Supporting Information: Excited-State Vibronic Dynamics of Bacteriorhodopsin from 2D Electronic Photon Echo Spectroscopy and Multi-Configurational Quantum Chemistry

Samer Gozem,[†,‡,∇] Philip J. M. Johnson,[¶,§,‖,∇] Alexei Halpin,[¶,§] Hoi Ling Luk,[†] Takefumi Morizumi,[⊥] Valentyn I. Prokhorenko,[#] Oliver P. Ernst,[⊥,@] Massimo Olivucci,[*,†,△] and R. J. Dwayne Miller[*,¶,§]

†Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403

‡Current address: Department of Chemistry, Georgia State University, Atlanta, GA 30302.

¶Department of Chemistry, University of Toronto, Toronto, Ontario, Canada.

§Department of Physics, University of Toronto, Toronto, Ontario, Canada.

‖Current address: Laboratory for Nonlinear Optics, Paul Scherrer Institut, CH-5232 Villigen, Switzerland.

⊥Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada

#Max Planck Institute for the Structure and Dynamics of Matter, Atomically Resolved Dynamics Division, Building 99 (CFEL), Luruper Chaussee 149, 22761 Hamburg, Germany

@Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

△Department of Biotechnology, Chemistry and Pharmacology, Universitá di Siena, via De Gasperi 2, I-53100 Siena, Italy

∇These authors contributed equally to this work.

E-mail: molivuc@bgsu.edu; dmiller@lphys.chem.utoronto.ca

Experimental

Pulse compression and characterization

Photon echo spectra were measured using the broadband output of a noncollinear optical parametric amplifier (NOPA) which generated pulses with a bandwidth spanning ∼525–600 nm. Phase distortions arising from dispersion through transmissive optics were accounted for through a combination of prism and deformable mirror compressors. The deformable mirror compressor consisted of a 425 g/mm grating blazed for 560 nm, an $f = 250$ mm spherical mirror, and a 13×3 array membrane-based deformable mirror (OKO Technologies) all in a $4f$ configuration, while the folded prism compressor consisted of two equilateral fused silica prisms separated by ~ 90 cm. A transient grating frequency resolved optical gating (FROG) instrument was used for both *in situ* pulse optimization and characterization. The optimally compressed NOPA pulse (fwhm ~ 11 fs) is shown in Fig. S1.

Phasing

The local oscillator delay (and therefore phasing of the heterodyne-detected photon echo spectra) was determined through the projection-slice theorem,[1] fitting for the local oscillator delay which minimizes the residuals between the pump-probe and the real part of the photon echo response at zero delays. The result of this process is shown in Fig. S2, where the pump-probe response (gray) and the photon echo signal (black) show excellent agreement (1.4% std of residuals) for a local oscillator delay of 422.52 fs.

Photon echo spectra

All real-valued photon echo spectra are shown in Fig. S3 for completeness. The waiting time t_2 is indicated in the top left of each 2D spectrum. Contours are drawn at 5% intervals. High frequency dynamics are plainly observable over the first 100 fs of waiting time, after which the onset of spectral diffusion results in broad, featureless 2D spectra. By $t_2 = 600$ fs,

i	A_i	τ_i, fs	ν_i, cm^{-1}	ϕ_i, rad
below diagonal				
1	0.91 ± 0.1	26 ± 5	–	–
2	0.86 ± 0.09	29 ± 4	945 ± 37	5.4 ± 0.2
above diagonal				
1	0.57 ± 0.03	∞	–	–
2	0.47 ± 0.12	42 ± 14	1634 ± 46	0 ± 0.3
3	1.9 ± 0.2	38 ± 4	936 ± 17	3.6 ± 0.1
4	3.9 ± 1.5	23 ± 5	219 ± 33	3.4 ± 0.1

the photoproduct absorption band is observable at the characteristic difference absorption frequency of $\sim 17\,000$ cm^{-1} (~ 585 nm).

Transient vibronic dynamics

To better quantify the vibronic features from the absorptive 2DPE spectra, the integrated spectral response as a function of the waiting time t_2 (Fig. S4, bottom panel, generated by integration over the indicated detection frequency ranges in the top and middle panels above) are fit to a function of the form

$$I(t_2) = A_1 \mathrm{e}^{-t_2/\tau_1} + \sum_{i=2}^{n} A_i \mathrm{e}^{-t_2/\tau_i} \sin\left(2\pi\nu_i t_2 + \phi_i\right),$$

where the first term corresponds to exponential population dynamics and the summation represents damped coherent nuclear motions. The cross-peak below the diagonal is adequately fit by a single damped sinusoid, while the dynamics above the diagonal require three damped sinusoids. All retrieved fit parameters can be found in the table above.

QM/MM calculations

Computational Methods and bR Model Construction

The dark-adapted bR models were prepared using the available 1.55 Å resolution crystallographic structures (PDB code: 1C3W).[2] As a general criteria and because we are exclusively interested in investigating the bR dynamics in the Franck-Condon region (i.e. during less than 100 fs after photon absorption), we targeted the construction of a QM/MM model capable of reproducing the experimentally observed absorption maximum of bR (568 nm). Membrane lipids at the protein surface were excluded in the model but crystallographic water molecules were maintained. To ensure that the global system charge is neutral, we added a suitable number of chloride ions near positively charged residues (such as Arg, His) on the protein surface and far away from the chromophore. Amino acid ionization states were determined with PROPKA 3.0.[3,4] For the case of the Asp85 and Asp212 residues, PROPKA predicts that they are protonated. However, a model with both residues protonated has a too red-shifted absorption (641 nm) with respect to the experimentally observed bR absorption of 568 nm. Moreover, one (or both) of these residues must serve as the counterion to the retinal protonated Schiff base. PROPKA predicts that Asp212 is more acidic than Asp85, and indeed deprotonation of the Asp212 residue results in a model with an S_0–S_1 energy gap (569 nm) that is in very good agreement with the experimental bR absorption. On the other hand, a model featuring ionized Asp212 and Asp85 residues resulted in a far too blue shifted absorption (455 nm). Therefore, we the model with a deprotonated Asp212 and a protonated Asp85 is used for the dynamics.

The retinal chromophore was treated quantum mechanically using the *ab initio* complete-active-space self consistent field (CASSCF) method.[5] The protein environment is described by the AMBER94 force field[6] with modified parameters for the Lys residue linked to the chromophore.[7,8] Electrostatic embedding was used to describe the interaction between the MM and QM subsystems using the ESPF approach.[9–11] CASSCF is a flexible multiconfigu-

rational method for an unbiased description of the electronic character on both excited and ground state (i.e. with no empirically derived parameters and avoiding single-reference wave-functions). The equilibrium structure was optimized at the single-root CASSCF(12,12)/6-31G*/AMBER level. QM/MM calculations were carried out with Molcas 7.8[12] and Tinker 5.1[13] programs and using the microiterations approach during geometry optimizations.[11] The QM/MM boundary is at the Lys C_{δ}-$C_{epsilon}$ bond and the link-atom scheme was used to treat the frontier between the QM and MM subsystems.[14] The partitioning of QM and MM atoms on our bR model is as follows:

- QM atoms: all retinal atoms, five atoms of the lysine chain connected to it (ϵ-nitrogen, ϵ-carbon and their hydrogen atoms and the link atom. There are 54 QM atoms in total.

- Explicit MM atoms: the remaining 9 atoms of the lysine side-chain are treated explicitly by Molcas (as opposed to other atoms which are treated by Tinker). These atoms are optimized along with the QM part and not by microiterations as the other MM atoms (see below).

- Active MM atoms: these are given by the set of all the side-chains or waters that has at least one atom that is within 4 Åfrom the QM atoms. These MM atoms are optimized using the microiterations approach.

- Inactive MM atoms: all the remaining atoms. These atoms contribute to the electrostatic environment around the chromophore but their geometries are kept frozen during QM/MM optimizations.

The protocol for constructing the final QM/MM model of bR starting from the crystal structure follows these steps:

- Addition of the hydrogens to crystallographic waters and the side chain polar atoms using DOWSER[15]

- Addition of the remaining hydrogen atoms using pdb2gmx command in GROMACS[16]
- MM minimization of all hydrogen atoms, keeping heavier atoms fixed in their original crystallographic positions
- The minimized structure is used to perform a 1 ns molecular dynamics at 298 K on any amino acid side chain featuring at least one atom within 6 Åfrom any retinal chromophore atom. The last snapshot of the molecular dynamics is used to perform a MM minimization on all hydrogen atoms.
- A QM/MM Hartree-Fock/3-21G single point calculation was performed to calculate ESPF charges on QM atoms
- Tinker minimization of ACTIVE MM atoms using the calculated ESPF charges, with the same settings used for hydrogen atoms minimization
- QM/MM Molcas/Tinker HF/3-21G optimization of the system with ACTIVE atoms relaxed
- QM/MM Molcas/Tinker CASSCF(12,12)/3-21G optimization of the system with ACTIVE atoms relaxed
- QM/MM Molcas/Tinker CASSCF(12,12)/6-31G* optimization of the system with ACTIVE atoms relaxed.

This protocol follows the steps The protocol above leads to a ground state equilibrium structure for bR. This structure serves as a starting point for the excited state dynamics on the S_1 state. The S_1 trajectory was computed using the velocity Verlet algorithm[17] with the forces necessary to propagate the Newton's equations of motion obtained from the two-root state-averaged (SA)-CASSCF/6-31G*/AMBER calculations over successive time steps of 1 fs. In order to obtain more quantitative energetics (i.e. excitation energies and excited state energy differences), we carry out single-point six-root CASPT2/6-31G*/AMBER

computations for each step along the bR trajectory.[18] This approach of using CASSCF geometries and trajectories and correcting the energies at the CASPT2 level of theory (the CASPT2//CASSCF approach) is a common compromise that allows calculation of gradients at an affordable level of theory while getting energetics at a quantitatively accurate level of theory.

Calculation of BLA, methyl rock, and π orbital overlap

The bond length alternation (BLA) values plotted in Fig. 4B are computed by taking the average bond length of all formal double bonds and subtracting them from the average bond length of all single bonds along the C5 - N π-conjugated backbone. The methyl rock value (see Fig. S5) is determined by taking the difference between the Me-C13=C14-H dihedral (β) and C12-C13=C14-C15 dihedral (α). Finally, the π-orbital overlap (τ) is defined as the average of α and β ($\alpha/2 + \beta/2$). Changes in the methyl rock and τ along the rPSBAT trajectory are shown in Fig. S5. Fig. S6 supplements Fig. 4 of the main manuscript by providing additional energy gaps and oscillator strength data along the trajectory. Finally, Fig. S7 provides structural information about the rPSBAT chromophore at 0 fs, 5 fs, and 12 fs of the trajectory.

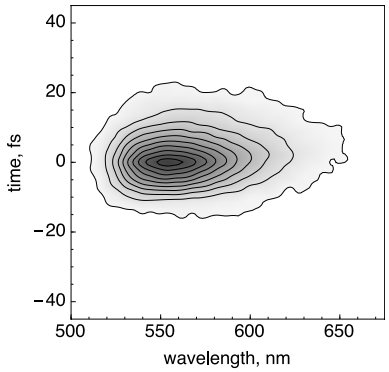

Figure S1: Measured transient grating FROG signal of the transform-limited pulses used for the collection of photon echo spectra of bacteriorhodopsin. The temporal FWHM is $\sim$ 11 fs. Contours are drawn at 10% intervals.

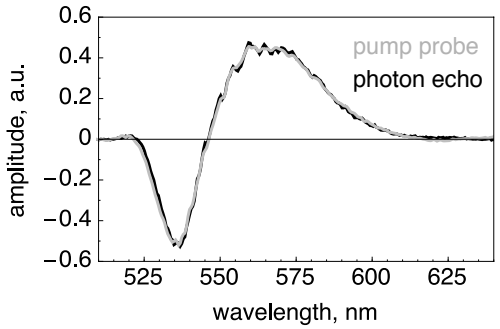

Figure S2: Projection-slice analysis of the pump-probe (grey) and photon echo (black) response of bacteriorhodopsin at zero delays to determine the temporal local oscillator offset and therefor the real valued response of the heterodyne-detected photon echo signals. The LO was determined to be offset by 422.52 fs.

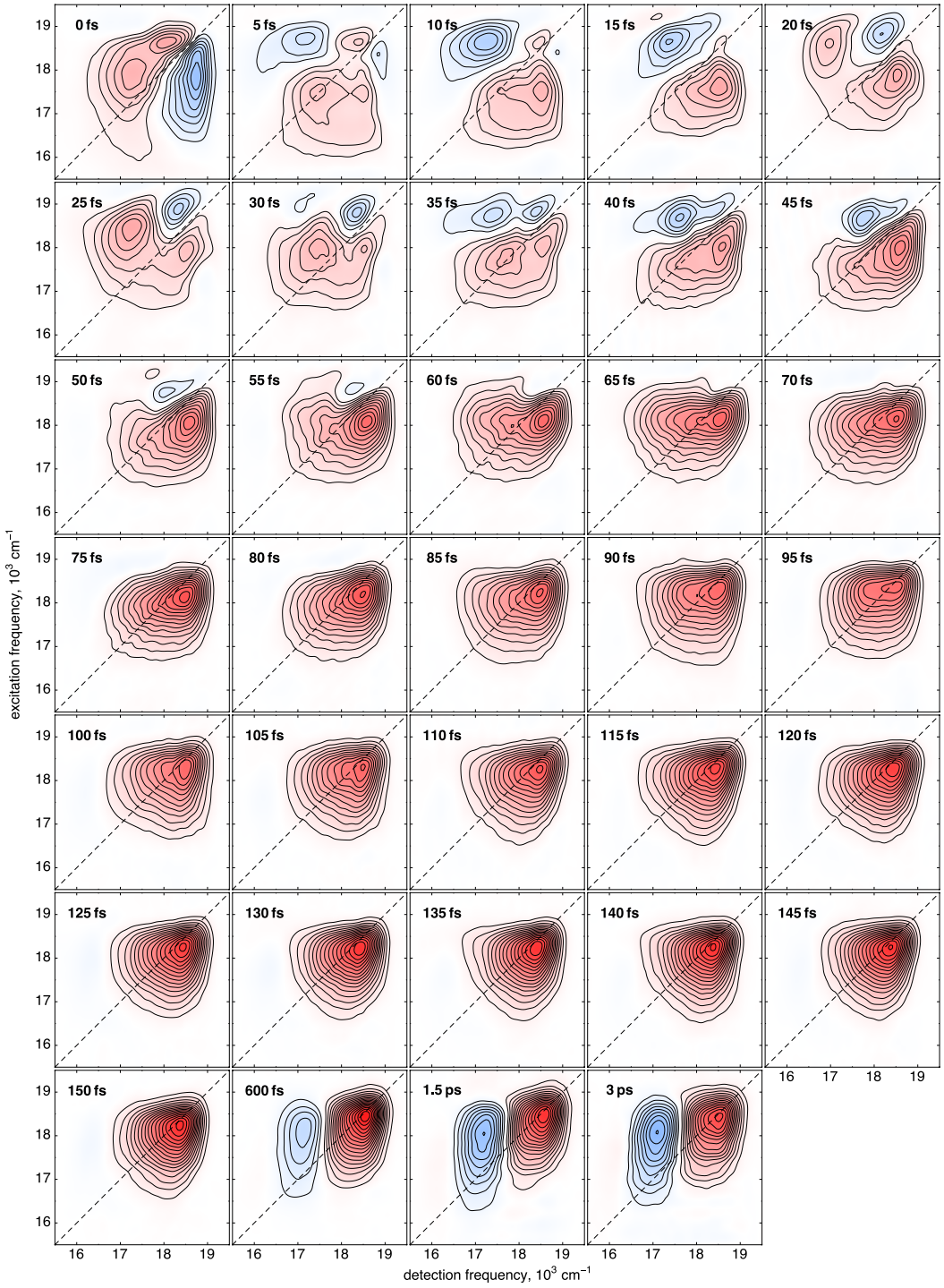

Figure S3: Real-valued 2D photon echo spectra of retinal isomerization in bacteriorhodopsin. The waiting time t_2 is indicated in the top left of each 2D spectrum, and contours are drawn at 5% intervals.

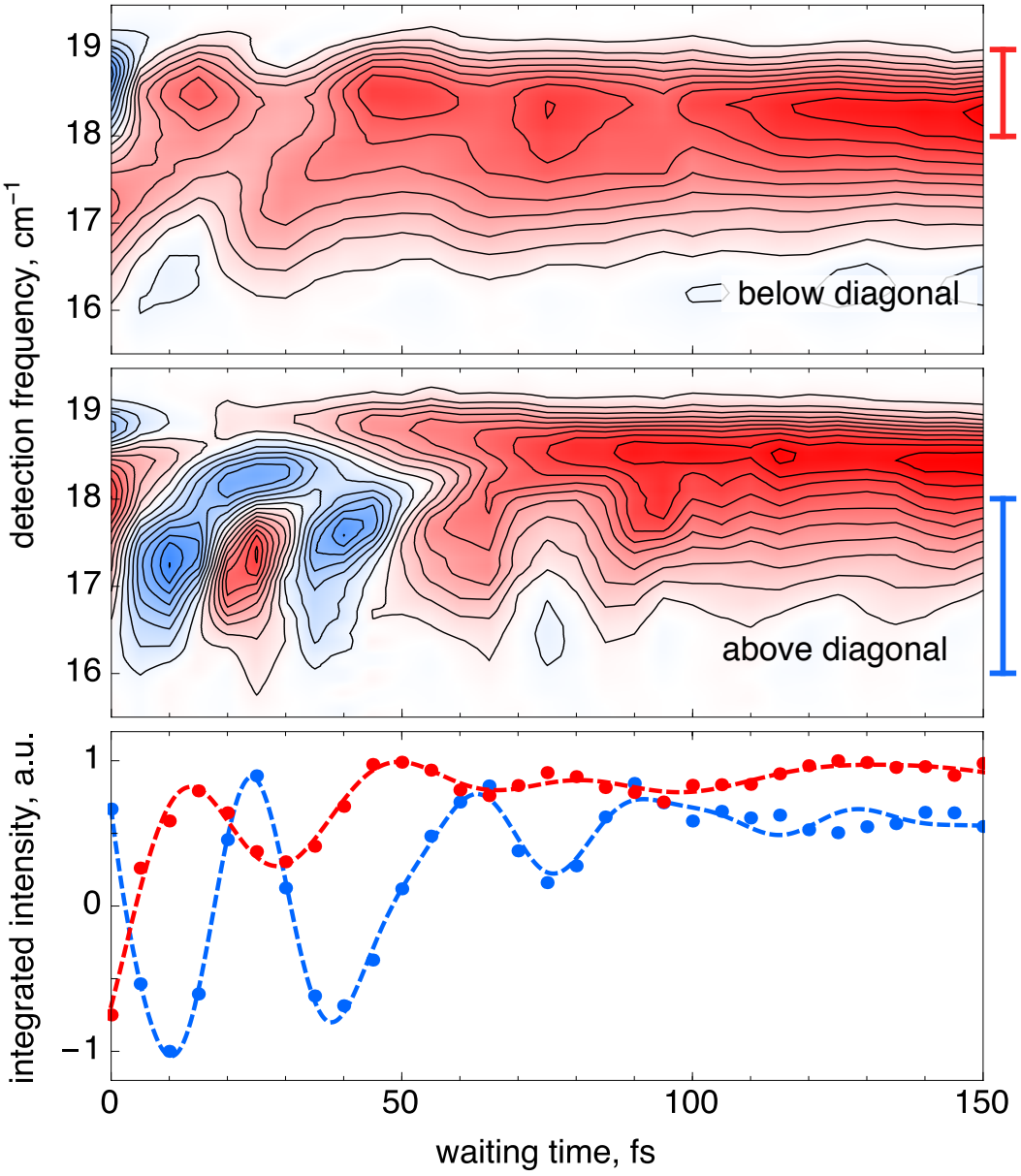

Figure S4: Pump-frequency-resolved spectral dynamics. (Top) Spectral response over excitation frequencies corresponding to the absorption maximum at $17,500 \pm 500\,\mathrm{cm}^{-1}$, highlighting the vibronic cross peak at a detection frequency of $18,500\ \mathrm{cm}^{-1}$. (Middle) Dynamics due to excitation of the high frequency vibronic shoulder at $18,500 \pm 500\,\mathrm{cm}^{-1}$, where broad, large amplitude oscillatory dynamics are resolved for $\sim$ 100 fs following photoexcitation. (Bottom) Spectrally-integrated dynamics of the vibronic cross peak (red data) and chromophore dynamics associated with excitation of this vibronic feature (blue data) over the ranges identified above by the corresponding bars along the right axes of the top and middle panels, with associated fits shown as dashed lines.

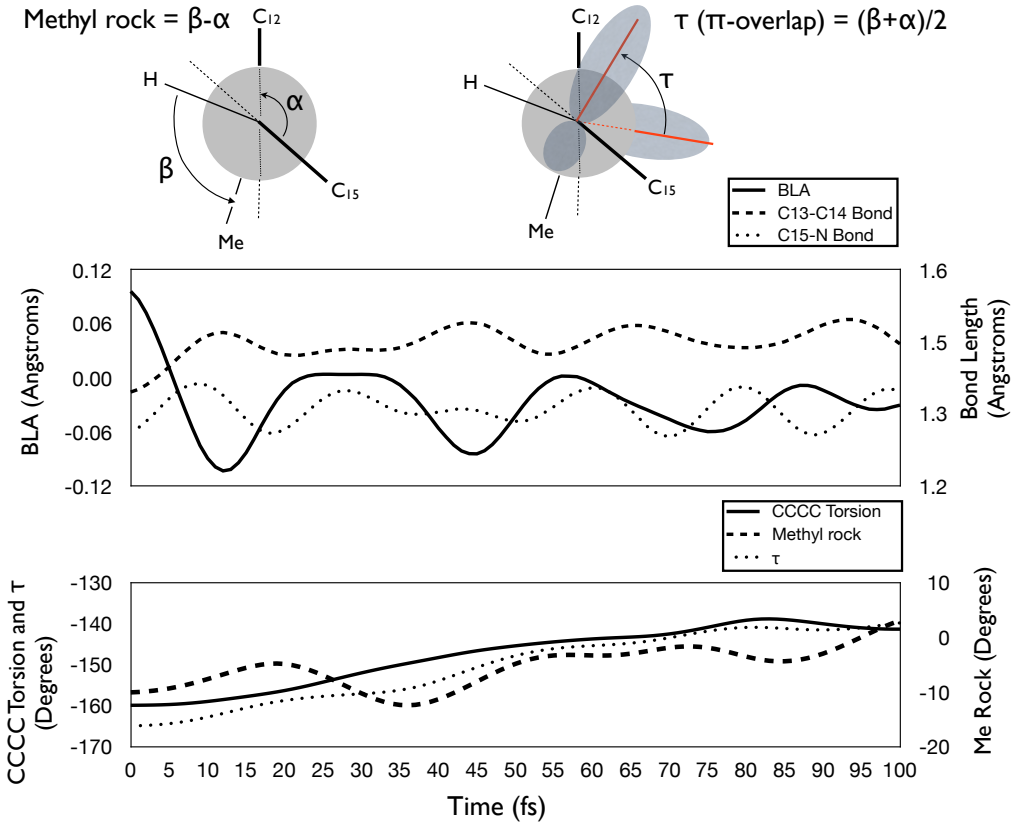

Figure S5: Top: Definition of the methyl rock (Me Rock) and π-orbital overlap (τ) parameters used in this work. Methyl rock is defined as $\beta - \alpha$, while τ is defined as $(\alpha + \beta)/2$. Center: Changes in the BLA coordinate (see SI text or main manuscript for definition of BLA) and in the C13–C14 and C15–N bond lengths along the bR S_1 trajectory. Bottom: Changes in the C12–C13–C14–C15 dihedral angle and in the methyl rock and τ coordinates along the same trajectory.

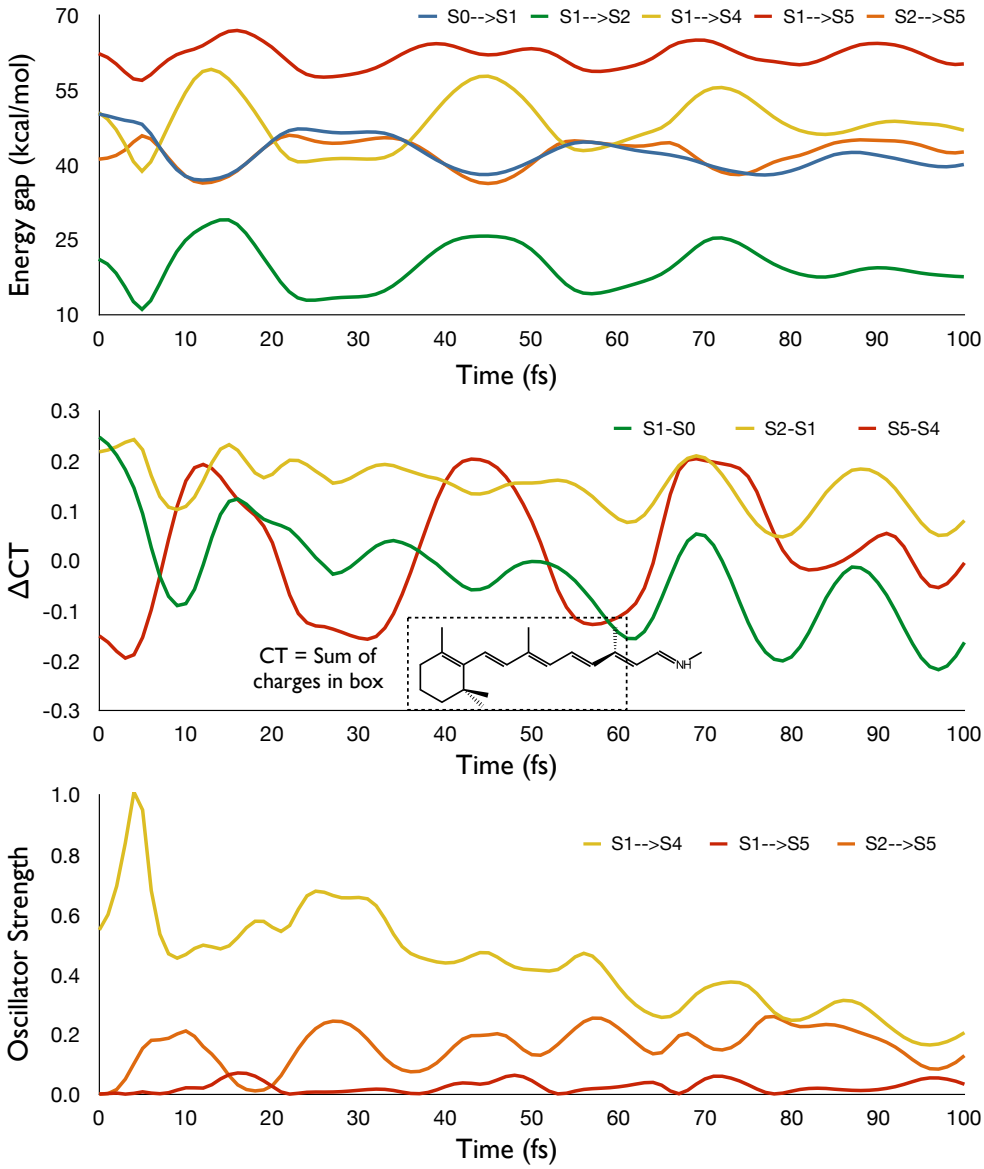

Figure S6: Top: Change in the S_0–S_1 (blue), S_1–S_4 (green), S_1–S_5 (yellow), and S_2–S_5 (red) energy gaps along the bR S_1 trajectory, computed at the 6-root CASPT2//CASSCF level of theory. Middle: Difference in the charge transfer (CT) character between S_1 and S_0 (green), S_2 and S_1 (yellow), and S_5 and S_4 (red) along the bR S_1 trajectory, computed at the 6-root SA-CASSCF level of theory. Charge transfer character is defined as the sum of the charges of atoms inside the boxed part of the structure. Bottom: Change in the S_1–S_4 (green), S_1–S_5 (yellow), and S_2–S_5 (red) oscillator strength along the same trajectory, computed at the same level of theory. Note that the S_1–S_5 transition is always forbidden or weak, and that S_1–S_4 is always allowed. However, S_2–S_5 is allowed when its energy nears the experimental detection window (ca. 46 kcal/mol and above).

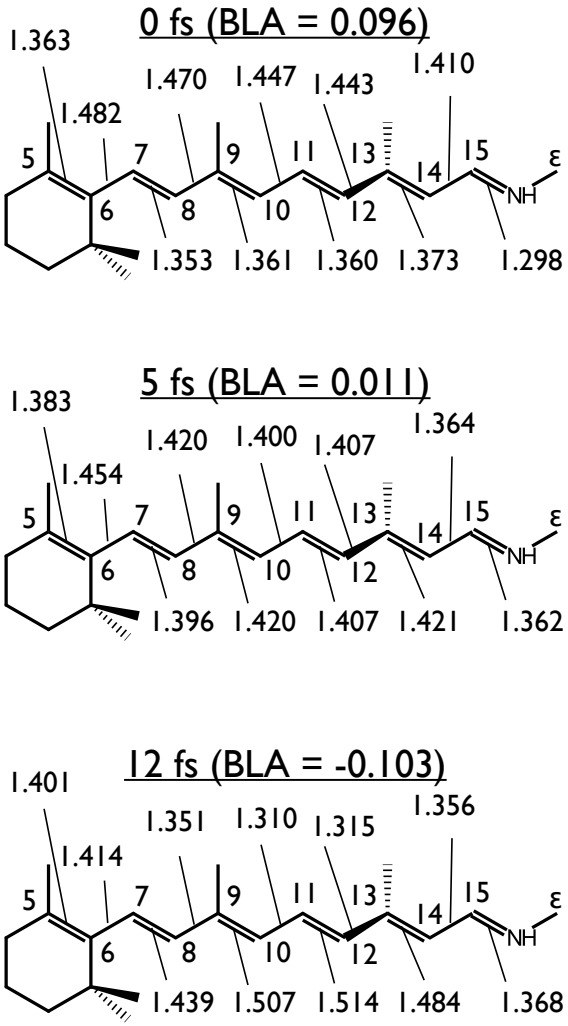

Figure S7: Bond lengths in Å of the bR chromophore at 0 fs (Franck-Condon point), 5 fs (when the S_1–S_2 energy gap first reaches a minimum), and at 12 fs (when the BLA value first reaches a minimum and the S_1–S_2 energy gap reaches a maximum. Note the inverted bond length alternation between the top structure (long formal single bonds and short formal double bonds) and the bottom structure (short formal single bonds and long formal double bonds). The S_2 state is the most stable when the single and double bond lengths are approximately equal (i.e., when BLA=0)

References

(1) Hybl, J. D.; Albrecht, F. A.; Jonas, D. M. Two-dimensional Fourier transform electronic spectroscopy. *J. Chem. Phys.* **2001**, *115*, 6606–6622.

(2) Luecke, H.; Schobert, B.; Richter, H. T.; Cartailler, J. P.; Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 Åresolution. *J. Mol. Biol.* **1999**, *291*, 899–911.

(3) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 704–721.

(4) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

(5) Roos, B. O.; Bruna, P.; Peyerimhoff, S. D.; Shepard, R.; Cooper, D. L.; Gerratt, J.; Raimondi, M. Ab Initio Methods in Quantum Chemistry, II. *Adv. Chem. Phys.* **1987**, *69*, 399–446.

(6) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(7) Ferré, N.; Olivucci, M. Probing the rhodopsin cavity with reduced retinal models at the CASPT2//CASSCF/AMBER level of theory. *J. Am. Chem. Soc.* **2003**, *125*, 6868–9.

(8) Ferré, N.; Cembran, A.; Garavelli, M.; Olivucci, M. Complete-active-space self-consistent-field/Amber parameterization of the Lys296–retinal–Glu113 rhodopsin chromophore-counterion system. *Theor. Chem. Acc.* **2004**, *112*, 335–341.

(9) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

(10) Ferré, N.; Ángyán, J. G. Approximate electrostatic interaction operator for QM/MM calculations. *Chem. Phys. Lett.* **2002**, *356*, 331–339.

(11) Melaccio, F.; Olivucci, M.; Lindh, R.; Ferré, N. Unique QM/MM potential energy surface exploration using microiterations. *Int. J. Quant. Chem.* **2011**, *111*, 3339–3346.

(12) Aquilante, F.; Vico, L. D.; Ferré, N.; Ghigo, G.; Malmqvist, P. A.; Neogrády, P.; Pedersen, T. B.; Pitonák, M.; Reiher, M.; Roos, B. O. et al. MOLCAS 7: the next generation. *J. Comput. Chem.* **2010**, *31*, 224–47.

(13) Ponder, J. W.; Richards, F. M. An efficient newton-like method for molecular mechanics energy minimization of large molecules. *J. Comp. Chem.* **1987**, *8*, 1016–1024.

(14) Singh, U. C.; Kollman, P. A. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH3Cl+ Cl- exchange reaction and gas phase protonation of polyethers. *J. Comp. Chem.* **1986**, *7*, 718–730.

(15) Hermans, J.; Xia, X.; Zhang, L.; Cavanaugh, D. Dowser program. 2014; `http://danger.med.unc.edu/hermans/dowser/dowser.htm`.

(16) Hess, B.; Kutzner, C.; Spoel, D. V. D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(17) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **1982**, *76*, 637.

(18) Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-order perturbation theory with a CASSCF reference function. *J. Phys. Chem.* **1990**, *94*, 5483–5488.