Modeling pH-dependent biomolecular photochemistry

Elisa Pieri,[*,†,‡] Oliver Weingart,[¶] Miquel Huix-Rotllant,[†] Vincent Ledentu,[†] Marco Garavelli,[§] and Nicolas Ferré[*,†]

†*Aix-Marseille Univ, CNRS, Institut de Chimie Radicalaire, Marseille, France*

‡*Present address: Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States*

¶*Institut für Theoretische Chemie und Computerchemie, Heinrich Heine Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany*

§*Dipartimento di Chimica Industriale "Toso Montanari", Università degli Studi di Bologna, Viale del Risorgimento, 4, 40136 Bologna, Italy*

E-mail: elipieri@unc.edu; nicolas.ferre@univ-amu.fr

Abstract

The tuning mechanism of pH can be extremely challenging to model computationally in complex biological systems, especially with respect to photochemical properties. This article reports a protocol aimed at modeling pH-dependent photodynamics, using a combination of constant-pH molecular dynamics and semi-classical nonadiabatic molecular dynamics simulations. With retinal photoisomerization in Anabaena Sensory Rhodopsin (ASR) as a testbed, we show that our protocol produces pH-dependent photochemical properties such as the isomerization quantum yield or decay rates. We decompose our results in single titrated residue contributions, identifying some key tuning amino acids. Additionally, we assess the validity of the single protonation state

picture to represent the system at a given pH and propose the most populated protein charge state as a compromise between cost and accuracy.

1 Introduction

The pH is one of the main intensive properties that can influence the outcome of chemical processes in biomolecules, e.g., in terms of rate and yield.[1,2] While a famous mechanism of action is proton transfer and related acid-base equilibria, pH changes can also impact transformations in a more indirect fashion. For instance, the photoisomerization in rhodopsin proteins is pH-dependent, even if the protonation state of the retinal chromophore does not change under non-extreme pH conditions.[3–6] In fact, changing the macroscopic pH value translates microscopically into potentially modifying the protonation state, and hence the electric charge, of each titratable group in the (macro)molecule. In photoactive proteins, this modifies the electrostatic potential acting on the chromophore and eventually tunes the energy gap between the ground and some excited electronic state, especially when charge transfer states are involved.

The modeling of the pH dependence of excited state properties in photoactive proteins faces several challenges. Besides the usual question regarding the electronic structure method required to reach a good accuracy for excitation energies, excited electronic state characters, and potential energy surface topologies,[7–9] the most challenging issue involves statistics. In fact, not only does the conformational space of the system need to be sampled extensively through, for instance, classical molecular dynamics, but the ensemble of protonation microstates also requires proper sampling.[10] When the number of titrated residues is large, as is usually the case with medium and large-size proteins, the latter sampling becomes incredibly demanding since the ensemble size scales exponentially with the number of titratable residues. However, successfully modeling the impact of pH on protein photochemical properties would provide a comprehensive atomistic-level understanding of the pH tuning

mechanisms and crucial information to develop or enhance control of protein activity and rationally design improved mutants.

Some pH-dependent photochemical properties have been successfully modeled and studied *in silico*. One example is the pH-dependence of absorption spectra in polypeptides and proteins.[11–15] A convenient approach consists of modeling the titration of only the residues whose change of protonation state modifies the excitation energy significantly, which could be identified using a minimal electrostatic model.[12] Alternatively, constant-pH molecular dynamics (CpHMD) simulations can be used to sample simultaneously the conformational and protonation state spaces.[16,17] This yields thousands of snapshots that can be used for hybrid quantum mechanics/molecular mechanics (QM/MM) calculations of excitation energies to convolve absorption spectra. This workflow (called CpHMD-then-QM/MM)[13] has been successfully applied to the pH-dependent absorption spectrum of ASR.[14] For this protein, Pieri et al. reproduced the pH-induced variation of the maximum absorption wavelength $\lambda_{\max}$, and analyzed its molecular origin in terms of titrated amino acids. The robustness of such an approach has also been demonstrated in the case of a titratable chromophore, as it is the case for the oxyluciferin luminophore involved in the bioluminescence phenomena.[18,19]

While it is possible to obtain theoretical absorption spectra with the tools mentioned above, the pH-dependent quantum yield and excited state lifetime of photo-active proteins have never (to our knowledge) been accurately modeled before. In fact, typically a pH value is modeled by choosing (and fixing) the protonation state of each titratable amino acid in the protein; in other words, only one protonation microstate is assumed to be representative of the system at a given pH value, which can be a gross simplification. In this work, we want to assess whether the methodology we developed[13] and successfully applied to study pH-dependent $\lambda_{\max}$ variations,[14,18,19] which considers large numbers of microstates, can be extended to study the pH-dependent photochemistry of photo-active biomolecules. Therefore, using Anabaena Sensory Rhodopsin (ASR) as a testbed, we hereafter present an enhanced CpHMD-then-QM/MM nonadiabatic molecular dynamics protocol meant to in-

vestigate the effects of changing pH on its excited state lifetime and isomerization quantum yield. This new computational protocol is specially designed to address the following main questions:

1. When is a single protonation microstate sufficient to reproduce photochemical properties at a given pH value? When it is not sufficient, how the conventional models could be modified?

2. How much can these properties vary when the pH changes?

3. Is it possible to rationalize the molecular origin of such effects?

In a nutshell, thousands of carefully selected snapshots extracted from CpHMD trajectories are used as initial conditions for a statistically relevant ensemble of QM/MM semi-classical non-adiabatic MD simulations. We emphasize that our main goal is not to produce accurate quantum yields or excited state lifetimes, but rather to design a methodology suitable for qualitatively investigating how pH modifies photochemical properties of large (bio-) molecular systems. In the next section, we present the selected biomolecular model together with its titratable amino-acids, as well as the main strategies and tools we take advantage of. All the technical details, necessary for ensuring the reproducibility of our results, are given as Supporting Information.

2 Methods and Computational Details

The protein used to develop and test our protocol is Anabaena Sensory Rhodopsin (ASR),[20] a photoactive transmembrane protein in which the retinal chromophore can exist in two conformations (Figure 1): all-*trans* (AT) and 13-*cis* (13C).[21]

AT is the only conformation found in the dark-adapted ASR. Upon light absorption, all-*trans* retinal can photoisomerize to 13-*cis*. Therefore, light-adapted ASR exists as a mixture of AT and 13C conformations, the respective concentrations of which depend on

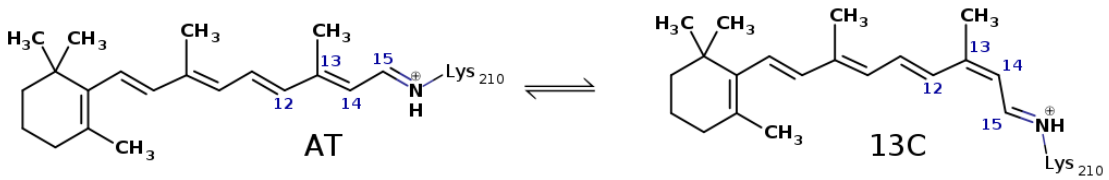

Figure 1: AT and 13C conformations of the retinal chromophore in ASR. Retinal is bound to the opsin through a protonated Schiff base with the Lysine 210 residue. The photoinduced isomerization involves the dihedral angle defined by carbon atoms 12, 13, 14, and 15.

the pH. As a consequence, the ASR photophysical and photochemical properties have been measured to be pH-dependent.[22,23] ASR absorption maximum wavelength ($\lambda_{\max}$) features a small but sizable (∼2 nm) red-shift between pH=3 and pH=5, then a larger (6 nm or more, depending on the isomer) blue-shift between pH=5 and basic pH. Two main properties are often used to characterize the ASR photochemical behavior:[24–26] its excited state lifetime and the retinal isomerization quantum yield.[27] At pH=7, the latter one is significantly larger for the AT→13C (0.38) than for the reverse transformation (0.24). However, as of today, available experimental information regarding the pH dependence of such properties in ASR remains scarce.

A macromolecule with n titratable sites has at least 2^n available protonation microstates, defined as combinations of individual protonation states; this number can be even larger in proteins since histidine residues feature three possible protonation states. Even if not all of the microstates are populated at a given pH value, the number of significantly populated microstates can be extremely high under non-extreme pH conditions.[14] Moreover, often the population is evenly distributed in the microstates, making it almost impossible to select a handful of microstates capable of accurately representing the system.

One way to circumvent this issue in simulations is reducing as much as possible the number of titrated sites. Our previous study indicates that, even though many minor contributions are present, only a few amino acids are capable of tuning the absorption maximum of ASR with their protonation state change.[14] For this reason, we decided to titrate only four residues at pH=3 (D57, E62, D98, and D120) and four residues at pH=7 (H21, E36,

D217, and H219). The pH=5 case represents a peculiarity since it stands exactly at the frontier between the pH range where the glutamic and aspartic acid residues typically finish their titration process and the one where histidine residues start theirs. For this reason, we decided to titrate a larger number of amino acids, for a total of twelve: E4, H8, D57, E62, H69, D98, E123, D125, E160, D217, H219, and D226. The location of the titrated residues is represented in Figure 2.

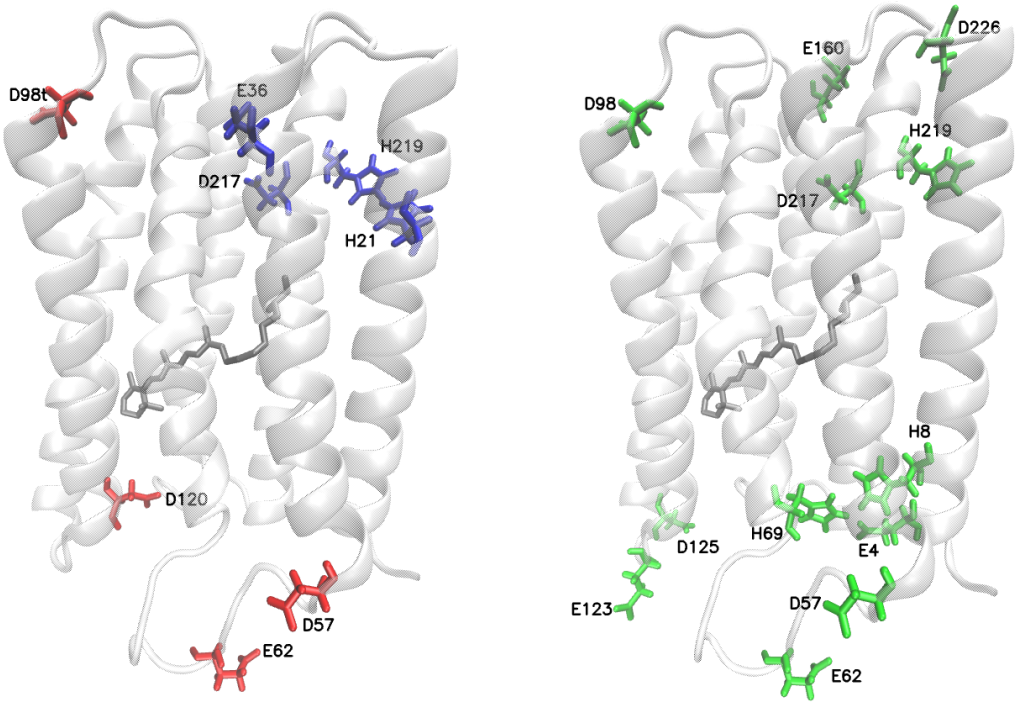

Figure 2: Location of the titrated residues in ASR at pH=3.0-4.5 (left, in red), pH=4.5-6.0 (right; in green) and pH=6.0-7.5 (left, in blue). The retinal+K210 moiety is represented in grey.

Our first step to get a reasonable sampling of the phase and protonation state spaces is CpHMD. Briefly, CpHMD is a molecular dynamics method in which the protonation state of some titratable amino-acids can be periodically changed, using a Metropolis decision based on the free energy of (de)protonation. The Replica-Exchange extension of CpHMD (pH-REMD) takes advantage of the embarrassingly parallel trajectories (one per pH value) to improve the convergence of the protonation microstate space sampling. This is achieved by periodically exchange information between "neighbor" trajectories.[28] After the system

preparation, heating, and equilibration (see Supporting Information), we calculated four 20 ns long CpHMD trajectories with pH-REMD (using a 0.5 pH step) in three distinct pH ranges, hereafter called windows, (3.0-4.5, 4.5-6.0, 6.0-7.5) titrating the aforementioned residues and keeping the protonation state of the remaining titratable sites fixed. The fixed protonation states were decided using the results presented in our previous article in order to reflect the most probable situation[14] (see Supporting Information). This step yielded 20000 snapshots per trajectory and provided us with statistical information about the population of the available microstates at each pH value.

To select initial conditions for the nonadiabatic dynamics, we used data from our previous work on ASR where we modeled pH-dependent absorption spectra.[14] In said work, we extracted 20,000 snapshots (intended as a geometry and a distribution of charges representing the corresponding protonation microstate) from each CpHMD simulation at pH=3, pH=5 and pH=7 and for both isomers, and computed the excitation energy using PM7 to model the electronic structure of the retinal chromophore. Therefore, for each isomer and pH value, we have a full dataset D consisting of 20,000 excitation energies. The initial condition selection procedure was conducted as follows:

1. We computed the average $\bar{\lambda}^{\mathrm{D}}_{\mathrm{max}}$ and standard deviation σ^{D} for the excitation energies in D.

2. We created a smaller dataset d containing 1,000 snapshots for each D. The selection of d was different depending on the pH value. At pH=3 and pH=7, we randomly chose a certain number of snapshot n_i within each populated microstate i. n_i is proportional to the population N_i of microstate i:

$$n_i = \frac{N_i * 1,000}{20,000} = \frac{N_i}{20}$$

where n_i is rounded to the closest integer. This procedure allowed us to maintain in d the same microstate distribution as in D. At pH=5, each microstate is represented

by a number of snapshots between 0 and ∼20: such a spread out population makes it impossible to use the same selection scheme mentioned above. Therefore, at this pH value snapshots were selected randomly and regardless of their microstate of belonging.

3. We computed $\bar{\lambda}^{\mathrm{d}}_{\mathrm{max}}$ and σ^{d} and compared these values to $\bar{\lambda}^{\mathrm{D}}_{\mathrm{max}}$ and σ^{D}. In all cases, d and D proved to have an extremely similar distribution of excitation energies, since $\left|\bar{\lambda}^{\mathrm{D}}_{\mathrm{max}} - \bar{\lambda}^{\mathrm{d}}_{\mathrm{max}}\right| < 0.01\,\mathrm{eV}$ and $|\sigma^{\mathrm{D}} - \sigma^{\mathrm{d}}| < 0.01\,\mathrm{eV}$, ultimately verifying that our smaller datasets are a good representation of the larger datasets.

Using the selected 1000 structures per pH value as initial conditions, we perform excited state semi-classical QM/MM-MD simulations using COBRAMM 2.0[29] with the retinal chromophore as the QM part and the remaining protein and lipid layer as MM. It is worth mentioning that pH is essentially maintained since the protonation state of each residue is fixed when ASR is promoted to its excited state. This restriction is essentially motivated by the proton transfer time scale, much slower than the retinal isomerization one and by the absence of retinal deprotonation in the considered pH range. Retinal itself and residues within 5 Å of the retinal chromophore were movable during our photodynamics simulations, while the rest of the system was frozen. The initial velocities were set to zero (see Supporting Information for details). The trajectories were propagated for up to 2.5 ps on QM/MM potential energy surfaces, starting on S_1 and allowing hops to S_0 and S_2 using Tully's fewest switches surface hopping algorithm.[30] S_2 was included to describe the strong state mixing already demonstrated in the case of ASR.[31,32] In fact, describing S_1 as a charge-transfer excited state and S_2 as a diradical state in the Franck-Condon region, Manathunga et al.'s CASPT2//CASSCF semi-classical MD simulations highlighted their proximity during the early part of the photo-induced retinal isomerization,[32] leading to state mixing between S_1 and S_2. The chosen QM level of theory scheme is OM3/MRCI,[33,34] since it has been shown to provide good excited state properties[35] and has been already used to study rhodopsins yielding reasonably good results with an acceptable compromise between accuracy and computational effort;[36,37] this choice is motivated by the extremely large amount of trajectories

(6000) we needed to calculate. The QM region is limited to the ASR retinal chromophore; the MM interactions are modeled using the Amber forcefield[38] and the QM/MM ones with an electrostatic embedding approach, i.e., including MM point charge electrostatic potential in the semi-empirical Hamiltonian, as implemented in COBRAMM.[29]

The resulting data sets are labeled as pH3-AT, pH3-13C, pH5-AT, pH5-13C, pH7-AT, and pH7-13C. Notice that about 80% of the trajectories have been used for later analysis, the rest is discarded for reasons explained in Supporting Information, section 1. We analyzed each ensemble of trajectories using three different statistical properties. The first one, the isomerization quantum yield (IQY), characterizes the efficiency of the retinal photoisomerization. In principle, this quantity can be directly compared with the experiment, if the number of trajectories is large enough (i.e., statistically converged). The two other properties are related to the excited state lifetime: the average hopping time between S_1 and S_0 and the time constants fitted to reproduce the $S_1 \rightarrow S_0$ population decay time evolution. The former is only defined theoretically (in the framework of surface-hopping MD simulations) and it does not provide information about the future evolution of the trajectory (i.e. if the isomerization will succeed or get aborted). The time constants, based on mechanistic and kinetic assumptions we will present later, can, in principle, be compared with experimental equivalent ones. Moreover, each of them characterizes a particular decay channel.

3 Results

3.1 Retinal isomerization mechanism

The canonical retinal isomerization mechanism in microbial rhodopsins involves the highly selective AT↔13C reversible photoisomerization. After visible light absorption, the retinal structure changes: its bond length alternation (BLA, difference between the double bond lengths and the single bond lengths, from C_6 to C_δ) pattern is inverted, eventually triggering a complex space-conserving rotation around the C_{13}−C_{14} bond, happening on a sub-ps

time scale. Retinal can also in some cases twist around other bonds, but these usually lead to aborted isomerizations due to steric hindrance, confirming that the cavity shape enforces selectivity for the dihedral involved in the photoisomerization.[6,39–41] Since a pH change translates into modified electrostatic interactions between retinal and its surroundings amino acids, we may first question the validity of the above-mentioned isomerization mechanism between pH=3 and pH=7. Our simulations show this isomerization mechanism is conserved across pH values and independently of the initial retinal configuration (AT, figure 3 or 13C, figure 4 in Supporting Information).

The BLA values are essentially positive at $t = 0$, with an average value close to 0.02 Å (see figure 3 for the case of the AT→13C retinal isomerization and Supporting Information for the opposite case). The BLA quickly becomes negative (about -0.02 Å) after a few tens of fs. Then, it rapidly goes back to its initial value (after about 300 fs), thanks to the increasing number of $S_1 \rightarrow S_0$ hops. Since trajectories are stopped rapidly after having decayed to S_0, the BLA value at a large time scale cannot reflect the necessary bond length oscillations. Nevertheless, it should be emphasized that the reported BLA variation takes place in less than 200 fs, i.e. it is ultrafast.

In the AT→13C case, the initial $C_{13}-C_{14}$ torsion angles are distributed between $-150°$ and $-190°$. After about 200 fs, two branches become apparent. In one, the torsion angles remain in the same range as the initial values, indicating unsuccessful (e.g., aborted torsions around other bonds) isomerizations. The other branch features torsion angles rapidly evolving towards $-90°$, corresponding to successful isomerizations. The dihedral and BLA time evolution is perfectly in line with the ones reported for rhodopsins.[42] This conclusion is also valid for the 13C→AT case (figure 4 in Supporting Information). Accordingly, changing the pH does not alter the usual retinal isomerization mechanism. It is worth noting that successful isomerizations occur more often at long times when the starting retinal isomer is AT.

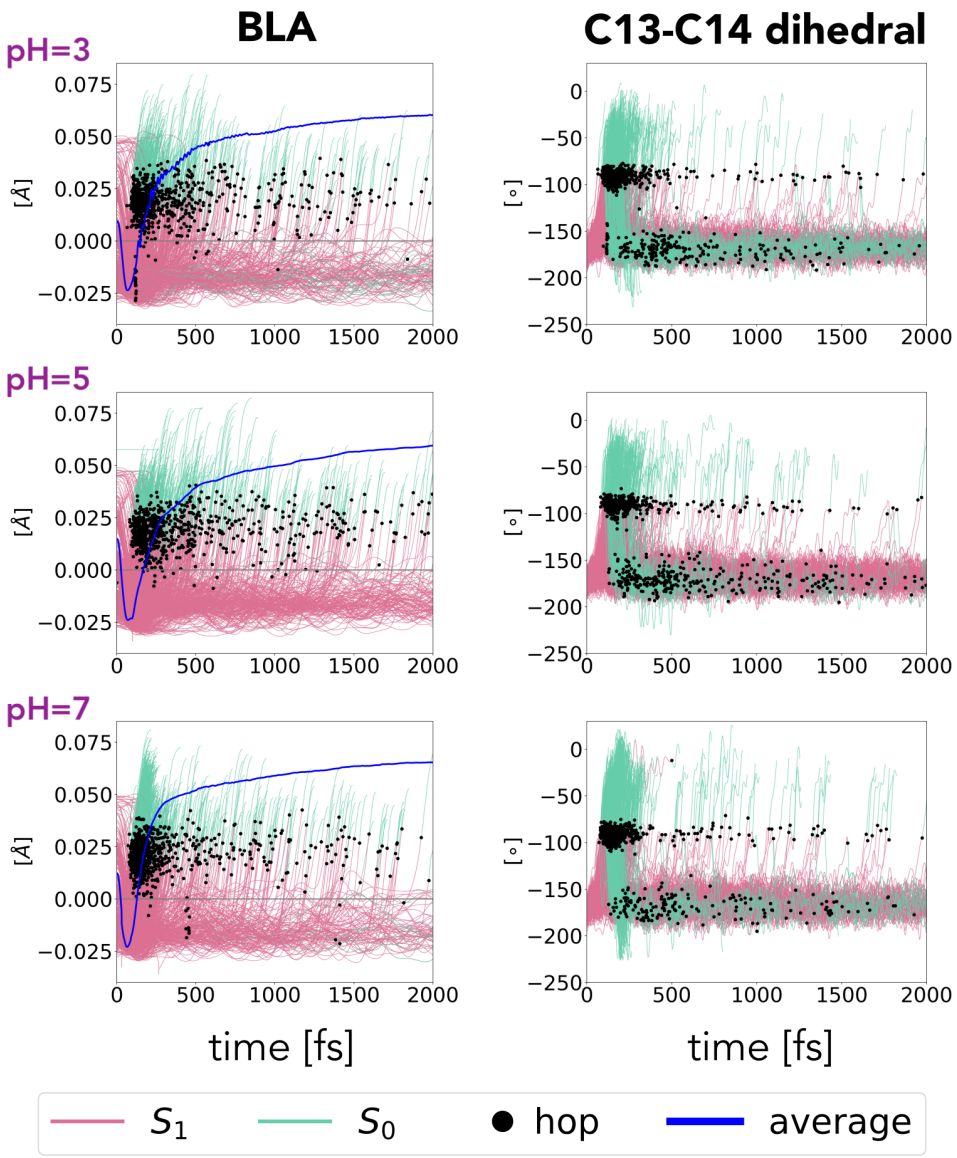

Figure 3: Time evolution of BLA (left) and torsion dihedral angle (right) during the AT→13C retinal isomerization. S_1 (pink) and S_0 (green) parts of a trajectory are separated by a hop point (black circle). The BLA instantaneous average values are also plotted in blue. Please note that, since the trajectories are stopped shortly after reaching S_0, their last BLA value is frozen for the remainder of the averaging to avoid noise and discontinuities.

3.2 Isomerization quantum yield

Having found that the isomerization mechanism is not affected by the pH, we now investigate how pH modifies the IQY, the quantity that characterizes the efficiency of the retinal photo-isomerization. In rhodopsins, this value typically ranges between a few percent and

almost 0.7 in bovine rhodopsin or bacteriorhodopsin.[6] The ASR IQY is still debated: it is estimated to be either 0.17 (0.06) or 0.38 (0.24) for AT→13C (13C→AT) at neutral pH.[25,27] However, the IQY is known to be pH-dependent in several related systems, like bacteriorhodopsin.[4] Accordingly, we hereafter investigate how the pH can influence the ASR IQY, using the CpHMD-then-QM/MM protocol.[13] Computationally, the IQY can be calculated as the number of reactive trajectories, i.e., the ones featuring a completed isomerization around the C_{13}=C_{14} bond, divided by the number of trajectories for each model. We report this quantity for each model in Table 1.

Table 1: Isomerization quantum yields (IQY) calculated using the *Full* set of trajectories, the *Direct* (i.e., trajectories which hop directly from S_1 to S_0) subset and the *Indirect* (i.e., trajectories which visit S_2 before going back to S_1 and then S_0) subset. Uncertainties are calculated as $\sqrt{\text{IQY}(1-\text{IQY})/n}$ with n the total number of valid trajectories.[43]

	AT→13C			**13C→AT**		
pH	3	5	7	3	5	7
Full	0.43 ± 0.02	0.39 ± 0.02	0.56 ± 0.02	0.33 ± 0.02	0.35 ± 0.02	0.36 ± 0.02
Indirect	0.46 ± 0.05	0.46 ± 0.04	0.63 ± 0.04	0.39 ± 0.04	0.43 ± 0.04	0.50 ± 0.04
Direct	0.43 ± 0.02	0.38 ± 0.02	0.54 ± 0.02	0.31 ± 0.02	0.34 ± 0.02	0.33 ± 0.02
Dir:Indir ratio	637:115 (5.5)	640:124 (5.2)	615:172 (3.6)	670:140 (4.8)	632:136 (4.6)	614:126 (4.9)

The 13C→AT IQY is always smaller than the AT→13C one, independent of the pH value. This result is in line with experimental values at pH=7 (0.24 ± 0.03 and 0.36 ± 0.06, respectively).[27] However, the computed IQY values at pH=7 are much larger than the experimental ones, probably evidencing the relatively low quality of the semi-empirical Hamiltonian-based potential energy surface. Nevertheless, the level of theory chosen as a compromise allows us to compare properties at different pH values and extract trends. Indeed, while the 13C→AT IQY seems pH-independent, the AT→13C one dramatically increases when the pH becomes neutral.

The origin of pH-dependent IQY variations can be traced back by considering the impact of the S_1/S_2 mixing.[31,32] For this purpose, we split each ensemble of trajectories in (i) a direct subset that contains trajectories that hop directly from S_1 to S_0; (ii) an indirect subset for the trajectories that hop from S_1 to S_2 before hopping back to S_1 and eventually to S_0 (no

direct hop between S_2 and S_0 has been observed). Each subset can be characterized by its respective IQYs, as reported in Table 1.

The indirect IQY is larger than the direct one at all pH values and retinal isomers. This result suggests that S_2 assists the nuclear wavepacket to prepare for a successful isomerization. This is especially true when the pH is neutral, with a direct to indirect IQY increase equal to 0.09 (AT→13C) and 0.17 (13C→AT). While independent of the isomer, the IQY enhancement due to the permanence on S_2 does depend on pH. For instance, we find a 0.1 increase of the AT→13C IQY at pH=5 and 7, while the IQY is three times smaller at pH=3. In the 13C→AT case, the largest IQY enhancement is found at pH=7 with a spectacular 0.17 increase in the isomerization efficiency with respect to the direct subset.

This "S_2-assisted" mechanism efficiently increases the IQY; to our knowledge, such an effect was never reported before and further studies could give additional insights into its mechanism and importance. However, the majority of the trajectories do not visit S_2, as shown in the last row of Table 1: the ratio between the numbers of trajectories in the direct and in the indirect subsets is comprised between 3.6 and 5.5. Similar to the IQY, this ratio is pH-independent in the case of the 13C→AT isomerization. However the AT→13C ratio varies significantly with the pH: it is larger than 5 at pH=3 and 5, while it is reduced to 3.6 at pH=7. We conclude that the S_1/S_2 state mixing has a noticeable (and often pH-dependent) impact on the retinal IQY in our simulations of the ASR photochemistry.

We observe a direct correlation between the IQY and the number of trajectories hopping to S_0 when the C_{13}=C_{14} dihedral angle is close to 90° (Figure 4): unsurprisingly, when the chromophore twists around the correct bond, the overall probability of completing the isomerization increases. The 13C→AT intercept value, i.e., $x = 0$ in Figure 4, is close to 0, as expected. However, the AT→13C intercept is -0.15, suggesting that a linear relationship cannot fully represent the physics involved in the process. Extrapolating to 100% of C_{13}=C_{14} dihedral angles close to 90° (i.e., $x = 1$ in Figure 4) results in IQYs close to 0.7 as an upper limit for the ASR photoisomerization in the present simulations.

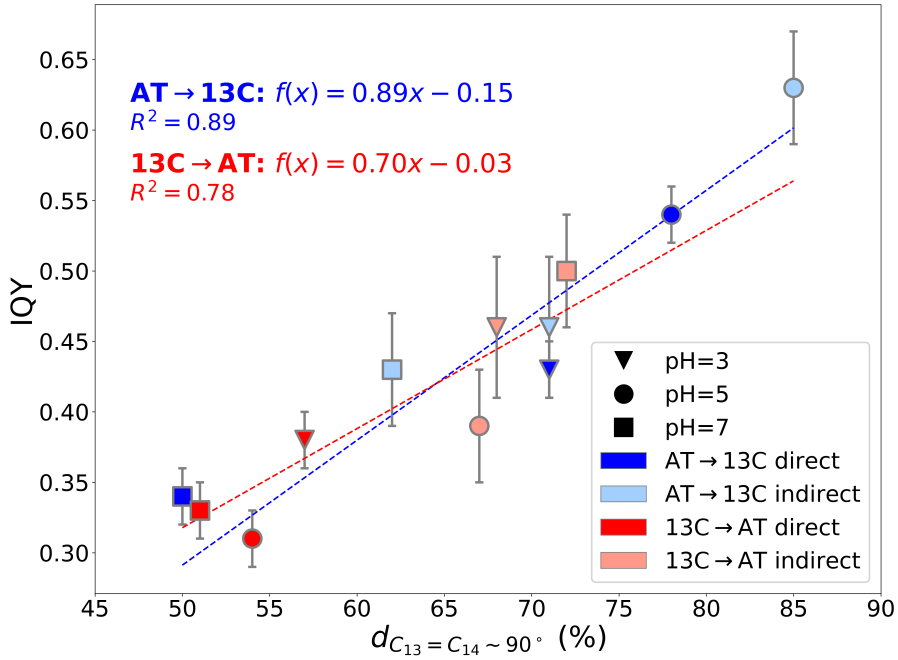

Figure 4: Correlation between the IQY obtained for the direct and indirect subsets of trajectories and the amount of them hopping to the ground state with a C_{13}=C_{14} dihedral angle close to 90° ($d_{C_{13}=C_{14}\sim 90°}$). The IQY uncertainties, represented as vertical bars on top of each data point, are calculated as $\sqrt{\mathrm{IQY}(1-\mathrm{IQY})/n}$ with n the total number of trajectories in the considered subset.

Given (i) the pH-dependence of the IQY and (ii) the correlation between the IQY and the number of trajectories hopping at C_{13}=C_{14} torsion angles close to 90°, we can expect the latter to be also pH-dependent, as we demonstrate in the following.

3.3 Average hopping time

The hopping time (t_{hop}, defined as the time at which a given trajectory hops from S_1 to S_0 in the context of Tully's fewest switches surface hopping) is not an observable, i.e., it cannot be compared directly to any experimentally measured property. Nevertheless, within the non-adiabatic fewest-switches surface-hopping MD methodology, its average value, $\bar{t}_{\mathrm{hop}}$, can be related to the excited-state lifetime independently of the $S_1 \rightarrow S_0$ decay channels, at variance with decay time constants which are introduced in user-defined kinetic models.

While we study the latter in the next section, we start here by showing how pH influences the t_{hop} distribution and $\bar{t}_{\mathrm{hop}}$ values (per retinal isomer and per pH value).

The distribution of t_{hop} against the C_{13}=C_{14} dihedral angles at the moment of a hop is presented in Figure 5. Independently of the pH value and isomer, we can distinguish two behaviors. In one group (named "alternative subset" in the following), the hops occur when the C_{13}=C_{14} dihedral angle has a similar value to the one it had at $t = 0$ (yellow set in Figure 5). The vast majority of this subset is nonreactive, i.e., is characterized by aborted isomerizations around bonds other than C_{13}=C_{14}; the corresponding t_{hop} spread between 200 fs and 2 ps, with many of them occurring after 750 fs. In the second group (indicated as "C_{13}=C_{14} subset" in the following), the C_{13}=C_{14} dihedral angle value at the moment of a hop is close to 90°, indicating an attempted isomerization around the "correct" bond. While t_{hop} for this subset can be as large as 2 ps, most of the corresponding trajectories hop between 100 fs and 400 fs. This subset can be further split into a reactive subset (i.e., successful isomerization, blue set in Figure 5) and a nonreactive subset (i.e., aborted isomerization, red set in Figure 5). The kernel density functions at the right side of each panel in Figure 5 show that aborted isomerizations at large timings are mainly due to the alternative subset. Accordingly, it is difficult to foresee a noticeable excited state lifetime difference between the reactive and unreactive subsets.

While pH does not modify the global picture drawn in Figure 5, its effect is more evident when looking at $\bar{t}_{\mathrm{hop}}$ (Figure 6). Within the full set of trajectories, this property is pH-dependent only when retinal's initial conformation is AT (see Figure 6A): $\bar{t}_{\mathrm{hop}}$ gets delayed at pH=5 compared to both pH=3 and pH=7. Also visible in Figure 5, the 13C $\bar{t}_{\mathrm{hop}}$ has a narrower distribution and is always shorter than the AT one for any pH value, but their difference is particularly large (135 fs) at pH=5.

We can determine $\bar{t}_{\mathrm{hop}}$ values for the reactive, nonreactive, and alternative subsets (Figure 6B). Confirming what is observed in Figure 5, in the latter case $\bar{t}_{\mathrm{hop}}$ is always larger than 500 fs, while the reactive and nonreactive subsets feature $\bar{t}_{\mathrm{hop}}$ shorter than 300 fs. This

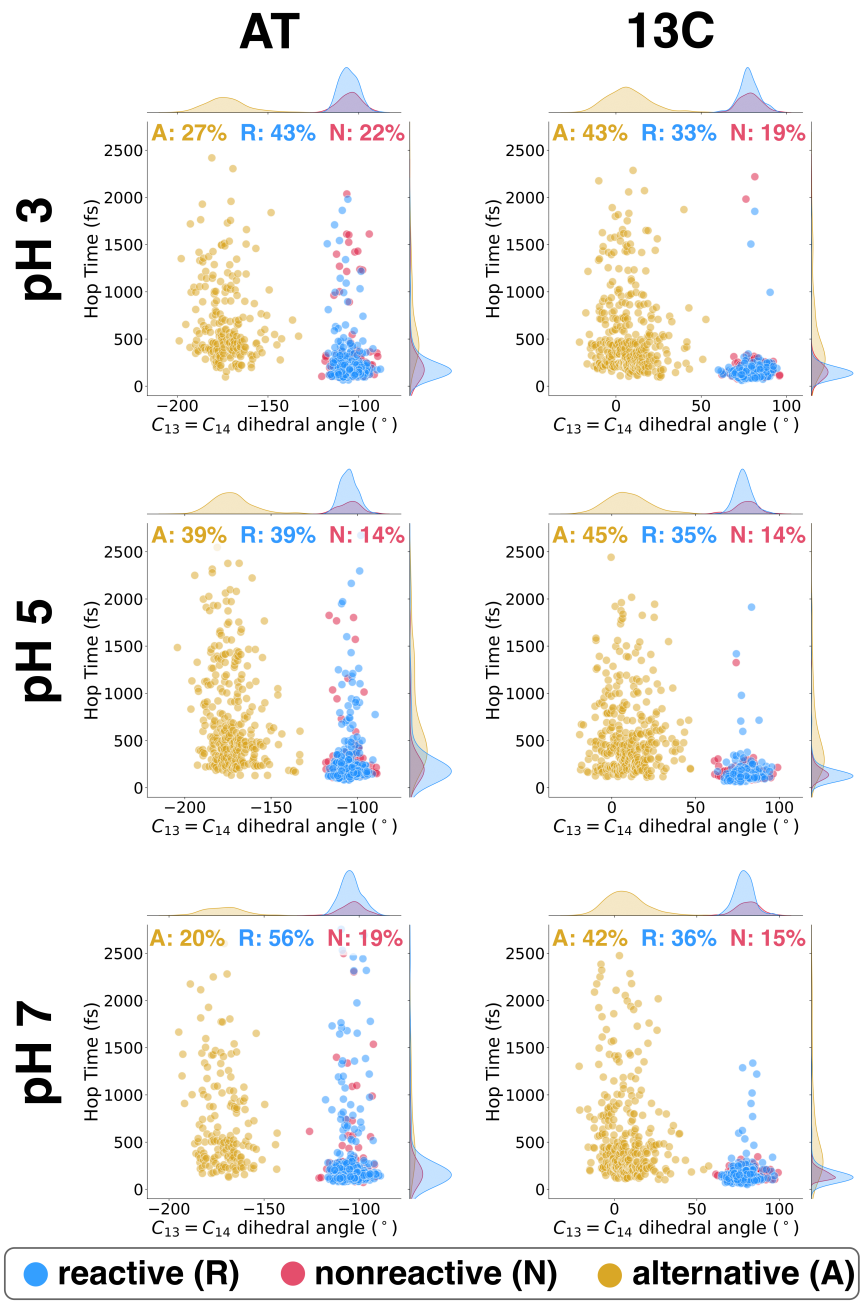

Figure 5: t_{hop} as a function of the torsion angle around the C_{13}=C_{14} bond as scatter plots and kernel distribution functions. The total set of surface hopping trajectories is divided into reactive, nonreactive, and alternative as mentioned in the text. The subset size percentages are calculated with respect to the total set of trajectories, which include the ones that do not hop within 2 ps (see Supporting Information for their population size).

result suggests two different excited state lifetimes, one shorter corresponding to attempted isomerizations around the C_{13}=C_{14} bond, and one longer for the aborted isomerizations

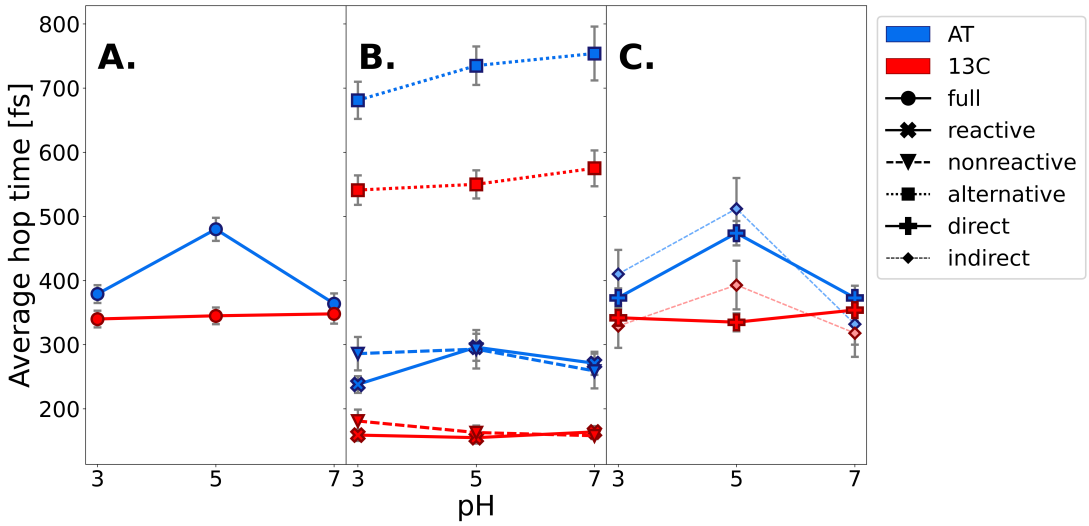

Figure 6: $S_1 \rightarrow S_0$ $\bar{t}_{\text{hop}}$ as a function of pH for AT and 13C models; standard errors are given as grey error bars. (A) Full set of valid trajectories. (B) Reactive, nonreactive, alternative channels. (C) Direct and indirect subsets.

around any other double bond in the retinal. Interestingly, panel 6B also shows clearly that the AT isomer has a higher $\bar{t}_{\text{hop}}$ than 13C, independent of the deactivation mechanism or the pH value.

We also analyzed $\bar{t}_{\text{hop}}$ within the direct (i.e., trajectories not visiting S_2) and indirect (i.e., trajectories visiting S_2) subsets (Figure 6C). Intuition suggests that indirect trajectories may take longer to decay to the ground state, given the detour. However, this is not always the case: in the pH7-AT, pH3-13C, and pH7-13C sets, the indirect subset proves to get, on average, faster decay. This indicates that under certain circumstances (e.g., the pH or the chromophore isomer), the nuclear wavepacket split between S_1 and S_2 before recombining on S_1 may favor rapid decay to the ground state. A similar population transfer scheme has already been documented *in silico* in a Channelrhodopsin Chimera, where it was hypothesized it could explain the long-time component in the protein's photokinetics.[37] However, to our knowledge, this is the first time that evidence of an accelerating action induced by visiting S_2 is found, an intriguing hypothesis that deserves further exploration in the future.

In general, the pH has the same impact on the direct and indirect subsets and on the

two isomers: $\bar{t}_{\mathrm{hop}}$ increases by several tens of fs from pH=3 to pH=5, then decreases by a slightly bigger amount at pH=7. The 13C direct case constitutes an exception, since its $\bar{t}_{\mathrm{hop}}$ is virtually unaffected by the pH, as evidenced also by the analysis performed on the whole 13C set (Figure 6A).

3.4 Decay rates

Similar to other rhodopsins,[44] ASR's photochemistry involves three electronic states (S_0, S_1 and S_2), at least in the Franck-Condon region where S_1 and S_2 are close in energy;[45] such a three-state model can be complex to reproduce with standard kinetic approaches, which assume only first-order processes (if all the three states are populated at the same time). In our case, though, a close inspection of the three state populations reveals that the population transfers to S_0 only after S_2 gets depopulated. Since our goal is to assess pH effects on the ASR photochemistry, we considered the $S_1 \rightarrow S_0$ part of the population time evolution only and denote t_{start} the lower bound of the corresponding time window.

Since we want to study how pH can change the retinal photodynamics, we have derived the most compact, yet effective, kinetic model for the $S_1 \rightarrow S_0$ decay. As extensively detailed in Supporting Information, it is based on two distinct S_1 populations, from now on denoted P^1_{fast} and P^1_{slow}, which eventually decay to the ground state and increase its population P^0. This model has been already used to experimentally fit decay rates in microbial rhodopsins, as exemplified by Hasson *et al.*[44] in bacteriorhodopsin. This kinetic model translates to the following first-order differential equations:

$$\begin{aligned} \frac{\mathrm{d}P^1_{fast}}{\mathrm{d}t} &= -k_{fast}P^1_{fast}(t) \\ \frac{\mathrm{d}P^1_{slow}}{\mathrm{d}t} &= -k_{slow}P^1_{slow}(t) \\ \frac{\mathrm{d}P^0}{\mathrm{d}t} &= k_{fast}P^1_{fast}(t) + k_{slow}P^1_{slow}(t) \end{aligned} \tag{1}$$

Such a model contains four unknown parameters: the two decay constants k_{fast} and k_{slow},

as well as the initial populations $P^1_{fast}(t = t_{start})$ and $P^1_{slow}(t = t_{start})$. The quality of this model depends heavily on the initial guess for the four parameters to optimize. It is difficult to estimate it from the inspection of the trajectories. Hence, we have resorted to an indirect, yet controlled, multi-step approach as described in Supporting Information. Most notably, we take advantage of the result presented in the previous section, i.e. the apparent different time scales for the alternative subset and for the reactive and unreactive subsets, to produce a good guess for the parameters in the bi-exponential model using mono-exponential models. The resulting models are of excellent quality, as demonstrated by the low root mean square deviation values with respect to the S_1 population evolution.

In ASR, the AT→13C and 13C→AT photoisomerizations are expected to be characterized by different values of both decay constants and initial populations. While AT→13C exhibits an energy barrier on S_1 which can be lowered by means of the retinal–opsin interactions, 13C→AT is essentially barrierless.[46] We reported the fitted time constants $\tau_i = 1/k_i$ in Table 2 with the corresponding populations at t_{start}, i.e., the time at which the ground state population starts to increase - we included the corresponding S_1 population time evolution and their fitted curves in Supporting Information. Independently of the retinal isomer and the pH value, two distinct channels, represented by one exponential function each, allow to fit the 2 ps S_1 population evolution. The main decay channel is ultrafast, with a time constant always comprised between 90 and 150 fs, while the secondary channel is much slower (1.0 to 2.0 ps). The inspection of the initial populations at t_{start} reveals that both the fast and slow channels are already populated, even if the fast one is always the principal one.

While the comparison of the fitted decay time constants with the ones coming out of ultrafast transient absorption experiments is tedious (decay models are not always the same), we stress that our computed shortest excited state lifetimes (i.e. $t_{start} + \tau_{fast}$) lie in the range 160–230 fs, in good agreement with experiments.[25,26]

We repeated the same fitting procedure within the direct and indirect subsets (Table 2) to complement the data obtained in the previous sections. Predictably, the indirect

Table 2: Fitted time-constants τ_i (in fs) and initial populations P_i^1 of a bi-exponential model (Equations 1) modeling the $S_1 \rightarrow S_0$ decay in ASR. t_{start}, the time at which S_0 starts to be populated, indicates the lower bound of the time window used in the fitting procedure, 2000 fs being the upper bound.

	AT					**13C**				
	t_{start}	P^1_{fast}	τ_{fast}	P^1_{slow}	τ_{slow}	t_{start}	P^1_{fast}	τ_{fast}	P^1_{slow}	τ_{slow}
					pH=3					
Full	64	0.79	143	0.21	1368	60	0.84	129	0.16	1369
Direct	64	0.80	146	0.20	1461	60	0.83	130	0.17	1201
Indirect	108	0.83	87	0.17	913	83	0.89	100	0.11	2238
					pH=5					
Full	81	0.72	146	0.28	1318	61	0.79	126	0.21	1096
Direct	81	0.73	148	0.27	1257	61	0.80	130	0.20	1099
Indirect	107	0.74	107	0.26	1498	86	0.79	83	0.21	1021
					pH=7					
Full	71	0.86	91	0.14	1276	49	0.83	139	0.17	1677
Direct	71	0.85	93	0.15	1196	49	0.82	148	0.18	1560
Indirect	86	0.91	73	0.09	1546	90	0.91	73	0.09	2130

subset is always characterized by a t_{start} larger than the one in the direct subset or full set. It is also characterized by time constants that are significantly different from the ones in the corresponding direct subset or in the full set. In particular, the lifetime of the fastest decay channel is always shorter when the system visits S_2. Irrespective of the pH or of the retinal isomer, our results confirm that the S_1/S_2 mixing modifies significantly the retinal photodynamics.[47] Nevertheless, in the present case where the size of the direct subset is 3 to 5 times larger than the indirect subset, the latter cannot significantly alter the overall retinal photo-induced molecular dynamics.

The pH does not modify the essential features of the ASR decay: two channels with significantly different rates are present, the faster one being more populated than the slower one. However, interesting pH effects are observed in AT retinal. Going from acidic to neutral pH not only does the population in the fast channel increase (from 72% to 86%), but it also accelerates the overall dynamics, hence reducing the global excited state lifetime. Large pH-induced variations are also found in the indirect subsets, where P^1_{fast} can vary between 74% and 91% for AT retinal, and between 79% and 91% for 13C retinal. In the case of

AT retinal, the $S_1 \rightarrow S_0$ decay is accelerated when the pH becomes neutral: the fast time constant is significantly reduced (from about 145 fs to 90 fs) and the corresponding initial population increases to more than 70%.

3.5 Residue-based analysis

We want to showcase another type of analysis that could be performed with the workflow employed in this work. Up to now, we have investigated how pH impacts photochemically relevant properties, like IQY or $S_1 \rightarrow S_0$ decay time constants. In this section, we focus on identifying the molecular origin of such pH dependence by analyzing the effect of individual amino acids.

Changing the pH translates into changing the ratio between the protonated and deprotonated forms of titrated amino acids. In order to disentangle the effects of the two forms, we can compare the photochemical behavior at very different pH values, where we can assume the amino acid is either always protonated or always deprotonated. However, very different pH values imply very different charge distributions around the chromophore, hence making the comparison less reliable.

Here, we have chosen a slightly different strategy to statistically take into account the pH-dependent electrostatic environment in the cavity: we split the trajectory ensemble at a given pH into two subsets, one where a given amino acid is protonated, the other one where the same amino acid is deprotonated. Such a splitting can be repeated for each titrated residue, yielding many pairs of subsets for further analysis. $\bar{t}_{\mathrm{hop}}$, IQY values, and fitted $S_1 \rightarrow S_0$ decay time constants are given in Supporting Information for all the titrated amino acids. Here, we discuss the amino acids that are responsible for the biggest variations in these quantities (Table 3).

D120. With a low pK_{a} value,[14] this aspartic acid was already identified as the main amino acid responsible for the tiny red-shift in ASR λ_{max} when going from pH=3 to pH=5. Interest-

Table 3: Impact (in %, relative to the value in the protonated form, see Supporting Information) on $\bar{t}_{\mathrm{hop}}$, IQY and fitted time constant τ_{fast} of the deprotonation of several titrated amino acids, whose deprotonated fractions at the considered pH are indicated as x_{dep}. In the case of D98, the deprotonated dataset size at pH3 for the 13C isomer is too small to allow reliable fitting.

Model	$\boldsymbol{x_{dep}}$	$\bar{t}_{\mathrm{hop}}$	**IQY**	$\boldsymbol{\tau_{fast}}$
		D120		
pH3-AT	0.90	+19	-2	-9
pH3-13C	0.17	+21	-34	+25
		D57		
pH3-AT	0.21	+10	-5	-39
pH5-AT	0.72	-25	+35	-31
pH3-13C	0.11	+10	-18	-22
pH5-13C	0.67	+16	-20	+27
		D98		
pH3-AT	0.19	-1	+7	-11
pH5-AT	0.89	+53	-25	+27
pH3-13C	<0.001	N/A	N/A	N/A
pH5-13C	0.13	+30	-51	+47
		D217		
pH5-AT	0.12	-7	+8	-33
pH7-AT	0.93	+22	-3	+18
pH5-13C	0.23	+15	-19	+9
pH7-13C	0.85	-6	-10	-14

ingly, in the present simulations, D120 is 90% deprotonated at pH=3 in the AT conformation, while it is only 17% deprotonated in 13C. This allows us to analyze the deprotonation effect at the same pH value, revealing a conformation-independent 20% increase of the $\bar{t}_{\mathrm{hop}}$. Since this value cannot distinguish between the decay channels, we hypothesize that the D120 deprotonation results in a flatter S_1 potential energy surface at pH=3. In the case of 13C retinal, D120's deprotonation also induces a large IQY decrease (34%). This is in agreement with the general lower efficiency described previously for the 13C→AT isomerization at higher pH values: at more neutral pH, the presence of the less efficient deprotonated D120 increases. These effects are large enough to conclude that the D120 protonation state impacts quantitatively ASR's photochemistry.

D57. Similarly to D120, the titration of this aspartic acid induces a $\lambda_{\max}$ red-shift at acidic pH.[14] Its higher pK_a value (4.94)[14] implies that its titration process barely starts at pH=3 and is fully ongoing at pH=5. In fact, in both isomers, D57 is substantially protonated at pH=3, while the ratio between the protonated/deprotonated forms is closer to one at pH=5. The protonation state of D57 has a large impact on the photochemistry of the AT isomer at pH=5. Upon D57's deprotonation, the $\bar{t}_{\text{hop}}$ decreases by 25%, IQY increases by 35%, while decay time constants become smaller by about 30%. Therefore, at pH=5, the AT→13C photoisomerization is much more efficient and faster if a negative charge is located on D57. The 13C→AT is more moderately affected by the protonation state of D57.

D98. This aspartic acid represents a more extreme case than D57. An accurate pK_a value is difficult to achieve since its side-chain continuously flips between opposing dielectric environments represented by solvent and membrane.[14] Given our estimation of its pK_a to 3.75 in our previous study, we only titrated it at pH=3 and pH=5. Regarding AT retinal, D98 is 81% protonated at pH=3 and becomes 89% deprotonated at pH=5. While its deprotonation does not induce large photochemical modifications at pH=3, increasing the pH lowers the AT→13C efficiency and speed (+53% for $\bar{t}_{\text{hop}}$, -25% for the IQY, +27% for τ_{fast}). In the case of the 13C retinal isomer, D98 is exclusively protonated at pH=3 and becomes only 13% deprotonated at pH=5, indicating a larger pK_a value for the 13C isomer. Consistently with the AT case, the D98 deprotonation process is predicted to largely decrease the 13C→AT efficiency (+30% for $\bar{t}_{\text{hop}}$, -51% for the IQY, +47% for τ_{fast}). Given the position of D98, relatively distant from the chromophore, these results confirm the importance of the long-range electrostatic interactions in photodynamics.

D217. While H21, E36, H219, and D217 constitute a cluster the titration of which is responsible for the $\lambda_{\max}$ blue-shift occurring between pH=5 and pH=7,[14] we can assume D217 is the leading titration site in the quartet.[14,23] Our results indicate retinal isomer-dependent D217 deprotonation effects. In fact, upon its deprotonation, the AT→13C IQY does not

change with the pH, while the reverse transformation becomes somehow less efficient by 10 to 19%. The pH also seems to play a part, likely due to the changes in the surrounding electrostatic environment. For instance, while the AT→13C becomes 33% slower at pH=5, it is 18% faster at pH=7. Conversely, D217 deprotonation induces a 9% faster 13C→AT transformation at pH=5 and a 14% slower one at pH=7. From these results alone, it is hard to disentangle what is the exact role of D217. However, as already mentioned, D217 is part of a cluster of residues titrating in the same pH range, suggesting correlated deprotonation processes. Accordingly, we should also consider the effect due to their deprotonation, resulting in $2^2 \times 3^2 = 36$ different protonation microstates. A possible way to study the protonation state change effects for the cluster is to consider it as a unique titration site, and use the most abundant microstate to represent the protein at pH=7. While this is a perfectly viable interpretation, the most populated microstate for this cluster also corresponds almost exactly with the most populated protein charge state, as described in the next section. In this work, we chose to explore the latter option due to its broader relevance across different windows and its potential to establish a new analytical framework.

3.6 Exploration of most populated protonation microstate and charge state

As mentioned in the introduction, currently the standard procedure to model the nonadiabatic dynamics of a protein at a given pH relies on picking and fixing the protonation state of every titratable amino acid, or, in other words, using only one protonation microstate. In this work, we explored the advantages and drawbacks of using a more realistic microstate distribution. The resulting large dataset allows us to compare the results and accuracy of the single microstate picture compared to the full set.

We start by examining how representative of the entire system the most populated protonation microstate is (see Figure 7B). Our CpHMD simulations indicate that at pH values where not many amino acids titrate, the most populated microstate accounts for more than

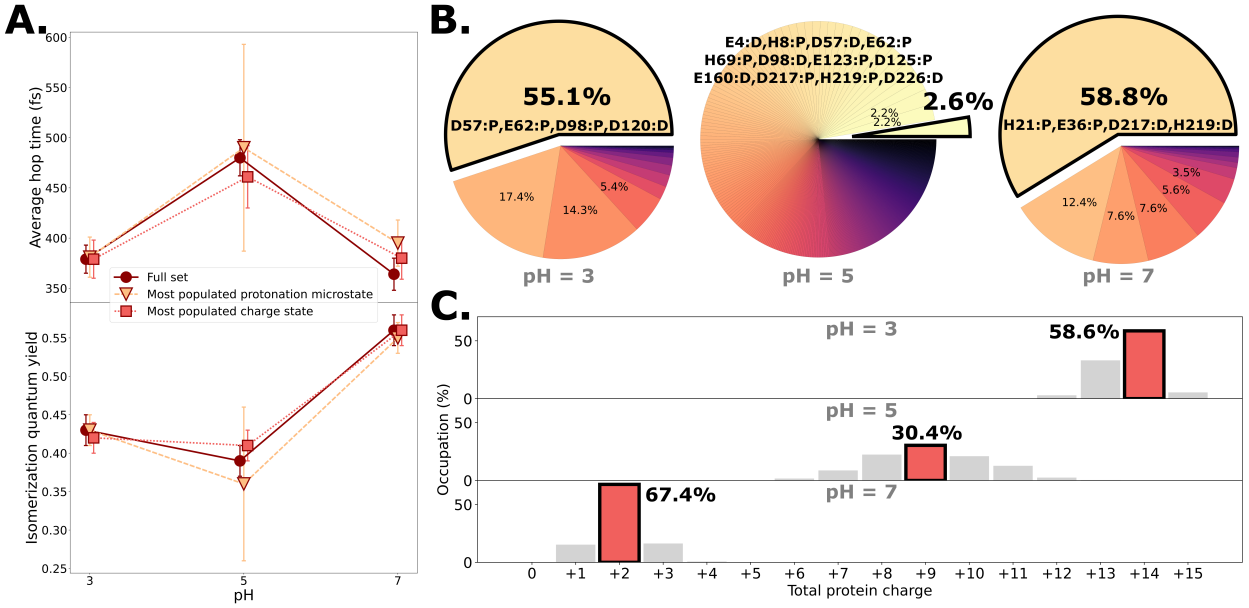

Figure 7: Comparison of the results for the AT isomer when considering the full dataset, only the most populated protonation microstate or the most populated protein charge state. (A) Upper panel: average hop times in the three datasets and standard errors as error bars; lower panel: IQY in the three datasets and relative standard errors as error bars. The datasets have been slightly staggered with respect to the pH to ease the comparison. (B) A pie chart representation of the population of each protonation microstate according to our CpHMD simulation; only populations higher than 2% are reported; the most populated microstate is reported in the text, where "P" stands for "protonated" and "D" for deprotonated. (C) A histogram depiction of the protein charge state population distribution along our CpHMD simulations; the highest populated charge state is indicated in red.

50% of the population. As such, at pH=3, where the majority of the titratable residues are still largely protonated, and at pH=7, where ASP and GLU residues are typically deprotonated, while TYR, CYS, LYS, and ARG residues are mostly protonated, one microstate might constitute a reasonable proxy to represent the system. However, at intermediate pH windows where several amino acids might actively titrate like at pH=5, the most populated microstate might lose its meaning, as the population it represents might be incredibly tiny (e.g., 2.6%). In other words, at certain pH values, the population distribution over protonation microstates might be much larger. Therefore, picking one of these microstates that carries such a small fraction of the total population might lead to artifacts.

In our case, when choosing the most populated microstate, the $\bar{t}_{\text{hop}}$ gets overestimated

by 10-30 fs at both pH=5 and pH=7, and the IQY gets underestimated by 0.03 at pH=5 (see Figure 7A). This confirms a loss of accuracy when switching to the single microstate picture. However, one needs to consider the impressive computational cost of a study that includes comprehensive protonation space sampling like the present one and evaluate what the best compromise is for the case in question.

A possibility might be sampling from the total protein charge space: one might use the relatively affordable CpHMD simulations to analyze the distribution of protein charge. In fact, one can conceptualize pH changes as changing the protonation microstate distribution, or consider the protein as a large macromolecule for which we can calculate the total charge at each step of the CpHMD. The advantage of this method is that sampling is confined to a smaller microstate set, saving computational resources while still maintaining a more accurate, realistic description of pH.

In our case, the most populated charge state represents a similar (but higher) portion of the population with respect to the most populated microstate at pH=3 and pH=7 (see Figure 7C). However, the most populated charge state at pH=5 carries 30% of the population, compared to the meager 2% of the most populated microstate, ensuring a more comprehensive description of our system. In fact, while the most populated charge microstate does not perfectly reproduce the $\bar{t}_{\mathrm{hop}}$ or IQY at all pH values compared to the full set, it yields generally closer results than the microstate set ($\bar{t}_{\mathrm{hop}}$ values within 15 fs and IQY within 0.02 of the full set ones).

In conclusion, we suggest that sampling from a fixed total charge state instead of a fixed protonation microstate might constitute a good compromise to compute reasonably accurate pH-dependent properties at the only added cost of initially running CpHMD.

3.7 Future improvements

The results presented above demonstrate that the here-introduced enhanced CpHMD-then-QM/MM protocol is a valuable tool for investigating how pH can tune the photochemical

properties of a photoactive protein, like ASR. Our workflow not only allowed us to estimate pH-dependent excited state lifetimes and isomerization quantum yields but also gave us insights into the individual role of the titratable amino acids and the importance of sampling from the full protonation and conformational space. However, improvements can be expected by refining some aspects:

- The CpHMD technique is computationally expensive, even when combined with the replica-exchange approach to accelerate convergence. The CpHMD simulations performed in this work, which involve the replica exchange technique and the introduction of a membrane (which requires more expensive Generalized Born calculations in the protonation state change attempts due to the larger number of non-solvent atoms), required an estimated 2 million CPU hours. Given the typical size of a medium protein protonation space, the pre-selection of a subset of titratable amino acids, e.g., using a simple approach like the Minimal Electrostatic Model,[12] significantly reduces the computational cost. While there is a higher risk of trapping in local minima - especially when several titrated residues exchange a proton - such a compromise could prove valid. Complementary to this protonation space reduction, CpHMD simulations could be constrained in order to sample only microstates belonging to the most populated charge state.

- Typically, running non-adiabatic MD requires accepting a compromise. This is especially true in our case, where we required thousands of QM/MM trajectories. The semi-empirical Hamiltonian chosen in this study is a good compromise between accuracy and computational efficiency. In the future, more accurate and/or faster electronic structure methods could be used thanks to recent advances in machine learning or GPU-accelerated code.

- We extracted snapshots (structure and protonation microstate) from CpHMD trajectories and used them as initial conditions for non-adiabatic semi-classical MD. Even

if we carefully chose these structures to qualitatively reproduce the protein absorption spectrum[14] at several pH values, we set their velocities to zero, i.e., the trajectories were produced at 0 K. At the same time, most of the protein remained frozen during the excited state dynamics. This approximation results in biased ballistic trajectories, initially following the steepest descent pathways, resulting in large momenta after a few time steps. We assumed that zeroed velocities at the MM coordinates would span the same conformational space that non-zero velocities at the QM coordinates would do; this is undoubtedly a crude approximation, even if chosen in the interest of computational affordability. In fact, a more accurate course of action would be to run ground state QM/MM molecular dynamics from the MM snapshots (including their velocities, or sampling them from a Boltzmann distribution) to correct possible force-field inaccuracies in the chromophore geometry and obtain initial velocities; this approach would result in a molecular model closer to what is achieved experimentally, e.g., in time-resolved ultrafast spectroscopies. However, we could not afford to do so for 6000 trajectories; interestingly the approximation we used yielded decay constants that are comparable to the ones typically reported experimentally in rhodopsins, probably due to error cancellation.

4 Conclusions

We have introduced an enhanced version of our previously published CpHMD-then-QM/MM protocol,[13] aiming at investigating how pH can modify the photochemical properties of a photoactive biomolecule. We used CpHMD to extensively sample initial conditions (positions and protonation states) for semi-classical non-adiabatic molecular dynamics simulations. With this protocol, we were able to analyze the pH-dependent deactivation mechanism and calculate the isomerization quantum yield and time constants for the excited state lifetime.

In our model, the retinal isomerization mechanism does not depend on pH in the 3-7

range: we observe an inversion of the BLA starting immediately after photoexcitation to allow torsion around a former double bond, which returns to the original value soon after decaying back to the ground state. The isomerization can be attempted around several double bonds of the retinal chromophore, but only the ones around the C13=C14 double bond are successful thanks to the constraining protein environment.

The AT→13C isomerization seems usually more efficient than the 13C→AT one in terms of quantum yield. Additionally, while the former tends to increase towards neutral pH values, the latter is pH-independent. In all cases, we observed a part of the trajectories briefly visiting S_2; interestingly, this less common mechanism leads to higher IQYs in our model (+0.03 to +0.14). If the IQY increase by almost 50% that we calculated when the pH changes from 5 to 7 (AT retinal isomer) was confirmed experimentally, it would also imply a major correlation between pH and excited state lifetime that needs to be explored.

Similarly to the IQY, the $\bar{t}_{\mathrm{hop}}$ differs between the isomers. The 13C value is virtually pH-independent (340 fs), while the AT $\bar{t}_{\mathrm{hop}}$ increases at pH=5 and is generally higher than the 13C one. The time efficiency of the different deactivation channels is clear: while reactive and nonreactive isomerizations are characterized by lower $\bar{t}_{\mathrm{hop}}$ values (below 300 fs), the aborted isomerizations around bonds other than C13=C14 lead to typically higher $\bar{t}_{\mathrm{hop}}$ values (above 550 fs). Visiting S_2 might also have a slightly accelerating effect in mildly acidic pH conditions.

Our calculations also revealed a bi-exponential decay mechanism, with one component characterized by a faster time constant (100-150 fs) and typically carrying 70%-90% of the population, and a slower, less used one (1 ps or more). pH effects can be observed in, for instance, the AT isomer, where the excited state lifetime decreases at pH=7 as a combination of an increase in the population of the fast decay channel and a decrease in its time constant.

Our large dataset also offered the chance to analyze the impact of protonation state changes in individual amino acids. This revealed that some residues, like D57, affect the photochemical properties somewhat drastically. However, trends are sometimes difficult to

extract, especially when the changes in the protonation state in one residue are correlated with other residues (i.e., interacting amino acids). As a matter of fact, such clusters would need to be considered as a poly-acid featuring several effective pK_{a} values that cannot be assigned to any particular amino acid.

The "protonation microstate" and the "protein charge state" are two useful ways to conceptualize and investigate pH-dependent properties in complex systems. We showed that the former, while widely used to model proteins, may prove insufficient to accurately describe the problem in certain situations, mainly due to the small fraction of the total population it may represent. Therefore, using a reduced set of weighted microstates that yield the most populated protein charge state might be an attractive option to retain a more realistic description of the problem at a low additional computational cost.

To conclude, adding CpHMD as a preliminary step to the simulation of the photochemistry of proteins can provide rich insights into the complex mechanisms that regulate the events following photoabsorption. The precious information extracted can be used to understand and tune the photoactivity of proteins, and, more ambitiously, suggest possible mutation sites to design tunable and/or improved mutants.

Acknowledgement

EP and NF thank the French Agence Nationale de la Recherche for funding (grant ANR-14-CE35-0015-02, project FEMTO-ASR). The Centre de Calcul Intensif d'Aix-Marseille is acknowledged for granting access to its high performance computing resources. Computational support and infrastructure was also provided by the "Centre for Information and Media Technology (ZIM)" at the Heinrich Heine University.

Supporting Information Available

The Supporting Information is available free of charge on the ACS Publications website.

Content: computational details, analysis of each ensemble of trajectories, BLA and dihedral torsions in the 13C sets, fitting S_1 decay time evolution, and residue-based analysis.

References

(1) Connors, K. A. *Chemical Kinetics: The Study of Reaction rates in solution*; Wiley, 1990.

(2) Han, J.; Burgess, K. Fluorescent Indicators for Intracellular pH. *Chem. Rev.* **2010**, *110*, 2709–2728.

(3) Zhukovsky, E. A.; Oprian, D. D. Effect of carboxylic acid side chains on the absorption maximum of visual pigments. *Science* **1989**, *246*, 928–930.

(4) Logunov, S. L.; Song, L.; El-Sayed, M. A. pH Dependence of the Rate and Quantum Yield of the Retinal Photoisomerization in Bacteriorhodopsin. *J. Phys. Chem.* **1994**, *98*, 10674–10677.

(5) Mathies, R. A.; Lugtenburg, J. In *Handbook of Biological Physics*; Stavenga, D. G., de Grip, W. J., Pugh Jr., E. N., Eds.; Elsevier Science B.V., 2000; Chapter The primary photoreaction of rhodopsin, pp 55–90.

(6) Ernst, O. P.; Lodowski, D. T.; Elstner, M.; Hegemann, P.; Brown, L. S.; Kandori, H. Microbial and Animal Rhodopsins: Structures, Functions, and Molecular Mechanisms. *Chem. Rev.* **2014**, *114*, 126–163.

(7) Subotnik, J. E.; Alguire, E. C.; Ou, Q.; Landry, B. R.; Fatehi, S. The Requisite Electronic Structure Theory To Describe Photoexcited Nonadiabatic Dynamics: Nonadiabatic Derivative Couplings and Diabatic Electronic Couplings. *Acc. Chem. Res.* **2015**, *48*, 1340–1350.

(8) Ferré, N., Filatov, M., Huix-Rotllant, M., Eds. *Density-Functional Methods for Excited States*; Topics in Current Chemistry; Springer International Publishing, 2016; Vol. 368.

(9) González, L., Lindh, R., Eds. *Quantum Chemistry and Dynamics of Excited States*; Wiley, 2020.

(10) Khaniya, U.; Mao, J.; Wei, R. J.; Gunner, M. R. Characterizing Protein Protonation Microstates Using Monte Carlo Sampling. *J. Phys. Chem. B* **2022**, *126*, 2476–2485.

(11) Laricheva, E. N.; Goh, G. B.; Dickson, A.; Brooks, C. L. pH-Dependent Transient Conformational States Control Optical Properties in Cyan Fluorescent Protein. *J. Am. Chem. Soc.* **2015**, *137*, 2892–2900.

(12) Stenrup, M.; Pieri, E.; Ledentu, V.; Ferré, N. pH-Dependent absorption spectrum of a protein: a minimal electrostatic model of Anabaena sensory rhodopsin. *Phys. Chem. Chem. Phys.* **2017**, *19*, 14073–14084.

(13) Pieri, E.; Ledentu, V.; Huix-Rotllant, M.; Ferré, N. Sampling the protonation states: the pH-dependent UV absorption spectrum of a polypeptide dyad. *Phys. Chem. Chem. Phys.* **2018**, *20*, 23252–23261.

(14) Pieri, E.; Ledentu, V.; Sahlin, M.; Dehez, F.; Olivucci, M.; Ferré, N. CpHMD-Then-QM/MM Identification of the Amino Acids Responsible for the Anabaena Sensory Rhodopsin pH-Dependent Electronic Absorption Spectrum. *J. Chem. Theory Comput.* **2019**, *15*, 4535–4546.

(15) Pedraza-González, L.; Cignoni, E.; D'Ascenzi, J.; Cupellini, L.; Mennucci, B. How the pH Controls Photoprotection in the Light-Harvesting Complex of Mosses. *J. Am. Chem. Soc.* **2023**, *145*, 7482–7494.

(16) Swails, J. M.; Roitberg, A. E. Enhancing Conformation and Protonation State Sampling

of Hen Egg White Lysozyme Using pH Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 4393–4404.

(17) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10*, 1341–1352.

(18) de Almeida Barbosa, N. M.; Zemmouche, M.; Gosset, P.; García-Iriepa, C.; Ledentu, V.; Navizet, I.; Didier, P.; Ferré, N. pH-Dependent Absorption Spectrum of Oxyluciferin Analogues in the Presence of Adenosine Monophosphate. *ChemPhotoChem* **2019**, *3*, 1219–1230.

(19) de Almeida Barbosa, N. M.; Gosset, P.; Réal, E.; Ledentu, V.; Didier, P.; Ferré, N. pH-Dependent absorption spectrum of oxyluciferin analogues in the active site of firefly luciferase. *Phys. Chem. Chem. Phys.* **2020**, *22*, 21731.

(20) Vogeley, L.; Sineshchekov, O. A.; Trivedi, V. D.; Sasaki, J.; Spudich, J. L.; Luecke, H. Anabaena Sensory Rhodopsin: A Photochromic Color Sensor at 2.0 Å. *Science* **2004**, *306*, 1390–1393.

(21) Sineshchekov, O. A.; Trivedi, V. D.; Sasaki, J.; Spudich, J. L. Photochromicity of Anabaena Sensory Rhodopsin, an Atypical Microbial Receptor with a cis-Retinal Light-adapted Form. *J. Biol. Chem.* **2005**, *280*, 14663–14668.

(22) Tahara, S.; Kato, Y.; Kandori, H.; Ohtani, H. PH-Dependent Photoreaction Pathway of the All-Trans Form of Anabaena Sensory Rhodopsin. *J. Phys. Chem. B* **2013**, *117*, 2053–2060.

(23) Rozin, R.; Wand, A.; Jung, K.-H.; Ruhman, S.; Sheves, M. pH Dependence of Anabaena Sensory Rhodopsin: Retinal Isomer Composition, Rate of Dark Adaptation, and Photochemistry. *J. Phys. Chem. B* **2014**, *118*, 8995–9006.

(24) Wand, A.; Rozin, R.; Eliash, T.; Jung, K.-H.; Sheves, M.; Ruhman, S. Asymmetric Toggling of a Natural Photoswitch: Ultrafast Spectroscopy of Anabaena Sensory Rhodopsin. *J. Am. Chem. Soc.* **2011**, *133*, 20922–20932.

(25) Cheminal, A.; Léonard, J.; Kim, S.-Y.; Jung, K.-H.; Kandori, H.; Haacke, S. 100 fs photo-isomerization with vibrational coherences but low quantum yield in Anabaena Sensory Rhodopsin. *Phys Chem Chem Phys* **2015**, *17*, 25429–25439.

(26) Roy, P. P.; Kato, Y.; Abe-Yoshizumi, R.; Pieri, E.; Ferré, N.; Kandori, H.; Buckup, T. Mapping the ultrafast vibrational dynamics of all-trans and 13-cis retinal isomerization in Anabaena Sensory Rhodopsin. *Phys. Chem. Chem. Phys.* **2018**, *20*, 30159–30173.

(27) Wada, Y.; Kawanabe, A.; Furutani, Y.; Kandori, H.; Ohtani, H. Quantum yields for the light adaptations in Anabaena sensory rhodopsin and bacteriorhodopsin. *Chem. Phys. Lett.* **2008**, *453*, 105–108.

(28) Sabri Dashti, D.; Meng, Y.; Roitberg, A. E. pH-Replica Exchange Molecular Dynamics in Proteins Using a Discrete Protonation Method. *J. Phys. Chem. B* **2012**, *116*, 8805–8811.

(29) Weingart, O.; Nenov, A.; Altoè, P.; Rivalta, I.; Segarra-Martí, J.; Dokukina, I.; Garavelli, M. COBRAMM 2.0 — A software interface for tailoring molecular electronic structure calculations and running nanoscale (QM/MM) simulations. *J. Mol. Model.* **2018**, *24*.

(30) Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.* **1990**, *93*, 1061–1071.

(31) Manathunga, M.; Yang, X.; Orozco-Gonzalez, Y.; Olivucci, M. Impact of Electronic State Mixing on the Photoisomerization Time Scale of the Retinal Chromophore. *J. Phys. Chem. Lett.* **2017**, *8*, 5222–5227.

(32) Manathunga, M.; Yang, X.; Olivucci, M. Electronic State Mixing Controls the Photoreactivity of a Rhodopsin with all-trans Chromophore Analogues. *J. Phys. Chem. Lett.* **2018**, *9*, 6350–6355.

(33) Fabiano, E.; Keal, T.; Thiel, W. Implementation of surface hopping molecular dynamics using semiempirical methods. *Chem. Phys.* **2008**, *349*, 334–347.

(34) Dral, P. O.; Wu, X.; Spörkel, L.; Koslowski, A.; Weber, W.; Steiger, R.; Scholten, M.; Thiel, W. Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Theory, Implementation, and Parameters. *J. Chem. Theory Comput.* **2016**, *12*, 1082–1096.

(35) Tuna, D.; Lu, Y.; Koslowski, A.; Thiel, W. Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Benchmarks of Electronically Excited States. *J. Chem. Theory Comput.* **2016**, *12*, 4400–4422.

(36) Calculating Absorption Shifts for Retinal Proteins: Computational Challenges. *J. Phys. Chem. B* **2005**, *109*, 3606–3615.

(37) Dokukina, I.; Nenov, A.; Garavelli, M.; Marian, C. M.; Weingart, O. QM/MM Photodynamics of Retinal in the Channelrhodopsin Chimera C1C2 with OM3/MRCI. *ChemPhotoChem* **2019**, *3*, 107–116.

(38) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(39) Ben-Nun, M.; Molnar, F.; Schulten, K.; Martínez, T. J. The role of intersection topography in bond selectivity of *cis-trans* photoisomerization. *Proc. Natl Acad. Sci.* **2002**, *99*, 1769–1773.

(40) Weingart, O.; Schapiro, I.; Buss, V. Photochemistry of Visual Pigment Chromophore Models by Ab Initio Molecular Dynamics. *J. Phys. Chem. B* **2007**, *111*, 3782–3788.

(41) Schapiro, I. The Origin of Bond Selectivity and Excited-State Reactivity in Retinal Analogues. *J. Phys. Chem. A* **2016**, *120*, 3353–3365.

(42) Gozem, S.; Luk, H. L.; Schapiro, I.; Olivucci, M. Theory and Simulation of the Ultrafast Double-Bond Isomerization of Biological Chromophores. *Chem. Rev.* **2017**, *117*, 13502–13565.

(43) Hinkley, D. R. C., D. V. *Theoretical Statistics*; Chapman and Hall/CRC: New York, 1979.

(44) Hasson, K.; Gai, F.; Anfinrud, P. A. The photoisomerization of retinal in bacteriorhodopsin: Experimental evidence for a three-state model. *Proc. Natl Acad. Sci.* **1996**, *93*, 15124–15129.

(45) Schapiro, I.; Ruhman, S. Ultrafast photochemistry of Anabaena Sensory Rhodopsin: Experiment and theory. *Biochim. Biophys. Acta Bioenerg.* **2014**, *1837*, 589–597.

(46) Kiefer, H. V.; Gruber, E.; Langeland, J.; Kusochek, P. A.; Bochenkova, A. V.; Andersen, L. H. Intrinsic photoisomerization dynamics of protonated Schiff-base retinal. *Nat. Commun.* **2019**, *10*.

(47) Yang, X.; Manathunga, M.; Gozem, S.; Léonard, J.; Andruniów, T.; Olivucci, M. Quantum–classical simulations of rhodopsin reveal excited-state population splitting and its effects on quantum efficiency. *Nat. Chem.* **2022**, *14*, 441–449.

TOC Graphic