2012
Revised: August 1, 2013
Published: August 2, 2013
Figure 1. (A) Crystal structure of Rh (Protein Data Bank record 1U19, chain A). The 11-cis retinal chromophore (displayed in red) is buried in the binding pocket of the protein. (B) The all-trans protonated retinal Schiff base (PSBAT) linkage and the crystal structure of the KLE mutant of CRABPII (Protein Data Bank record 2G7B) binding a PSBAT chromophore (displayed in red). The structure is dominated by beta sheets forming a barrel shape, with a helix-turn-helix motif at the entrance to the binding pocket. The chromophore is partially exposed to solvent (displayed in blue).
CRABPII-based rhodopsin mimics can be easily engineered and crystallized, and are amenable to transient spectroscopies. Since crystal structures of natural rhodopsins cannot be routinely obtained due to difficulties in generating suitably diffracting crystals, these mimics provide ideal “laboratory” systems for investigating the possibility of engineering new PSBAT based fluorescent proteins. In fact, their crystal structure allows for the construction of quantum mechanics/molecular mechanics (QM/MM) models of CRABPII mutants that can be used to guide rational design approaches.

To engineer rhodopsin mimics into functioning probes, the availability of modeling strategies capable of assisting the search for mutants with desired \( \lambda_{\text{max}} \) values and photochemical reactivity (Figure 2) is highly desirable. For this reason, below we present a combined computational QM/MM and ultrafast kinetic experimental study of a CRABPII system representing a first step toward the exploration of the photophysics of these new protein systems. Note that this is a first exploratory and, possibly, crude attempt to model and study these engineered proteins, paving the way for future systematic studies of such systems. In particular, here we investigate three systems for which experimental data is available: the aforementioned PSB-KLE-CRABPII protein, the corresponding form (SB-KLE-CRABPII) hosting the unprotonated form of PSBAT (SBAT), and the KF-CRABPII (R132KY134F) mutant with an unbound retinal molecule held in the binding pocket. The results are contrasted with those of analogue studies of solvated PSBAT and Rh. We employ QM/MM models of these systems in an attempt to reproduce the observed \( \lambda_{\text{max}} \) value (process a in Figure 2) of PSB-KLE-CRABPII, PSB-KLE-CRABPII, and KF-CRABPII. We also explore the chromophore fluorescence (process b in Figure 2) in PSB-KLE-CRABPII, for which faint fluorescence was experimentally observed. Complementary time-resolved absorption spectroscopy and excited state QM/MM scans provide information about the PSB-KLE-CRABPII photoisomerization mechanism/dynamics and photoproduct generation (processes c and d in Figure 2, respectively).

2. METHODS

2.1. Computations. Protein models capable of addressing diverse processes such as light-absorption, light-emission, and photoisomerization require quantum chemical methods capable of describing both ground and electronically excited states of the chromophore with comparable accuracy. The \textit{ab initio} complete-active-space self-consistent-field (CASSCF) method is a multiconfigurational method offering enough flexibility for an unbiased description (i.e., with no empirically derived parameters and with multideterminant wave functions) of the electronic and equilibrium structures of the ground and excited states. Furthermore, the CASSCF wave function can be used for subsequent single-state and multistate multiconfigurational second-order perturbation theory (CASPT2) and MRCASPT2 computations to account for dynamic electron correlation when evaluating the energy gap between different electronic states or the energy barrier found for a given state. In this protocol, energies and equilibrium structures are determined at the CASPT2 and CASSCF levels of theory, respectively (we use the notation CASPT2//(CASSCF)), and using the 6-31G* basis set. It has been shown that a CASPT2/CASSCF based QM/MM protocol allows one to reproduce the \( \lambda_{\text{max}} \) of different proteins incorporating cationic, anionic, and neutral chromophores. For visual and sensory rhodopsins, the excitation energies associated with the experimentally observed \( \lambda_{\text{max}} \) values have been reproduced with a root-mean-square error of \( \sim 3 \text{kcal/mol} \). However, it has to be pointed out that these rhodopsins are less complex and less configurationally flexible than the artificial systems modeled in the present work. Therefore, the same type of QM/MM model may not display similar performances. Indeed, the evaluation of the applicability of such models to rhodopsin mimics is among the targets of the present work. However, although similar QM/MM studies on natural rhodopsins have been carried out with different quantum chemical protocols, CASPT2//CASSCF is a practical method for a consistent description of spectra, reaction paths, conical intersections, and trajectories. Furthermore, the error associated with such a method has been recently examined. When using CASSCF equilibrium structures, CASPT2 (with IPEA = 0) and a double-\( \zeta \) basis set with polarization function computed energy gaps are comparable to experiment. The technical reason, ascribed to a balanced error cancelation, has also been investigated quantitatively in a recent contribution. Moreover, it has also been shown that CASSCF equilibrium geometries optimized in a protein environment may be closer to the correct geometry than when optimized in the gas phase.

The QM/MM models (see sections 1.a, 1.b, and 1.e in the Supporting Information for details) of the CRABPII constructs were built on the basis of heavy atom coordinates from known crystal structures from the Protein Data Bank (2G79 and 2G7B) of 1.69 and 1.18 Å resolution, respectively. After addition and MM minimization of hydrogens, the protein was placed in a 10 Å by 10 Å solvent box of water, and all the hydrogens of the water molecules were optimized. Next, a 1 ns molecular dynamics (MD) simulation was carried out at the MM level with the NAMD software and AMBER parameters to sample the internal water orientations and to account for the fluctuations in the hydrogen bond network (HBN). The waters were unconstrained, but the protein and
chromophore atoms (with the exception of the E121 counterion side chain) were held fixed at their crystal structure positions during the MD. Ten geometries were sampled from the last 200 ps of the trajectory and were used for QM/MM geometry optimization at the CASSCF/AMBER level (with an active space of 12 electrons in 12 orbitals spanning the entire π-system of the chromophore) to generate a set of models suitable for subsequent CASPT2//CASSCF/AMBER calculations. All QM/MM calculations were performed using Molcas 7.56 and utilized electrostatic embedding with a hydrogen link atom at the truncated C–C bond of the QM/MM boundary. The QM segment included the chromophore, the C2 and N of the lysine side chain, and hydrogens attached to these two atoms. The chromophore, counterion side chain, and proximal waters within 5 Å of the chromophore were optimized. The rest of the protein was kept frozen during the optimization. Therefore, we make the assumption that the high-resolution crystal structure provides an acceptable average representation of the protein backbone and side-chain structures. The \( \lambda_{\text{max}} \) was evaluated using three-root MS-CASPT2 calculations to alleviate the state mixing problems that were observed at the CASPT2 level between \( S_1 \) and \( S_2 \). Due to possible artifacts arising in CASPT2 and MS-CASPT2 near degeneracies,64,87 a discussion of this point is reported in section 1.i of the Supporting Information. The scans are de

The resulting PSB-KLE-CRABPII, SB-KLE-CRABPII, and KF-CRABPII models used in this study were compared to those of (1) a Rh model67 built using the same protocol to serve as a reference and starting from the 1U19 crystal structure (see section 1.g in the Supporting Information),5 (2) a model of PSBAT in water (see section 1.h in the Supporting Information), (3) two point mutants (A36D and A36H) of PSB-KLE-CRABPIII that were created in silico to investigate the effect of introducing a negative and a positive charge near the poorly solvated region of the chromophore (see section 1.f in the Supporting Information). To account for the change in protein structure from the mutations, all side chains within 4.0 Å of A36 were optimized for each protein system.

All protein models are constructed with a substantially consistent, although approximated, protocol to allow for comparison. Of course, it was necessary to make some assumptions regarding the protonation state of E181 in Rh (see section 1.g of the Supporting Information) and to assume that the CASPT2//CASSCF/6-31G* method provides an adequate quantum chemistry level for the present exploratory investigation. On the other hand, even reasonably well reproduced quantities, such as \( \lambda_{\text{max}} \) and \( \lambda_{\text{max}}^2 \), obtained with the employed level have been reported to rely on error cancellation effects, as shown for a gas-phase minimal model of the retinal chromophore.81

To investigate the internal conversion and photoisomerization mechanisms of PSB-KLE-CRABPII, we computed relaxed scans along its excited state potential energy surface to get information on the reaction path structure. Because the isomerization could occur along different bonds (either C10—C11=C12—C13 or C12—C13=C14—C15), and could also occur along different directions (clockwise or counterclockwise due to the chirality of the protein environment), we computed four scans (in each direction along both the C10—C11=C12—C13 and C12—C13=C14—C15 dihedrals). The scans were generated at the CASPT2//CASSCF/AMBER level, consistent with similar scans reported for Rh models74 (CASPT2 is more reliable than MS-CASPT2 in the absence of state mixing).57,89–96 As the chromophore twists, the state mixing is reduced due to the divergence of the \( S_1 \) and \( S_2 \) energies (see Figure S8 and section 4.a of the Supporting Information). The scans are defined by 10° steps in the C10—C11=C12—C13 and C12—C13=C14—C15 dihedral angles of the PSBAT chromophore to track the corresponding isomerization mechanisms.

In order to reinforce the conclusions of the PSB-KLE-CRABPII excited state mapping, we also present gas-phase calculations of a reduced model of PSBAT (PSB3 in the following) featuring three conjugated double bonds: C5H=CH=CH=C3H=CH=CH=CH2.88–88,88,88 The light emission and evolution on the excited state potential energy surface were explored at the CASPT2//6-31G* level with an IPEA shift of 0.00 and an imaginary shift of 0.2. The employed active space comprises six electrons in the six orbitals spanning the PSB3 π-system.

Relaxed scans were performed along the C3=C2=C1=N dihedral (describing the twisting about a single bond) by constraining the dihedral value at several progressing values, and also constraining the H=C2=C1=H dihedral value in order to avoid pyramidalization. A second relaxed scan was computed with the same modalities following the double bond isomerization coordinate around the central C4=C3=C2=C1 dihedral. Any excited state minimum or maximum found along the scan was then optimized with no constraints.

2.2. Static and Transient Spectroscopic Measurements. Static fluorescence spectra were recorded with a Fluorolog-3 fluorometer (Instruments S. A., Inc.). All-trans retinal was purchased from Toronto Research Chemicals and was used as received. The emission spectrum of a 3 mM solution of the KLE-CRABII mutant in phosphate buffered saline (5 mM, pH 7.3) was recorded (\( \lambda_{\text{ex}} = 440 \text{ nm} \), spectrum range: 470–800 nm). One equivalent of all-trans retinal was added (3 µL of a 5 mM all-trans retinal solution in ethanol), the solution was mixed, and the emission spectrum was recorded. The PSB-KLE-CRABII protein–retinal complex was also recorded on the same instrument.

The full details of the experimental setup for the transient absorption signals have been described elsewhere.97 Briefly, the 400 nm pump pulse was generated by the frequency doubling of the fundamental 800 nm output of a Ti:sapphire laser system (Spectra Physics Tsunami and Spitfire Pro) in a β-barium borate crystal to produce 2 μJ pulses of 50 fs duration at the 400 nm excitation wavelength. The probe pulse was produced by passing a 1 μJ portion of the 800 nm fundamental beam through a slowly translating 2 mm thick CaF2 crystal to generate the broadband white light continuum used for probing. The pump beam was linearly polarized at the magic angle (54.7°) with respect to the probe beam to avoid reorientation effects from contributing to the signals. Samples were flowed through a custom-built cuvette constructed with 150 µm thick SiO2 windows to reduce interfering cross-phase modulation artifacts98,99 and to improve the time resolution by reducing pump–probe walk off. A peristaltic pump (Watson Marlow 401U/D) was used to flow the sample at a sufficient
speed to ensure replenishment of new sample between successive laser shots (1 ms). The absorbance of the samples was set to approximately 0.8 at the 800 μm path length of the cuvette.

PSBAT samples were prepared from all-trans retinal (Sigma) and n-butylamine (Aldrich) according to previously described methods.100 A molar excess of n-butylamine was added to an 11 mM solution of all-trans retinal in anhydrous diethyl ether. The mixture was allowed to react for 24 h over 3 Å molecular sieves to absorb the water produced from condensation. The solvent was then evaporated under reduced pressure, and the sample was washed with diethyl ether three to four times to remove unreacted amine. After redissolving the sample in methanol, it was subsequently protonated with an excess of HCl to generate PSBAT, which is analogous to the linkage between a Lys residue and the retinal chromophore in Rh and bR.101 The free unprotonated SBAT in methanol absorbs at 364 nm and shifts to 445 nm upon protonation.

The protocol, regarding CRABPII proteins, for binding all-trans retinal to the Lys132 residue of KLE-CRABPII as PSBAT, and the static spectroscopic properties of the PSB-KLE-CRABPII have been described previously.50,51 The PSB-KLE-CRABPII sample was suspended in a phosphate buffer solution of pH 7.4 and adjusted dropwise with 0.2 M HCl to pH 7.2 to ensure that the Schiff base was fully protonated as determined spectroscopically. The static absorption spectra of the samples measured before and after laser excitation exhibit no significant change in shape or amplitude.

3. RESULTS AND DISCUSSION

3.1. Femtosecond Transient Dynamics. The femtosecond transient spectra for PSB-KLE-CRABPII are contrasted with PSBAT in methanol in Figure 3. The PSBAT signals (Figure 3A) are nearly identical to those previously reported,18,19,102,103 with a negative ground state bleach (GSB) peaking at 425 nm and a negative stimulated emission (SE) band peaking at 610 nm, which both overlap a positive narrow excited state absorption (ESA) band at 505 nm. The rapid decay of these signals is attributed to the nonadiabatic quenching of the first singlet excited state (S1) population due to twisting.18,19,102,103 Concurrent with decay, this is a red-shifting of the SE (from 610 to 650 nm) and a blue-shifting of the ESA (from 505 to 490 nm) that is attributed to vibrational relaxation on the S1 surface.18,19,104–106 The PSB-KLE-CRABPII signals (Figure 3B) exhibit qualitatively similar spectroscopic features to the PSBAT signals with clearly resolvable GSB and ESA bands (although distinctly broader) and a SE that is observed only within the first ~500 fs after photoexcitation (600–675 nm).

Although both PSBAT in solution and PSB-KLE-CRABPII exhibit similar transient spectra, the quantitative differences of the dynamics between the two samples are clearer in the kinetic traces (Figure 4). While the PSBAT kinetics follow similar decay kinetics across all probe wavelengths, the evolution of the PSB-KLE-CRABPII signals varies strongly with probe wavelength (e.g., the 470 nm kinetics are significantly more slowly evolving than the 528 nm kinetics, which in turn evolves more slowly than the 600 and 637 nm kinetics). The greatest difference between the two samples lies in the low energy region (Figure 4C,D) with the PSB-KLE-CRABPII signals at 600 and 637 nm exhibiting considerably faster sub-ps decay of the SE with the growth of a new positive absorption. To extract a more complete picture of the underlying photodynamics, multiwavelength global analysis is used to simultaneously analyze the signals at all measured probe wavelengths107,108 and to fit the data into a small number of time-dependent interconnected populations with fixed spectra.

A simple sequential analysis with a linear model (Figure 5A, inset) was used for the PSBAT dynamics to extract the spectra and time scales of constituent populations in the data.109 Four populations (1–4 in the figure) are resolved, which track the quenching and vibrational relaxation dynamics with the first three populations ascribed to S1 evolution due to the presence of the SE and a weak final 250 ps component (blue curves) that is ascribed to the primary isomerized cis photoproduct. These dynamics can be ascribed to S1 to ground state (S0) quenching and vibrational relaxation with the red-shifting SE and blue-shifting ESA on S2 as discussed above. In contrast to the PSBAT in solution data, the PSB-KLE-CRABPII dynamics (Figure 5B) requires the use of a more complex nonlinear model that explicitly accounts for the microscopic evolution between coevolving populations. Multiple models were evaluated with the final model providing the best fit to the data within self-consistent constraints (e.g., comparable GSB
amplitudes). The six-compartment model involves a rapid (<100 fs) bifurcation of the Franck–Condon (FC$S_1$) population into two coexisting “relaxed” $S_1$ populations ($S_{1c}$ and $S_{1s}$; the reason behind our use of this nomenclature will be explained later) with identical spectra (red curve). While $S_{1c}$ rapidly decays (660 fs) into a red-absorbing $S_0$ photoproduct (which is tentatively assigned to formation of PSB11 or PSB13, red curve to green curve), $S_{1s}$ persists (∼8 ps) and exhibits vibrational relaxation (red curve to blue curve) similar to PSBAT in solution. The experiment does not allow for the identification of the photoproduct, but the labeling in Figure 5C is consistent with the mechanistic hypothesis and supporting computational results discussed in section 3.3 below. These results also indicate that $S_{1c}$ may decay via an unidentified excited state rotamer (single bond twisting), which is also unidentified in the modeling and eventually relaxes to $S_0$ reverting back to PSBAT. Unfortunately, this population overlaps strongly with the ESA of the excited state and is poorly spectrally and temporally resolved, so all photoproducts generated from the different excited state pathways are modeled with the same spectrum (green curve).

### 3.2. Vertical Absorption Energies

The PSB-KLE-CRABPII chromophore has an observed $\lambda_{\text{max}}$ of 449 nm.\(^{50,51}\) The reported $\lambda_{\text{max}}$ values of PSBAT and PSB11 in methanol are 445 and 442 nm,\(^{11}\) respectively. Note that experimental results are only available in methanol, since PSBAT is unstable in water. The $\lambda_{\text{max}}$ of PSBAT in methanol is almost identical to that of PSB-KLE-CRABPII, and largely blue-shifted with respect to that of Rh (498 nm).\(^{38}\) Our first goal was to see if QM/MM models could reproduce the corresponding vertical excitation energy tuning by the protein and solvent environment.

The initial PSB-KLE-CRABPII QM/MM model (i.e., without sampling the environment) yielded $\lambda_{\text{max}}$ values of 367 and 372 nm, at CASPT2 and MS-CASPT2 levels, respectively, corresponding to excitation energy errors of 14.2 and 13.2 kcal/mol, respectively, when compared to the experimental value of 449 nm. This inaccurate result is attributed to a poor modeling of the average location of the highly mobile parts of the protein, especially the complex HBN structure. Therefore, in order to sample the HBN networks, classical MD simulations with a parametrized chromophore (sections 1.c and 1.d in the Supporting Information) were used to obtain different starting solvent configurations. Ten “snapshots” were then obtained from different times from the MD simulation, which were then optimized with CASSCF/AMBER and used for excitation energy calculations. This method gave an average $\lambda_{\text{max}}$ value of 400 nm with CASPT2 and 431 nm with MS-CASPT2, corresponding to a 7.8 and 2.7 kcal/mol difference from the observed excitation energy, respectively (Figure 6A). The big difference between the CASPT2 and MS-CASPT2 excitation energies is a consequence of the fact that only the second method allows for state mixing (see section 2.1). However, due to the present MS-CASPT2 formulation, the correct excitation energies would be expected to lie somewhere in between the CASPT2 and MS-CASPT2 values (see the comparison with XMCQDPT2 in section 1.i of the Supporting Information).

The results above indicate that MD sampling is essential for improving the description of the environment around the chromophore in KLE-CRABPII. This is because, in configurationally and conformationally variable systems, the sampling allows us to describe the average effect of the environment which may differ significantly from that of a selected configuration/conformation. In our models, in spite of the improvement with respect to a single configuration, the result of a limited sampling still does not adequately reproduce the experimental data, pointing to the fact that further improvements of the model are required.

While the $pK_a$ of the Schiff base linkage is high in Rh and bR (13.3\(^{10}\) and 16.6\(^{11}\) respectively), meaning that it is protonated in both proteins, the Schiff base in our mimics has a $pK_a$ of 8.7 and can be either protonated or deprotonated at physiological pH.\(^{51}\) Therefore, our models should be able to describe both
protonated or deprotonated (SBAT) chromophores correctly. As such, the sampling protocol applied to PSB-KLE-CRABPII, which features a carboxylate and protonated Schiff base forming an ion pair, was also applied to the SB-KLE-CRABPII and KFCRABPII (Figure 6A) models featuring a carboxylate polarizing SBAT and a guanidinium group polarizing all-trans retinal, respectively. In spite of the chemical diversity of chromophores and environments, the three models reproduce the experimentally observed $\lambda_{\text{max}}$ values with similar errors (for calculated energies and oscillator strengths, see sections 3.b and 3.c of the Supporting Information). The chromophore—counterion distances are correlated to the degree of localization of the positive charge in the chromophore. The double bond alternation in the region of the Schiff base (−N=C15—C14=N=C13—) is different for PSBAT and SBAT, the latter of which exhibits a decreased bond alternation with a longer −N=C15 double bond and a shorter C15—C14 single bond (Figure S5 in the Supporting Information).

For comparison, we provide the results of the sampling for the absorption of PSBAT isolated in water (Figure 6A). We find that the model of PSBAT in solution yields a reasonable excitation energy compared to the experiment if we assume that the water and methanol environments are similar (the excitation energy computed for PSBAT in water is 67 kcal/mol compared to 64.3 kcal/mol in methanol). The similarity of the excitation energy for PSBAT in solution and in KLECRABPII (66 kcal/mol) then suggests that the solution and KLECRABPII environments have a similar color tuning effect. This is tentatively ascribed to the possibility that the solvent incorporated in the KLE-CRABPII chromophore cavity plays a dominant role in tuning the excitation energy (i.e., by creating a solvent-like microenvironment). Indeed, KLE-CRABPII has a cavity that allows a large number of solvent molecules to surround the chromophore, unlike Rh, whose chromophore cavity only hosts two1 or three5 water molecules. Note that this solvent-like environment is found in KLE-CRABPII but may not be found in other CRABPII mutants. Indeed, mutants of a similar rhodopsin mimic, based on the human cellular retinol binding protein II, have been engineered with intrinsic chromophore structures of PSB-KLE-CRABPII and PSBAT in water solution (solvated PSBAT). (B) Top: Change in the charge distribution along the backbone of the isolated chromophore of PSB-KLE-CRABPII upon photoexcitation. The size of each bubble in the diagram reflects the value of the charge on the corresponding center. The numbers in the diagram represent the minimum and maximum values of the Mulliken charges (with hydrogen charges summed together with the heavy atom charges). Bottom: An environment imposing a more positive potential (blue color) on the chromophore C≡N segment or a more negative potential on the β-ionone segment (red color) will red-shift the $\lambda_{\text{max}}$. Reversing the displayed electrostatic potential results in a blue-shifted absorption. (C) The effect of the environment on excitation energies is evaluated. The bold lines represent the $S_0\rightarrow S_1$ excitation of the extracted chromophores in the gas phase. The solvated PSBAT environment (water) and the PSB-KLE-CRABPII (KLE) environment both have a similar effect, and the shift due to single mutations is evidently due to the electrostatic changes in the environment rather than a change in chromophore geometry.

**Figure 6.** Electronic structure of the ground ($S_0$), spectroscopic excited state ($S_1$), and forbidden state ($S_2$) of the protonated Schiff base of retinal. (A) MS-CASPT2 excitation energies for the CASSCF/AMBER optimized structures obtained starting from 10 MD snapshots. The same data are presented for models containing not only the PSBAT chromophore (PSB-KLE-CRABPII) but also the corresponding unprotonated chromophore (SB-KLE-CRABPII), a mutant containing retinal (KF-CRABPII), and PSBAT in water solution (solvated PSBAT). (B) Top: Change in the charge distribution along the backbone of the isolated chromophore of PSB-KLE-CRABPII upon photoexcitation. The size of each bubble in the diagram reflects the value of the charge on the corresponding center. The numbers in the diagram represent the minimum and maximum values of the Mulliken charges (with hydrogen charges summed together with the heavy atom charges). Bottom: An environment imposing a more positive potential (blue color) on the chromophore C≡N segment or a more negative potential on the β-ionone segment (red color) will red-shift the $\lambda_{\text{max}}$. Reversing the displayed electrostatic potential results in a blue-shifted absorption. (C) The effect of the environment on excitation energies is evaluated. The bold lines represent the $S_0\rightarrow S_1$ excitation of the extracted chromophores in the gas phase. The solvated PSBAT environment (water) and the PSB-KLE-CRABPII (KLE) environment both have a similar effect, and the shift due to single mutations is evidently due to the electrostatic changes in the environment rather than a change in chromophore geometry.
both PSB11 and PSBAT and is substantially independent of the stereochemistry of the C11–C12 bond. This is consistent with the comparison of Rh to the mirror-image structure of PSBAT in solution that shows a considerable overlap of geometries (Figure 7B, dashed green curve). For Rh, our conclusion that the geometry of the chromophore in the protein has a small
effect on color tuning conflicts with the conclusions of Bravaya et al.\textsuperscript{57} In their study, a third water molecule was added to the model near the protonated Schiff base. It formed a hydrogen bond that caused a torsional distortion of 48° in C13=CH−CH=C15=N which in turn had a significant effect on the calculated excitation energy. Our model included only the two waters seen in the crystal structure and, as a result, had only a 15° distortion in the C13=CH−CH=C15=N dihedral (Figure 8A).

The blue-shifting effect of the environment on vertical excitation energies can be understood on the basis of the charge distributions of the electronic structures of the S_{0}, S_{a}, and S_{i} states in PSBAT (Figure 6B). Although the charge distributions of the S_{a} and S_{i} states are similar, the charge distribution of the charge-transfer (S_{t}) state is marked by a translocation of the positive charge from the −C15=N− Schiff base region toward the \( \beta \)-ionone ring. As shown in Figure 6C, our calculations indicate that the environments of the Rh, PSB-KLE-CRABPII, and solvent cavities blue-shift the \( \lambda_{\text{max}} \) values of the corresponding chromophores with respect to the gas phase. Furthermore, inspection of the data reveals the following: (i) This effect is only 5 kcal/mol. This relatively small change is interpreted in terms of a “counterion quenching” effect which refers to the fact that the large blue-shift induced by the counterion in the chromophore absorption is counter-balanced by a red-shift induced by the rest of the protein residues.\textsuperscript{62} However, it has to be pointed out that in a previous Rh model by Andruniow et al.\textsuperscript{62} we found an effect even smaller than 5 kcal/mol. This is explained by the differences between the present Rh model and the Andruniow et al. model (e.g., the use of the 1U19 instead of the 1HZX crystallographic structures), indicating that the exact magnitude of the counterion quenching depends on how the model has been constructed.\textsuperscript{3,4,7,51,111,113} Also, notice that the documented counterion quenching refers to Rh (i.e., bovine rhodopsin) and shall not be generalized to all rhodopsins, which feature \( \lambda_{\text{max}} \) values ranging from 350 to 630 nm.\textsuperscript{45} (ii) The blue-shift induced by both the KLE-CRABPII cavity and solvent cavity is larger than the one computed for the Rh cavity. In KLE-CRABPII, this is ascribed to the presence of a carboxylate counterion whose effect is not counterbalanced by the rest of the protein residues. In fact, as explained above, this system displays a solvent-like microenvironment where the solvent is screening the effect of the protein residues but not the effect of the counterion. Notice that, as previously reported,\textsuperscript{114−117} in the solvent environment the effect of the counterion is replaced by a cluster of solvent molecules with their negative poles (e.g., oxygens in H_{2}O and MeOH) oriented toward the positively charged Schiff base group. This also results in a shielding of the real counterion (e.g., Cl−) called the “leveling effect”.\textsuperscript{178}

The effects described above can be easily visualized and quantitatively studied by mapping the electrostatic potential imposed by the protein pocket on the van der Waals surface of the chromophore (Figure 8). In all cases, the environment blue-shifting effect is due to a negative potential projected on the positive −C15=N− Schiff base charge (red region). The maps can also explain the red-shift observed in Rh relative to the PSB-KLE-CRABPII and solvated models. The excitation energy of Rh is 6.3 kcal/mol red-shifted from PSB-KLE-CRABPII due to the weaker negative electrostatic potential projected on the protonated Schiff base region of the chromophore in Rh and the increased negative potential around the C7 and C8 atoms near the \( \beta \)-ionone ring (Figure 8A) which partially counterbalances the effect of the counterion.
(indicated as $^{C\text{T}}S_1$ and $^{L\text{E}}S_1$) appear within 150 fs and are tentatively associated with the same reactive and nonreactive species observed in methanol. In the following, this assignment will be discussed on the basis of a computational investigation of the $S_1$ potential energy surface.

In order to simulate the fluorescence (process b in Figure 2), the average “relaxed” $S_1$ structure(s) must be determined. To determine the nature of the $S_1$ species and take into account the PSBAT and HBN structural variability, a short (ca. 50 fs) QM/MM trajectory was computed to describe the fast initial $S_1$ relaxation of PSB-KLE-CRABPII and starting from frame 3 in Figure 6A. Only the PSBAT and waters were allowed to relax (to sample the HBN network), while the protein was held fixed. Rapid oscillations in the bond length alternation (BLA; defined as the difference between the average double bond and average single bond lengths) of $S_1$ PSBAT were observed during the simulation, indicating that the structural changes mainly involve stretching modes. The BLA associated with the two extremes of the oscillation are markedly different, with the first exhibiting an almost “even” BLA pattern featuring single and double bonds having similar bond lengths (the BLA value is close to zero); the other extreme is a highly “inverted” BLA pattern (Figures 9A and S6 of the Supporting Information depict these geometries, and the coordinates are presented in sections 2.f and 2.g of the Supporting Information). In order to look for possible $S_1$ minima featuring different BLA patterns, two trajectory snapshots displaying extreme BLA values were used as starting points for classical MD simulations to sample the HBN structure (as done above for the absorption maxima). For each MD computation, 10 evenly spaced snapshots were selected for subsequent QM/MM geometry optimizations. The resulting 20 $S_1$ equilibrium structures produced, exclusively,
structures with an inverted BLA value and yielded an average vertical excitation energy of 41 kcal/mol (706 nm). This number has to be compared with the observed emission $\lambda_{\text{max}}$ values (see below). As would be expected with the inverted BLA, all structures feature a positive charge translocated toward the $\beta$-ionone ring and therefore are part of a charge-transfer (CT) population. The original central double bonds (e.g., C11=C12 and C13=C14) now have single bond orders and can twist easily. These excited state minima are therefore associated with a reactive (with respect to double bond isomerization, see section 3.4) CT $S_1$ species.

The experimental fluorescence spectrum, however, has a shoulder that suggests the presence of two bands (Figure 9C). A two-peak deconvolution performed with PeakFit (section 3.4 in the Supporting Information) resolves two bands centered at 600 and 665 nm (Figure 9C) that were associated with the CT $S_1$ and LE $S_1$ populations, respectively, and were presented in the ultrafast data in Figure 5. The 665 nm maximum is associated with the computed 704 nm $\lambda_{\text{max}}$ value. The reactive population CT $S_1$ would therefore correlate with a CT population characterized by an inverted BLA. The second 600 nm maximum would correspond to the alternative LE $S_1$ nonreactive population, but the CASSCF/AMBER calculations were not able to locate the corresponding equilibrium structure on the $S_1$ potential energy surface.

The fact that CASSCF/AMBER optimizations starting from 20 MD snapshots with very different BLA values all produced $S_1$ structures with similar BLA patterns could be due to the fact that CASSCF does not include dynamic electron correlation, thus failing to locate alternative $S_1$ minima. For this reason, we evaluated the CASPT2 energy profile along a CASSCF/AMBER “relaxed” scan starting from 1 of the 20 optimized structures (point 4 in Figure 9B was selected because its energy is closest to the average energy) and moving along points with decreasing BLA values (i.e., points displaying single and double bond lengths yielding BLA values getting closer to that of the $S_0$ equilibrium structure). As the results in Figure 9D show, this energy profile indicates the existence of a $S_1$ potential energy surface with two flat minima. While the first minimum corresponds to the CT structure described above (excitation energies in the 656–717 nm range), the second features a locally excited (LE) structure characterized by an even BLA in the central part of the molecule and by a positive charge delocalized on the same fragment. The LE structure is energetically located slightly below the CT structure, with a computed emissive energy gap in the 575–579 nm range that falls close to the primary fluorescence band $\lambda_{\text{max}}$ at 600 nm (Figure 9C). A stabilization of LE structures upon inclusion of the dynamic correlation is supported by reported CASPT2 geometry optimizations on computationally affordable reduced retinal chromophore models and in a recent QM/MM study of 11-cis locked bovine Rh. These studies demonstrate the existence of an energy minimum displaying an even BLA in the central part of the molecule.

Due to the far too high cost of CASPT2 optimizations, the CASPT2 BLA scan of Figure 9D was performed on geometries optimized with a CASSCF-based methodology. As a consequence, the flat LE minimum needs further investigation. To address this problem, the PSB3 (see section 2.1) reduced model of PSBAT was used to study the $S_1$ potential energy surface, and to provide further support for the existence of the two excited state species. PSB3 exhibits similar excited state behavior to PSBAT and allows for the use of CASPT2 geometry optimization due to its smaller size (Figure 10). LE and CT flat regions were indeed found on the $S_1$ energy surface. Due to its double bond/single bond inversion, the CT structure can readily isomerize. In fact, the twisting deformation about the central C2=C3 double bond (see angle $\alpha$ in Figure 10) produces a short flat $S_1$ region that gives access to a lower lying conical intersection between the $S_1$ and $S_0$ potential energy surfaces. The LE structure, on the other hand, is located at the center of a flat $S_1$ region mainly spanned by the twisting deformation about a single C1—C2 bond (see angle $\beta$ in Figure 10). Therefore, the evolution of LE does not seem to
lead to an isomerization product but to a lower energy conformer with a ca. 105° twisted single bond.

The LE and CT species proposed above will display different photophysics, possibly producing a dual-fluorescence signal if they slowly interconvert (with respect to the fluorescence lifetime). This hypothesis seems to be consistent with the assignment of CT and LE to the observed CTS1 and LES1 populations, respectively. This possibility is also consistent with recent computational and experimental reports looking at the excited state behavior of PSBAT in methanol, which argue for the existence of two different S1 populations with different spectral and temporal properties. One state was assigned to a nonreactive S1 intermediate (not leading to isomerization) with a 2.8 ps lifetime in methanol and the other to a reactive intermediate with a lifetime of 5.2 ps. In addition, averaged solvent electrostatic potential/molecular dynamics (ASEP/MD) computations (a QM/MM approach that uses a mean field approximation) at the CASPT2//CASSCF level of a PSBAT model in methanol support the possibility of dual fluorescence in these systems. The fluorescent species would have, according to the authors, CT character and covalent/diradical character due to near degeneracy and crossing of the S1 and S2 potential energy surfaces that display different electronic structures. However, to explain a dual fluorescence, these species should both have high oscillator strength and the S2 state must thus have a mixed character supposedly close to the one indicated as a LE state in the present contribution.

3.4. Excited State Evolution. Both the transient dynamics of PSBAT in solution and the PS-KLE-CRABPII signals (see section 3.1) indicate that photoproducts are generated upon S1 decay. To determine which bond could potentially isomerize in the PS-KLE-CRABPII pocket, relaxed S1 scans along both C10—C11=C12—C13 and C12—C13=C14—C15 dihedrals were computed starting at the CT minimum of Figure 9D (corresponding to point 4 in Figure 9B) where the central double bonds have single bond character and are prepared for isomerization. No relaxed scans have been computed for the LE pathway.

In Figure 11, we present results of four scans computed in different directions (clockwise or anticlockwise) along two different dihedral angles (around C11=C12 or C13=C14). An S0 and S1 degeneracy was reached corresponding to a conical intersection after as little as 60° rotation for the C13=C14 twisting (Figure 11B), consistent with what has been reported for bR and ASR models, or after as much as 90° rotation for the C11=C12 twisting, consistent with what has been reported for Rh.

Three of the four scans indicate that the S1 population is driven toward a conical intersection leading to rapid decay to S0. Indeed, counterclockwise rotation around either C11=C12 and C13=C14 and clockwise rotation around C13=C14 occurs along a flat potential energy surface with less than 1.5 kcal/mol barriers. In contrast, the clockwise rotation around C11=C12 has a ca. 4 kcal/mol barrier and should be, according to our results, less competitive. The origin of this barrier is ascribed to the steric interactions of P39 with the chromophore’s methyl group that swings out to within 2.27 Å of the P39 side chain at a 90° twisted geometry (Figure S9, Supporting Information). The fact that the scans indicate only one possible direction of rotation for formation of the PSB11 product and two possible directions of rotation for formation of PSB13 product from the CT minimum supports the tentative photoproduction assignments in Figure 5.

The results above indicate that the S1 population may decay via different reactive processes (C11=C12 and C13=C14 twisting) controlled by small barriers located along flat potential energy surface regions. These processes would limit the S1 lifetime and the fluorescence efficiency (e.g., process b in...
Figure 2). However, notice that the approximations made do not presently allow the correct magnitude of the barrier to be determined (which is expected to be at most on the order of a few kcal/mol), whether rotation around $C_{11} = C_{12}$ or $C_{13} = C_{14}$ is preferred or if the following decay leads to an observable photoproduct. Also, the decay of the LE population is not investigated here. This may result in a twisted conformer (i.e., about a single bond) of the excited state chromophore (contributing to the fluorescence with a hypothetically further red-shifted emission with respect to the ones reported in Figure 9C and consistent with the PSB3 model of Figure 10). For the moment, this remains a speculation that may be the target of future investigations.

3.5. Photoproducts. The PSB11 and PSB13 photoproducts were modeled assuming that the protein remains substantially unchanged so that only the HBN and isomerized chromophore are relaxed during the calculation. Accordingly, the last structures of the counterclockwise 11-cis and 13-cis scans (Figure 11A and B, respectively) were relaxed on $S_0$ to generate the PSB11 and PSB13 equilibrium structures (Figure 12C, left). The PSB13 product was found to be 4.3 kcal/mol more stable than the PSB11 product at the CASPT2 level. The photoproduct $\lambda_{max}$ values were calculated at the CASPT2 level to be at 466 and 492 nm for PSB11 and
PSB13, respectively. These transitions are 5 and 8 kcal/mol red-shifted from the calculated 431 nm $\lambda_{\text{max}}$ of the PSB-KLE-CRABPII reactant, respectively. Since the protein pocket of CRABPII is adapted to bind the PSBAT chromophore (i.e., all-trans isomer), the isomerized products are presumably unstable at room temperature and hence would revert back to the original structure. Indeed, the optimized geometries of PSB11 and PSB13 are significantly distorted, with dihedrals $-27$ and $-9^\circ$ twisted out of plane, respectively. These deformations extend to other parts of the chromophore backbone. In fact, from Figure 12C (bottom), it can be seen that the isomerization leads, in both cases, to large skeletal changes in the chromophore, while the segment starting at C9 and comprising the $\beta$-ionone ring seems to undergo a limited displacement. This happens even though the $\beta$-ionone ring sticks out of the protein pocket and would be, in principle, allowed to rotate.

4. CONCLUSIONS

We have constructed tentative QM/MM models for a set of CRABPII-based rhodopsin mimics. It is shown that the presence of many solvent molecules in the protein cavity of the mimics makes the modeling difficult and decreases its accuracy. This suggests that, in spite of the reported improvement upon a limited MD sampling, further effort has to be made to consistently model artificial and natural rhodopsins mainly but not exclusively, due to the need to properly describe the loosely bound solvent molecules in the protein cavity. This may be achieved by more sampling (e.g., longer dynamics with more snapshots) or by improving the quality of the sampling (e.g., sampling a larger part of the system by relaxing cavity side chains).

The $\lambda_{\text{max}}$ and geometries of the chromophore in solution have been found to be similar to those of the chromophore bound in KLE-CRABPII. Also, time-resolved spectroscopy of PSB-KLE-CRABPII reveals different time scales that, similar to what has been documented for PSBAT in solution,123 are assigned to formation and decay (via internal conversion and photoisomerization) of two independent or slowly interconverting $S_1$ species. The similarity of the spectroscopy, geometry, and excited state reactivity of PSBAT in the solvent and in KLE-CRABPII suggests that KLE-CRABPII offers a solvent-like microenvironment, at least more so than in typical rhodopsins like Rh.

The weak stationary fluorescence observed for PSB-KLE-CRABPII could be studied by sampling the $S_1$ potential energy surface via a short excited state QM/MM trajectory. In line with the result of time-resolved spectroscopy, the computations indicate that the observed emission band contains the contribution of two $S_1$ species whose relative stability could not be established. One species (LE) has a delocalized charge and could only be located along a CASPT2 energy profile. The other species (CT) features a charge located close to the $\beta$-ionone ring and shows an inverted BLA pattern. Countering $S_1$ species. The similarity of the spectroscopy, geometry, and excited state reactivity of two independent or slowly interconverting $S_1$ species could only be located along a CASPT2 energy profile. The results revised above support the use of QM/MM models for the in silico screening of spectral and reactivity properties of CRABPII-based rhodopsin mimics as well as of their mutants. It is hoped that further work in this area could lead to valuable tools for the systematic design and production of genetically encodable probes or light-responsive actuators with desired properties. However, this seems to primarily require an improved sampling of the complex HBN which links the internal solvent molecules with the cavity residues. Accordingly, a longer sampling which also involves the cavity side chains is likely to represent the initial direction for future work on this topic.

ASSOCIATED CONTENT

Supporting Information

Methodological details (Figures S1−S4 and Tables S1−S4). Cartesian coordinates of the optimized structures. Energy and oscillator strength values (Figures S5 and S6 and Tables S5−S10). Relaxed scans (Figures S7−S10). Experimental details (Figure S11). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dlarsen@ucdavis.edu (D.S.L.); molivuc@bgsu.edu (M.O.). Phone: 530-754-9075 (D.S.L.); 419-372-7606 (M.O.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are grateful to Dr. Alexander Tarnovsky and Dr. Elena Laricheva for useful comments. This work was partly supported with a Career Development Award (CA0016/2007-C) from the Human Frontiers Science Program Organization (HFSPO) to D.S.L., with NIH grant (R01-GM067311) to B.B. Further support was provided by the Bowling Green State University with RIG FY2010 Grant to M.O. who is also grateful to the Center for Photochemical Sciences and the School of Arts & Sciences of Bowling Green State University for start-up funds, to the National Science Foundation for the CHE-1152070 grant, and to the HFSPO for grant RGP0049/2012. We are grateful to the Ohio Supercomputer Center and NSF-TeraGrid (XSEDE) for granted computer time for multireference calculations. The COST-CMTS Action CM1002 “Convergent Distributed Environment for Computational Spectroscopy (CODECS)” is also acknowledged.

REFERENCES


