Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202200139
Chemistry Europe
European Chemical Societies Publishing
www.chemeurj.org

Deciphering the Spectral Tuning Mechanism in Proteorhodopsin: The Dominant Role of Electrostatics Instead of Chromophore Geometry

Jonathan R. Church[+],[a] Gil S. Amoyal[+],[a] Veniamin A. Borin,[a] Suliman Adam,[a] Jógvan Magnus Haugaard Olsen,[b] and Igor Schapiro*[a]

Abstract: Proteorhodopsin (PR) is a photoactive proton pump found in marine bacteria. There are two phenotypes of PR exhibiting an environmental adaptation to the ocean's depth which tunes their maximum absorption: blue-absorbing proteorhodopsin (BPR) and green-absorbing proteorhodopsin (GPR). This blue/green color-shift is controlled by a glutamine to leucine substitution at position 105 which accounts for a 20 nm shift. Typically, spectral tuning in rhodopsins is rationalized by the external point charge model but the Q105L mutation is charge neutral. To study this tuning mechanism, we employed the hybrid QM/MM method with sampling from molecular dynamics. Our results reveal that the positive partial charge of glutamine near the C_{14}–C_{15} bond of retinal shortens the effective conjugation length of the chromophore compared to the leucine residue. The derived mechanism can be applied to explain the color regulation in other retinal proteins and can serve as a guideline for rational design of spectral shifts.

Introduction

Proteorhodopsins (PRs) constitute a family of proteins that act as light-driven proton pumps. They were first identified by a metagenomic screening of an uncultivated marine γ-proteobacterium from the SAR86 group by Béjà et al. in 2000.[1] Since their initial discovery, PRs have been found in bacteria, archaea, and eukaryotes.[2–4] Currently it is estimated that 50–70% of marine microbes within the photic zone of the ocean contain PRs.[5–8] It has been suggested that PRs play a key role in marine solar energy conversion due to their wide presence in marine life and their unique ability to absorb sunlight.[8] This energy conversion by marine life is extremely important as photoactivated processes in marine systems have been shown to account for 40% of the annual global carbon fixation.[9]

PRs belong to the microbial rhodopsins, which share a high structural similarity, and are characterized by seven transmembrane helices (TMHs) (Figure 1).[10] These proteins carry a retinal (RET) chromophore linked to the opsin via a conserved lysine which together form a retinal protonated Schiff base (RPSB) (Figure 1).[1] PRs use the bound chromophore to absorb light in the visible range, which then isomerizes and initiates a proton transport process. This process occurs over a series of steps in a photocycle and is mediated through the interactions between the RPSB, water molecules and nearby amino acid sidechains.[11–13]

PR has evolved into two main groups in order to optimize light absorption at various depths of the ocean.[4] They are

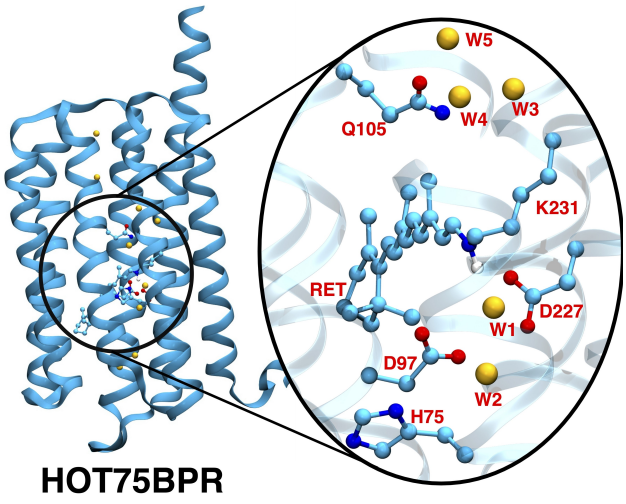

Figure 1. Blue-absorbing proteorhodopsin from HOT75 (HOT75BPR) wild type. The zoom-in shows the binding pocket of the retinal with amino acid sidechains H75, Q105, counterions D97 and D227 as well as the nearby water molecules (W1–W5) in yellow.

[a] *Dr. J. R. Church,[+] G. S. Amoyal,[+] Dr. V. A. Borin, Dr. S. Adam, Prof. I. Schapiro*
Fritz Haber Center for Molecular Dynamics Research, Institute of Chemistry
The Hebrew University of Jerusalem
Jerusalem 9190401 (Israel)
E-mail: Igor.Schapiro@mail.huji.ac.il

[b] *Prof. J. M. H. Olsen*
DTU Chemistry
Technical University of Denmark
DK-2800 Kongens Lyngby (Denmark)

[+] *These authors contributed equally to this work.*

Supporting information for this article is available on the WWW under https://doi.org/10.1002/chem.202200139

grouped by their absorption maxima: green-absorbing PRs (GPRs) ($\lambda_{max} \approx 525$ nm), which are found at the shallow ocean's surface, and blue-absorbing PRs (BPRs) ($\lambda_{max} \approx 490$ nm), which are typically found at greater depths.[7,14] The residue in position 105 has been experimentally identified to be a key residue responsible for the spectral shift. This specific position is documented to cause a ~20 nm bathochromic shift upon a glutamine (Q) to leucine (L) mutation.[13,15,24–28,16–23] Experimental and computational studies have been performed to elucidate the exact nature of the Q105L color-tuning mechanism.[13,22,23,25,26,28,29] It was found that the color switching Q105L mutation affects the C_{14}–C_{15} bond length of the retinal polyene chain.[21,22] In general, the bond length alternation (BLA) or bond order alternation (BOA), correlate strongly with the absorption maxima of rhodopsins.[30–32] The BOA is defined as the difference between the average bond order (BO) of the single bonds subtracted from the average BO of the double bonds.[32] Likewise, the BLA is the difference in the average length of single bonds from those of double bonds.[32–34] These factors are often used to quantify the amount of conjugation in the retinal; however, it still remains unclear how such a small change causes a 20 nm spectral shift. Typically, the spectral tuning in retinal proteins is explained on the basis of the external point charge model by Honig and coworkers. It is based on the observation that the excitation of retinal is accompanied by a transfer of the positive charge on the Schiff base moiety towards the β-ionone ring (Figure 2).[35–38] Hence, this model has been used to explain shifts in the absorption maximum when placing a charged residue at either end of the chromophore by noting the differential effect on the ground (S_0) and excited state (S_1) energies. However, in BPR position 105 is located further away from the Schiff base and both Q105 and L105 are neutral. The spectral tuning mechanism of the Q105L mutation in PRs presents a challenge to the external point charge model.

A prerequisite for a computational study of the color tuning mechanism is a high-resolution protein structure.[39] Earlier studies have relied on homology models based on the amino acid sequence of the Monterey Bay GPR (PDB code: 2L6X) using bacteriorhodopsin as a template.[40–43] However, homology models can often impede predictions due to the sensitivity of the chromophore to small displacements of the surrounding side chains.[44–46] Moreover, the uncertainty about the protonation state of the D97 counterion in GPR (pKa value of ~7) can alter the result of the simulation.[43] BPRs on the other hand have been successfully resolved, including structures with key mutations. Three crystal structures of BPR were solved by Ran et al.[10] Two were resolved from different locations: Hawaii at a depth of 75 meters (HOT75BPR) and the Mediterranean at 12 meters (Med12BPR, PDB: 4JQ6). Two important mutated variants of HOT75BPR (or simply BPR) were crystallized: the counterion mutation D97N (PDB: 4KLY) and the double mutant D97N/Q105L (PDB: 4KNF).[10] The possibility to use experimentally resolved crystallographic structures, rather than homology models, as well as the availability of key mutants, serves as a great starting point for computational studies.[22,23]

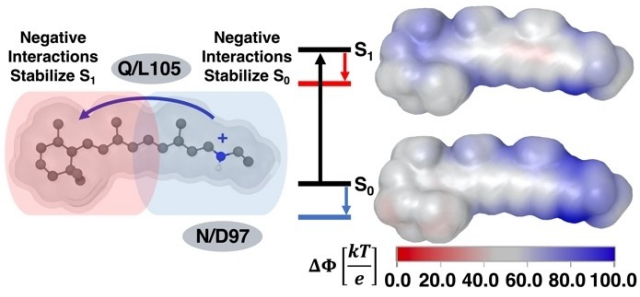

Figure 2. Illustration of the external point charge model for the retinal chromophore by Honig and coworkers.[36–38] **Right**: the positive charge of the retinal chromophore in the ground state (S_0) is localized near the Schiff base. Upon excitation to S_1 the positive charge is partially transferred towards the β-ionone ring. **Left**: Interactions with the protein environment near the Schiff base or the β-ionone ring can impact the stability of S_0 and S_1. These interactions can lead to shifts in the absorption maximum. The locations of the key residues examined in this study including position 105 and 97 are shown.

The aim of this study is to use hybrid QM/MM methods to investigate the origin of the spectral tuning in the BPR which cannot be explained by the external point charge model by Honig and coworkers. To address this question, we have used spectra constructed from 100 structures sampled from hybrid quantum mechanics/molecular mechanics (QM/MM) molecular dynamics (MD) simulations. We performed these simulations by treating the QM region with the semi-empirical DFTB2+D method (See Experimental Section). The excitation energies were computed with time-dependent density functional theory (TD-DFT) with the CAM–B3LYP functional as well as an ab initio wave function method (RI-ADC(2)). In addition to using an electrostatic embedding scheme [EE], we also employ polarizable embedding [PE] to model the protein environment due to the nature of the mutation from glutamine to leucine. Polarizable embedding is key to understanding the spectral tuning in PR as the Q105L mutation does not change the charge of the binding pocket and involves switching a polar residue with a non-polar residue. We examined the effect that the Q105L mutation has on the protein-chromophore interactions and the resulting changes to the UV-Vis absorption spectrum of HOT75BPR (Figure 1). The D97N mutant was also analyzed as a quality check because this residue is one of the counterions of the RPSB and has a strong impact on the absorption spectrum.[47]

Results and Discussion

The absorption spectra are presented relative to the computed absorption maximum of WT in Figure 3A. The experimental trends of the spectral shifts were qualitatively reproduced independent of the method.[13,17,27]

Two trends become apparent: (i) the spectral shifts at the RI-ADC(2) level of theory are larger than those at TD-CAM–B3LYP level and (ii) PE results in smaller shifts than EE. The difference between the shifts produced by the RI-ADC(2) and TD-CAM–B3LYP can be explained by the increased sensitivity of the RI-ADC(2) method with respect to the basis set

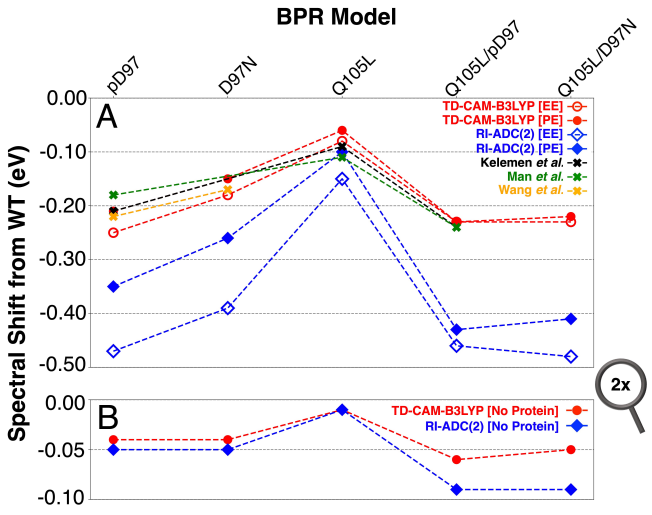

Figure 3. Spectral shifts (in eV) relative to the WT absorption maximum. These shifts were obtained from either TD-CAM–B3LYP or RI-ADC(2) with the cc-pVDZ basis set. (A) Spectral shifts obtained in the protein environment using the [EE] or [PE] QM/MM approach and experimental values for comparison.[13,17,27] (B) Spectral shifts obtained after the removal of the protein environment.

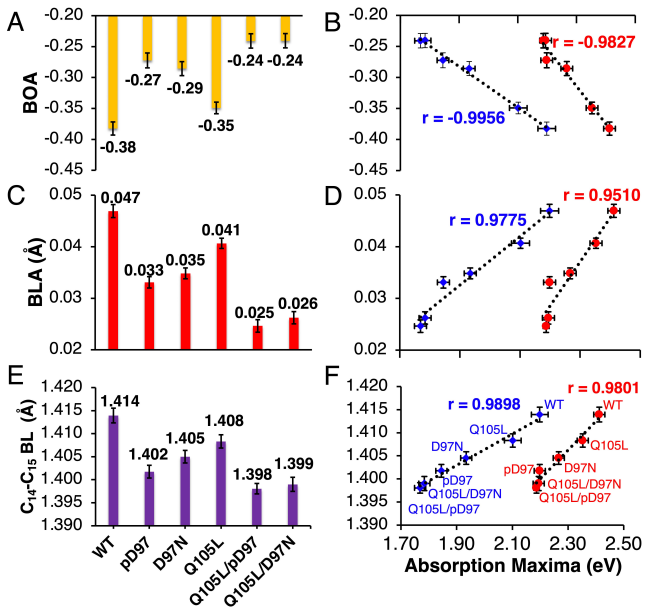

Figure 4. Structural factors obtained as averages from QM/MM–MD trajectories. (A) retinal bond order alternation (BOA) values of each variant and (B) their correlation with the absorption maxima obtained using RI-ADC(2) [PE] (blue) and TD-CAM–B3LYP [PE] (red) along with the Pearson correlation coefficients. (C) retinal bond length alternation (BLA) values of each variant and (D) their correlation with the absorption maxima. (E) Average C_{14}–C_{15} bond lengths (BL) and (F) their correlation with the absorption maxima. Also presented are error bars determined using the Confidence Interval at 95 % for each value.

size.[48] However, the current results capture the qualitative trends while remaining computational feasible.

Neutralizing the counterion D97 by either mutating it to N97 or by protonating it (pD97) under acidic conditions leads to a red shift of 0.17 eV and 0.19–0.22 eV, respectively.[13,17,27] This large shift due to the elimination of a negative charge near the positive Schiff base can be rationalized by the external point charge model of the spectral tuning (Figure 2).[35–38] The presence of the negative D97 next to the protonated Schiff base stabilizes the ground state more than the excited state, leading to a blue shift compared to D97N or pD97. In contrast, the Q105L mutant was measured to cause a red shift of 0.09–0.11 eV, which is the smallest shift among the 5 variants. This is because Q105L does not alter the charge of the binding pocket and due to the central location of Q/L105 with respect to the retinal. This shift is reproduced by all the computational methods. The Q105L mutant in combination with pD97 further enhances the red shift to 0.24–0.26 eV in comparison with the pure effect of pD97 (0.19–0.22 eV). Although the shift between Q105L/pD97 and pD97 is relatively small, it is only reproduced correctly using the [PE]-QM/MM approach. In contrast, the [EE]-QM/MM method leads to a minute blue shift. The absorption maximum of the double mutant Q105L/D97N has not been measured, however, our prediction is that it will create a shift close to Q105/pD97. The ability of the computational methodology used in this study to yield reliable results allowed further investigation into the factors leading to the observed spectral shifts. This includes the BOA, BLA and C_{14}–C_{15} bond length which have been reported to correlate with the absorption maxima.[21,22,30] To gain further insight we analyzed structural factors from retinal geometries used for the spectrum generation (Figure 4A,C,E and Table S3 in Supporting Information).

The WT geometries show the longest C_{14}–C_{15} bond length and the largest BLA and BOA values among the variants studied in this work (Figure 4A,C). The BOA, BLA and C_{14}–C_{15} bond length for WT, Q105L, D97N, pD97, Q105L/D97N and Q105L/pD97 were found to decrease in magnitude between these variants in that order. The calculated absorption maxima were found to decrease (in eV) in the same order (Figure 4B,D,F). The absorption maxima of the variants resulted in high Pearson correlation coefficients ($|r| > 0.95$) for the BOA, BLA and the C_{14}–C_{15} bond length. Such correlations between the retinal structural factors and the predicted absorption maxima were reported by Choongkeun et al. for BPR and Adam et al. for Channelrhodopsin C1C2.[22,30] This high correlation raised the question of whether the change of the bond length, which is induced by the protein environment, is sufficient to explain the spectral shift as discussed by Mao et al.[21] Hence, we recomputed the absorption maxima by omitting the protein environment from the simulation, while keeping the retinal geometry unchanged.

Upon removal of the protein environment, the relative shifts were greatly reduced for all models, as shown in Figure 3B. This indicates that the C_{14}–C_{15} bond length elongation is not the origin of the spectral tuning which is instead controlled by the direct electrostatic interactions between the protein and the retinal chromophore. A similar conclusion was reached for other photoreceptor proteins, where the geometric distortions of the chromophore are secondary to electrostatic effects.[49–51] Work

by Collete et al. on visual rhodopsins also showed that the primary color tuning effect originated from the electrostatic effects of the protein environment on the retinal chromophore, including those with charge neutral mutations.[52] Despite the smaller magnitude of the spectral shifts of our gas-phase models, the trends between the variants are comparable to those in the protein environment. However, the difference between pD97 and D97N as well as between Q105L/pD97 and Q105L/D97N is significantly reduced for the isolated retinal. The red shift for Q105L has also nearly vanished, from 0.06–0.15 eV in the protein to 0.01 eV in the gas phase.

To better understand the protein-chromophore interactions the electrostatic potential of the environment was visualized on the retinal.[53,54] For each BPR variant the protein environment was found to generate a slightly positive electrostatic potential near the β-ionone ring and a negative potential near the protonated Schiff base of the chromophore (Figure 5A,B and Figure S5). WT and Q105L were found to have the largest negative electrostatic potential near the protonated Schiff base (Figure 5A,B). The D97N and Q105L/D97N environments on the other hand produced a smaller negative electrostatic potential near the Schiff base due to the elimination of the negative charge of the counterion (Figure S5).

Protonation of the D97 residue (pD97 and Q105L/pD97) leads to similar electrostatic potentials as the D97N analogs, in line with comparable spectral shift for these variants. These results are also in agreement with work by Shen et al. who found that the counterion in position 97 plays a key role in tuning the excitation energies of BPR.[55] In accordance with the external point charge model, a weaker negative electrostatic interaction near the Schiff base would yield red-shifted absorption maxima.[35,36,56] However, the Q105L mutant does not change the charge in the protein environment. To isolate the effect of the residue at position 105 the electrostatic potential was also generated using only the sidechain of this residue (Figure 5C, D).

This visualization shows that Q105 produces a positive electrostatic potential near the retinal polyene chain near the C_{14}–C_{15} bond. This positive potential vanishes upon mutating Q105L. To rationalize why this position has such a pronounced effect on the absorption maximum, we inspected the difference between S_1 and S_0 electron density of retinal (Figure 5E, F).The electron density difference (EDD) shows the electron redistribution from the β-ionone ring to the Schiff base.[57] The EDD of the WT model has a greater S_1–S_0 difference near the C_{14}–C_{15} atoms. The wave functions of these models were also analyzed by computing the average difference between the electron and hole populations of a fragmented retinal chromophore as outlined in the Experimental Section.[58,59] The WT model was found to have a greater difference between the electron and hole populations near the C_{14}–C_{15} atoms (0.150) compared to the Q105L model (0.129). These results along with the EDD plots of Figure 5E,F reveal that this is due to the positive electrostatic potential from Q105. These results suggest that the electron density in the S_1 state of Q105L is slightly more delocalized than that of the WT model, which explains the red

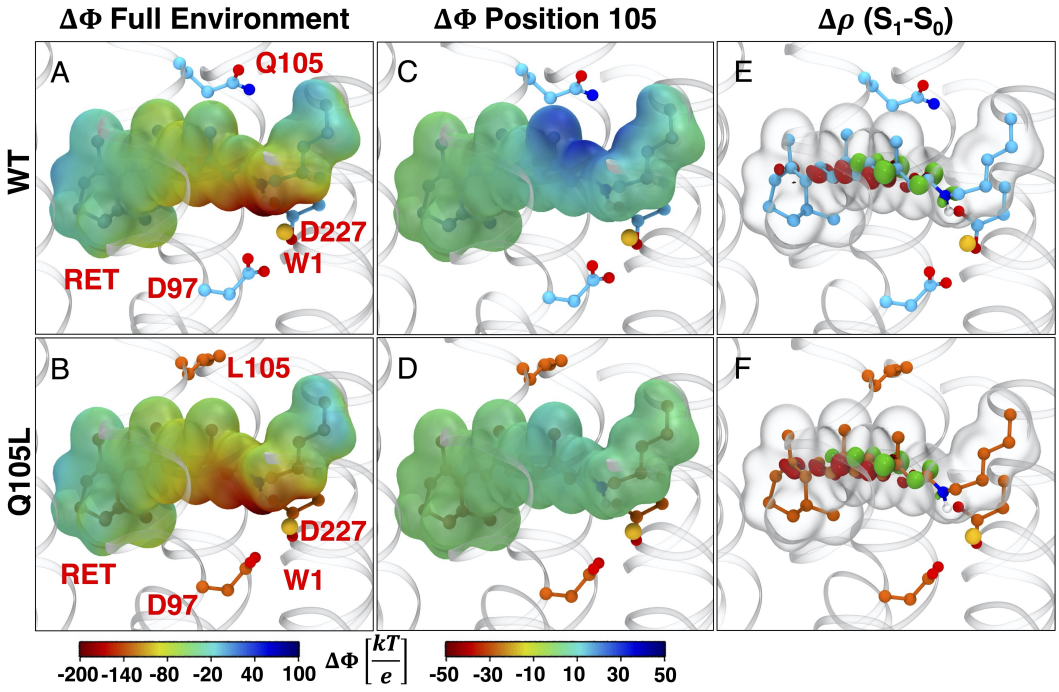

Figure 5. Electrostatic potentials generated for WT and Q105L using APBS in VMD.[53,54] (A,B) The electrostatic potential of the full protein environment on the chromophore was examined. (C,D) The electrostatic potential produced by position 105 was also examined to determine how the Q105L mutation impacts the absorption maximum. (E,F) The S_1–S_0 density difference was calculated to observe the effect of the environment on the charge delocalization upon excitation. The density difference was calculated for the snapshot of each model with the lowest RMSD to the average structure. Here green indicates a location with a net increase in electron density upon excitation and red indicates a net decrease.

shift of the mutant. The positive electrostatic potential from Q105 appears to localize the electron density in the excited state near the C_{14}–C_{15} bond, producing a blue-shift in the WT absorption maxima.

Conclusion

In summary, we have studied the color switching mutation Q105L in BPR by hybrid QM/MM simulations. Our results show that a change in the electrostatic potential from the residue in position 105 near the retinal chromophore is the primary cause of the spectral shift. This is unexpected because the Q105L mutation is charge neutral, but the close proximity of this residue enables Q105 to produce a small positive electrostatic potential near the center of the polyene chain. We believe that this spectral tuning mechanism can be utilized for rational design of retinal proteins with the desired absorption maximum. By exploring the electrostatic maps, one could strategically place polar residues to achieve a desired shift.

Experimental Section

Model Generation

The computational models of this study are based on crystal structures of blue-absorbing Proteorhodopsin (BPR). Two structures were taken from HOT75BPR (PDB code: 4KLY and 4KNF) and one was obtained from Med12BPR (PDB code: 4JQ6).[10] In the subsequent calculations chain (B) of the pentamer of HOT75BPR was chosen because this chain as well as chain C had the highest number of resolved residues for generating a full atomistic model.[10] Missing loops were constructed using the Modeller program by modelling the missing components using both crystal structures 4KLY and 4KNF.[10,60] In both HOT75BPR crystal structures the counterion D97 was mutated to N97, therefore the chain B of the hexameric Med12BPR crystal structure was used as a template for the orientation of the D97 in the reverse mutant N97D. These two HOT75BPR structures and their N97D mutants resulted in four models D97N (based on 4KLY), Q105L/D97N (based on 4KNF), WT, Q105L (Figure S1). The C15=N bond between retinal and lysine in the 4KLY and 4KNF crystal structures is highly contorted, probably due to the lack of D97 counterion. Using these geometries as a starting point for molecular dynamics resulted in a syn conformation instead of anti (Figure S2 B–C). These geometries were replaced with the one from the Med12BPR structure to produce structures with a C15=N bond in an *anti* configuration. Additionally, the D227 residue was taken from the Med12BPR structure. The original D227 residues were located much farther away from the Schiff base due to the contorted nature of the C15=N bond in the original crystal structures (Figure S2). The position of this residue is very important due to that fact that it acts as the primary counterion to the retinal and thus plays a key role in determining the absorption spectrum. Waters were also added based on Dowser predictions, with two notable waters adding to the binding pocket above the retinal near Q/L105 (Figure S1 and S2). The D97 residue in WT and Q105L was also protonated for direct comparison to low pH experimental results, resulting in two additional models pD97 and Q105L/pD97.The PDB2PQR web server was utilized to calculate the pKa value of each titratable residue and determine their protonation state at pH 7 leaving each titratable glutamate and aspartate unprotonated except for E108.[61]

The geometries of the five models were optimized using the hybrid quantum mechanics/molecular mechanics (QM/MM) method with constraints on the backbone and heavy atoms of sidechains outside of the binding pocket. The QM region consisted of the retinal chromophore along with the N/D97, Q/L105, D227, and K231 sidechains cut between C_β-C_α and capped with a hydrogen atom. Nearby water molecules (W1, W2, W3 and if applicable W4) were also included in the QM region. During the optimization process the MM region was composed of the remaining protein and solvent molecules which were further subdivided into relaxed and frozen portions. The relaxed MM region included the binding pocket residues of the chromophore which were defined as the sidechains which have at least one atom within a 6 Å distance to the retinal chromophore. The remaining protein and peptide backbone were fixed in space with 100 $kcal \cdot Å^{-2} \cdot mol^{-1}$ constraints. During the optimization process the QM region was treated at the B3LYP/Def2-SVP level of theory with Grimme's D3 dispersion correction using the TeraChem computational package.[62–65] The MM region was modeled using the AMBER ff14SB classical force field while the water molecules were treated using the TIP3P water model.[66,67]

After optimization, the models were subject to QM/MM molecular dynamics (MD) simulations with the AMBER 16 package.[68,69] During equilibration the QM region was reduced to the retinal chromophore with the linking K231 sidechain. This region was treated using the semi-empirical DFTB2+D method due to the ability of the method to produce a nanosecond MD trajectory efficiently while also providing reliable results.[70] The DFTB method itself has been successfully applied to other retinal systems, such as C1 C2 embedded in a membrane, and benchmarks of the method have shown it can produce results in close agreement to those obtained with the B3LYP functional with medium sized basis sets.[71,72] During the equilibration process each of the models were heated from 0 to 300 K over 0.02 ns with an integration step size of 1 fs with a 10 $kcal/mol \cdot Å^2$ restraint on all heavy atoms. An additional step of equilibration was then run for 0.02 ns at 300 K using the Langevin thermostat with the restraints relaxed for the heavy atoms of sidechains within 6 Å of the retinal chromophore. The equilibration steps were then followed by a 1 ns ground state QM/MM simulation with the same conditions and a time step of 1 fs. During the equilibration and production steps, SHAKE was used in the MM region to constrain the bond lengths of bonds involving hydrogen.

Spectra Generation

UV-Vis absorption spectra were then generated by sampling 100 conformations from the 1 ns QM/MM molecular dynamics (MD) trajectory every 10 ps followed by calculating the excitation energies. The excitation energies were determined with both time-dependent density functional theory (TD-DFT) and the second-order algebraic diagrammatic construction method with the resolution of identity approximation (RI-ADC(2)).[73,74] The TD-DFT excitation energies of each snapshot were determined for the 10 energetically lowest excited states using two embedding schemes for the QM/MM calculations, electrostatic embedding [EE] and polarizable embedding [PE]. The point charges of the environment were generated using the ff14SB AMBER force field.[66] The TD-DFT [EE] calculations were performed using the Orca program package at the CAM–B3LYP/cc-pVDZ level of theory without the use of the Tamm-Dancoff approximation (TDA).[75] The TD-DFT [PE] calculations were performed at the same level of theory with the Dalton program along with the external field effect (EEF) option.[76–79] The protein and solvent in the [EE] calculations were described using the same force field as in the MD simulations whereas potentials

based on atom-centered multipoles (up to and including quadrupoles) and dipole-dipole polarizabilities were used in the [PE] calculations. The multipoles and polarizabilities were derived for each snapshot by using a fragmentation scheme as detailed in Ref. [18]. The individual fragment calculations were performed using CAM–B3LYP together with an ANO-type recontracted 6–31 + G* basis set (loprop-6-31 + G*). The potential for the [PE] calculations were generated using the PyFraME package.[80] The Turbomole package was used to determine the excitation energies of each snapshot at the RI-ADC(2)/cc-pVDZ level of theory.[81] Each set of resulting stick spectra were then broadened assuming a Gaussian band shape and the final spectra were determined by averaging over the 100 conformations at each wavelength (Figures S5–6, Table S1).

Spectroscopic studies have shown that the UV-Vis absorption maximum of BPR in a basic environment (pH 9, 488 nm) does not shift from the value obtained at neutral pH (490 nm).[13] On the other hand in acidic conditions (pH 5, 528 nm) the absorption maxima of the D97N mutated proteins and the WT had similar λmax values, both of which were red shifted compared to the absorption maxima in basic conditions.[17,21,25,27,82] The experimental spectra in basic conditions were used as a reference to the WT and Q105L computed absorption spectra and the spectra obtained under acidic conditions were used as a reference for D97N and Q105L/D97N as well as pD97 and Q105L/pD97.[13] The absorption spectra obtained in this study are presented in Figure S5 and the absorption maxima are tabulated in Table S1 along with their corresponding experimental values. To estimate the direct electrostatic effect of the protein environmental on the absorption maxima, the UV-Vis absorption spectra were also determined by removing the protein from the excitation energy calculations (Figure S6). When neglecting the environment, both Dalton and Orca yielded similar excitation energies which differed at most by 1 nm (Table S1). Additionally, the chromophores and environments of the WT and Q105L models were swapped to observe the direct impact of the environment on the absorption maximum (Table S6).

Structural Factors

The structural factors of the retinal chromophore were determined from the DFTB2 + D trajectories. Here the bond length alternation is the difference between the single and double bonds between carbons of the polyene chain. Likewise the bond order alternation was determined in a similar fashion from the difference in the bond orders of the single and double bonds. The bond orders were generated by the Orca program which uses the Mayer bond order formulism.[83,84]

Electrostatic Potentials

The electrostatic potentials of the protein were visualized by projecting the point charges of the protein environment onto the chromophore. This was done by first averaging the positions of each atom over the DFTB2 + D trajectories and then zeroing the point charges of the retinal and lysine link residues. The electrostatic potential of the environment was then generated using the APBS program and visualized with VMD as shown in Figure 4 and Figure S7.[53,54] To examine the effect of the residue in position 105 a similar procedure was adopted where the point charges of only the Q/L105 residue were used to generate separate electrostatic maps for WT and Q105L mutant. For WT and Q105L, the snapshot of each variant with lowest RMSD from the average was used to calculate the electron density difference (EDD) between S_1 and S_0. The EDD was calculated using the RI-ADC(2) [EE] results from Turbomole and visualized with VMD.[54,81] The TheoDORE package was used to determine the average difference between the electron and hole populations of the retinal chromophore for the WT and Q105L models. The electron-hole analysis was performed using RI-ADC(2) for both embedding schemes along with Löwdin charges and the fragmentation pattern presented in Figure S14.[58,59] The chromophore was partitioned in this way because the lysine link is not conjugated (Fragment 3), the electron density difference figures and electrostatic potential maps of our work along with experimental evidence from Mao et al.[21] show that the primary effect of this mutation is located near the C14-C15 bond (fragment 2) and the remaining chromophore was grouped into fragment 1.

Trajectory Analysis

We have analyzed the structural differences between the QM/MM optimized geometries and the QM(DFTB2 + D)/MM trajectories at 300 K by plotting the root mean square deviations (RMSD) of the heavy atoms relative to their QM/MM optimized positions for both the binding pocket and retinal chromophore (Figure S8–13). The RMSD of each model shows only minor fluctuations over the 1 ns trajectory for both the retinal chromophore and 6 Å binding pocket. These results provide evidence that the models had been properly equilibrated after heating to 300 K and thus would provide reasonable structures when generating the absorption spectra. We also report the excitation energies of the individual snapshots from each trajectory which show that similar to the RMSD values there were only small fluctuations around the average absorption maximum.

Acknowledgements

I.S. thanks the DFG Collaborative Research Center 1078, project C6 for support. I.S. gratefully acknowledges funding by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant No. 678169 "PhotoMutant"). J.R.C thanks the Zuckerman STEM Leadership Program for their support. S.A. thanks the Minerva Stiftung for a post-doctoral fellowship within the framework of the Minerva Fellowship Program. J.M.H.O. thanks VILLUM FONDEN for financial support (grant no. 29478).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: proteorhodopsin · retinal · rhodopsin · QM/MM · spectral tuning

[1] O. Béja, L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich, E. F. DeLong, *Science.* **2000**, *289*, 1902–1906.

[2] O. Béja, E. N. Spudich, J. L. Spudich, M. Leclerc, E. F. DeLong, *Nature* **2001**, *411*, 786–789.
[3] N. U. Frigaard, A. Martinez, T. J. Mincer, E. F. DeLong, *Nature* **2006**, *439*, 847–850.
[4] J. L. Spudich, C. S. Yang, K. H. Jung, E. N. Spudich, *Annu. Rev. Cell Dev. Biol.* **2000**, *16*, 365–392.
[5] A. Rozenberg, K. Inoue, H. Kandori, O. Béjà, *Annu. Rev. Microbiol.* **2021**, *75*, 427–447.
[6] O. M. Finkel, O. Béjà, S. Belkin, *ISME J.* **2013**, *7*, 448–451.
[7] J. A. Fuhrman, M. S. Schwalbach, U. Stingl, *Nat. Rev. Microbiol.* **2008**, *6*, 488–494.
[8] L. Gómez-Consarnau, J. A. Raven, N. M. Levine, L. S. Cutter, D. Wang, B. Seegers, J. Arístegui, J. A. Fuhrman, J. M. Gasol, S. A. Sañudo-Wilhelmy, *Sci. Adv.* **2019**, *5*, 8855–8862.
[9] P. G. Falkowski, *Photosynth. Res.* **1994**, *39*, 235–258.
[10] T. Ran, G. Ozorowski, Y. Gao, O. A. Sineshchekov, W. Wang, J. L. Spudich, H. Luecke, *Acta Crystallogr. Sect. D* **2013**, *69*, 1965–1980.
[11] T. Marinetti, S. Subramaniam, T. Mogi, T. Marti, H. G. Khorana, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 529–533.
[12] J. K. Lanyi, *Annu. Rev. Physiol.* **2004**, *66*, 665–688.
[13] D. Man, W. Wang, G. Sabehi, L. Aravind, A. F. Post, R. Massana, E. N. Spudich, J. L. Spudich, O. Béjà, *EMBO J.* **2003**, *22*, 1725–1731.
[14] T. Friedrich, S. Geibel, R. Kalmbach, I. Chizhov, K. Ataka, J. Heberle, M. Engelhard, E. Bamberg, *J. Mol. Biol.* **2002**, *321*, 821–838.
[15] G. Sabehi, R. Massana, J. P. Bielawski, M. Rosenberg, E. F. Delong, O. Béjà, *Environ. Microbiol.* **2003**, *5*, 842–849.
[16] S. J. Giovannoni, L. Bibbs, J. C. Cho, M. D. Stapels, R. Desiderio, K. L. Vergin, M. S. Rappé, S. Laney, L. J. Wilhelm, H. J. Tripp, E. J. Mathur, D. F. Barofsky, *Nature* **2005**, *438*, 82–85.
[17] W. W. Wang, O. A. Sineshchekov, E. N. Spudich, J. L. Spudich, *J. Biol. Chem.* **2003**, *278*, 33985–33991.
[18] J. J. Amsden, J. M. Kralj, V. B. Bergo, E. N. Spudich, J. L. Spudich, K. J. Rothschild, *Biochemistry* **2008**, *47*, 11490–11498.
[19] N. Atamna-Ismaeel, G. Sabehi, I. Sharon, K. P. Witzel, M. Labrenz, K. Jürgens, T. Barkay, M. Stomp, J. Huisman, O. Beja, *ISME J.* **2008**, *2*, 656–662.
[20] S. Y. Kim, S. A. Waschuk, L. S. Brown, K. H. Jung, *Biochim. Biophys. Acta Bioenerg.* **2008**, *1777*, 504–513.
[21] J. Mao, N. N. Do, F. Scholz, L. Reggie, M. Mehler, A. Lakatos, Y. S. Ong, S. J. Ullrich, L. J. Brown, R. C. D. Brown, J. Becker-Baldus, J. Wachtveitl, C. Glaubitz, *J. Am. Chem. Soc.* **2014**, *136*, 17578–17590.
[22] C. Lee, S. Sekharan, B. Mertz, *J. Phys. Chem. B* **2019**, *123*, 10631–10641.
[23] S. Faramarzi, J. Feng, B. Mertz, *Biophys. J.* **2018**, *115*, 1240–1250.
[24] T. K. Maiti, K. Yamada, K. Inoue, H. Kandori, *Biochemistry* **2012**, *51*, 3198–3204.
[25] Y. Ozaki, T. Kawashima, R. Abe-Yoshizumi, H. Kandori, *Biochemistry* **2014**, *53*, 6032–6040.
[26] J. M. Kralj, E. N. Spudich, J. L. Spudich, K. J. Rothschild, *J. Phys. Chem. B* **2008**, *112*, 11770–11776.
[27] B. R. Kelemen, M. Du, R. B. Jensen, *Biochim. Biophys. Acta Biomembr.* **2003**, *1618*, 25–32.
[28] J. R. Hillebrecht, J. Galan, R. Rangarajan, L. Ramos, K. McCleary, D. E. Ward, J. A. Stuart, R. R. Birge, *Biochemistry* **2006**, *45*, 1579–1590.
[29] R. Rangarajan, J. F. Galan, G. Whited, R. R. Birge, *Biochemistry* **2007**, *46*, 12679–12686.
[30] S. Adam, C. Wiebeler, I. Schapiro, *J. Chem. Theory Comput.* **2021**, *17*, 6302–6313.
[31] M. Wanko, M. Hoffmann, T. Frauenheim, M. Elstner, *J. Comput.-Aided Mol. Des.* **2006**, *20*, 511–518.
[32] R. L. Gieseking, C. Risko, J. L. Brédas, *J. Phys. Chem. Lett.* **2015**, *6*, 2158–2162.
[33] C. H. Choi, M. Kertesz, A. Karpfen, *J. Chem. Phys.* **1997**, *107*, 6712–6721.
[34] D. Jacquemin, C. Adamo, *J. Chem. Theory Comput.* **2011**, *7*, 369–376.
[35] M. G. Motto, M. Sheves, K. Tsujimoto, V. Balogh-Nair, K. Nakanishi, *J. Am. Chem. Soc.* **1980**, *102*, 7947–7949.
[36] K. Nakanishi, V. Balogh-Nair, M. Amaboldi, K. Tsujimoto, B. Honig, *J. Am. Chem. Soc.* **1980**, *102*, 7945–7947.
[37] H. Barry, D. Uri, N. Koji, B. N. Valeria, M. A. Gawinowicz, A. Maria, M. G. Motto, *J. Am. Chem. Soc.* **1979**, *101*, 7084–7086.
[38] S. Mordechai, N. Koji, H. Barry, *J. Am. Chem. Soc.* **1979**, *101*, 7086–7088.
[39] S. Hirschi, D. Kalbermatter, Z. Ucurum, D. Fotiadis, *J. Struct. Biol.* **2020**, *4*, 100024.
[40] V. A. Borin, C. Wiebeler, I. Schapiro, *Faraday Discuss.* **2018**, *207*, 137–152.
[41] E. Krieger, S. B. Nabuurs, G. Vriend, in *Struct. Bioinforma.*, John Wiley and Sons Inc., **2005**, pp. 509–523.
[42] S. Hirschi, D. Kalbermatter, Z. Ucurum, T. Lemmin, D. Fotiadis, *Nat. Commun.* **2021**, *12*, 1–10.
[43] S. Reckel, D. Gottstein, J. Stehle, F. Löhr, M.-K. Verhoefen, M. Takeda, R. Silvers, M. Kainosho, C. Glaubitz, J. Wachtveitl, F. Bernhard, H. Schwalbe, P. Güntert, V. Dötsch, *Angew. Chem. Int. Ed.* **2011**, *5*, 11942–11946; *Angew. Chem.* **2011**, *123*, 12148–12152.
[44] K. Ginalski, L. Rychlewski, *Proteins Struct. Funct. Genet.* **2003**, *53*, 410–417.
[45] K. Ginalski, *Curr. Opin. Struct. Biol.* **2006**, *16*, 172–177.
[46] A. Kryshtafovych, Č. Venclovas, K. Fidelis, J. Moult, *Proteins Struct. Funct. Genet.* **2005**, *61*, 225–236.
[47] E. Kloppmann, T. Becker, G. M. Ullmann, *Proteins Struct. Funct. Genet.* **2005**, *61*, 953–965.
[48] A. D. Laurent, A. Blondel, D. Jacquemin, *Theor. Chem. Acc.* **2015**, *134*, 1–11.
[49] Y. Lahav, D. Noy, I. Schapiro, *Phys. Chem. Chem. Phys.* **2021**, *23*, 6544–6551.
[50] M. Bondanza, L. Cupellini, F. Lipparini, B. Mennucci, *Chem* **2020**, *6*, 187–203.
[51] A. Sirohiwal, F. Neese, D. A. Pantazis, *J. Am. Chem. Soc.* **2020**, *142*, 18174–18190.
[52] F. Collette, T. Renger, F. Müh, M. Schmidt Am Busch, *J. Phys. Chem. B* **2018**, *122*, 4828–4837.
[53] E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G. W. Wei, M. J. Holst, J. A. McCammon, N. A. Baker, *Protein Sci.* **2018**, *27*, 112–128.
[54] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33–38.
[55] C. Shen, X. Jin, W. J. Glover, X. He, *Molecules* **2021**, *26*, 4486–4505.
[56] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227–249.
[57] R. Mathies, L. Stryer, *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 2169–2173.
[58] F. Plasser, *J. Chem. Phys.* **2020**, *152*, 084108.
[59] F. Plasser, H. Lischka, *J. Chem. Theory Comput.* **2012**, *8*, 2777–2789.
[60] A. Šali, T. L. Blundell, *J. Mol. Biol.* **1993**, *234*, 779–815.
[61] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, N. A. Baker, *Nucleic Acids Res.* **2007**, *35*, W522–W525.
[62] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
[63] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
[64] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
[65] S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, S. I. L. Kokkila-Schumacher, N. Luehr, J. W. Snyder, C. Song, A. V. Titov, I. S. Ufimtsev, T. J. Martínez, *J. Chem. Phys.* **2020**, *152*, 224110.
[66] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
[67] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.
[68] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.* **2005**, *26*, 1668–1688.
[69] R. C. Walker, I. F. Crowley, D. A. Case, *J. Comput. Chem.* **2008**, *29*, 1019–1031.
[70] G. D. M. Seabra, R. C. Walker, M. Elstner, D. A. Case, A. E. Roitberg, *J. Phys. Chem. A* **2007**, *111*, 5655–5664.
[71] S. Adam, A. N. Bondar, *PLoS One* **2018**, *13*, e0201298.
[72] M. Gaus, Q. Cui, M. Elstner, *J. Chem. Theory Comput.* **2011**, *7*, 931–948.
[73] E. Runge, E. K. U. Gross, *Phys. Rev. Lett.* **1984**, *52*, 997–1000.
[74] C. Hättig, *Adv. Quantum Chem.* **2005**, *50*, 37–60.
[75] F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1327.
[76] J. M. H. Olsen, S. Reine, O. Vahtras, E. Kjellgren, P. Reinholdt, K. O. Hjorth Dundas, X. Li, J. Cukras, M. Ringholm, E. D. Hedegård, R. Di Remigio, N. H. List, R. Faber, B. N. Cabral Tenorio, R. Bast, T. B. Pedersen, Z. Rinkevicius, S. P. A. Sauer, K. V. Mikkelsen, J. Kongsted, S. Coriani, K. Ruud, T. Helgaker, H. J. A. Jensen, P. Norman, *J. Chem. Phys.* **2020**, *152*, 214115.
[77] C. Steinmann, P. Reinholdt, M. S. Nørby, J. Kongsted, J. M. H. Olsen, *Int. J. Quantum Chem.* **2019**, *119*, 1–20.
[78] K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I. M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V.

Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, H. Ågren, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 269–284.

[79] J. M. H. Olsen, N. H. List, C. Steinmann, A. H. Steindal, M. S. Nørby, P. Reinholdt, *PElib: Polarizable Embedding library (version 1.4.3)*, **2020**. DOI: 10.5281/zenodo.3967017. See https://gitlab.com/pe-software/pelib.

[80] J. M. H. Olsen and contributors, *PyFraME: Python framework for Fragment-based Multiscale Embedding (version 0.3.0)*, **2020**. DOI: 10.5281/zenodo.3820471. See https://gitlab.com/FraME-projects/PyFraME.

[81] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, F. Weigend, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 91–100.

[82] T. Köhler, I. Weber, C. Glaubitz, J. Wachtveitl, *Photochem. Photobiol.* **2017**, *93*, 762–771.

[83] I. Mayer, *J. Comput. Chem.* **2007**, *28*, 204–221.

[84] I. Mayer, *Chem. Phys. Lett.* **1983**, *97*, 270–274.

Manuscript received: January 15, 2022
Accepted manuscript online: March 20, 2022
Version of record online: April 5, 2022