



UNIVERSITÀ
DI SIENA
1240

UNIVERSITA' DEGLI STUDI DI SIENA
DIPARTIMENTO DI BIOTECNOLOGIE, CHIMICA E FARMACIA

DOTTORATO DI RICERCA IN
CHEMICAL AND PHARMACEUTICAL SCIENCES

CICLO XXXVI

COORDINATORE: Prof. Maurizio Taddei

TITOLO DELLA TESI:

QUANTUM CLASSICAL SIMULATIONS OF THE RETINAL CHROMOPHORE
PHOTOISOMERIZATION EVENT IN NEWLY DISCOVERED RHODOPSINS

SETTORE SCIENTIFICO-DISCIPLINARE: Chim/06

DOTTORANDO

Riccardo Palombo

TUTOR

Chiar.mo Prof. Massimo Olivucci

ANNO ACCADEMICO: 2020/2021



UNIVERSITÀ DI SIENA 1240

Dipartimento di Biotecnologie, Chimica e Farmacia

Dottorato in Chemical and Pharmaceutical Sciences

36° Ciclo

I hereby declare that the research presented in the thesis entitled "Quantum classical simulations of the retinal chromophore photoisomerization event in newly discovered rhodopsins", in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the Department of Biotechnology, Chemistry and Pharmacy of the University of Siena is an authentic record of my own work carried out during the period from November 2020 to November 2023 under the supervision of Prof. Massimo Olivucci, in the Laboratory for Computational Photochemistry and Photobiology (LCPP) . The matter presented in this thesis has not been submitted by me for the award of any other degree or diploma of this or any other institute or university.

PhD Candidate: Riccardo Palombo

Principal Investigator: Prof. Massimo Olivucci

Contents

1	Introduction	12
1.1	Objective of the Research	12
1.2	A Brief Rhodopsin Survey	13
1.2.1	Microbial and Animal Rhodopsins	14
1.3	Photochemistry of Rhodopsins	15
1.3.1	Color Tuning	16
1.3.2	Color Tuning in the Studied Systems	18
1.3.3	Photoisomerization Efficiency	18
1.3.4	Origin of the Fluorescence in Microbial Rhodopsin	21
1.4	Neorhodospin, an Unusual Fungal Fluorescent Rhodopsin with Optogenetic Potential	23
1.5	The Discovery of Heliorhopsins	24
2	Methods	27
2.1	Quantum Mechanics methods	27
2.1.1	The Hamiltonian operator	27
2.1.2	The Born-Oppenheimer approximation	29
2.1.3	The Hartree-Fock method	30
2.1.4	The CASSCF method	31
2.1.5	The CASPT2 method	33
2.2	Molecular Mechanics (MM) methods	33
2.3	Quantum Mechanics / Molecular Mechanics Methods	34
2.3.1	The a-ARM protocol	35
2.4	Semiclassical Molecular Dynamics	37
2.4.1	Room Temperature Simulation Protocol	39
3	Results	40
3.1	Red-shift and Fluorescence Enhancement of Neorhodopsin	40
3.1.1	Abstract	40
3.1.2	Introduction	41
3.1.3	Model Construction and Validation	43
3.1.4	Electronic Character of the Absorption and Emission	45

3.1.5	S ₁ relaxation	46
3.1.6	NeoR photoisomerization	47
3.1.7	Fluorescence Enhancement Mechanism	50
3.1.8	Resonance Raman Simulation	52
3.1.9	Conclusions	54
3.2	Photoisomerization Quantum Yield Modulation in the Heliorhodopsin Family	56
3.2.1	Abstract	56
3.2.2	Introduction	57
3.2.3	Model Construction and Validation	58
3.2.4	Excited State Dynamics and TAS Assignment	59
3.2.5	Aborted Bicycle-Pedal Isomerization Motion	62
3.2.6	Quantum Yield Modulation	63
3.2.7	Conclusions	64
4	Conclusions and Outlook	68
5	Articles	70
5.1	PAPER 1: Retinal Chromophore Charge Delocalization and Confinement explain the Extreme Photophysics of Neorhodopsin	70
5.2	PAPER 2: Experimental Assessment of the Electronic and Geometrical Structure of a Near-Infrared Absorbing and Highly Fluorescent Microbial Rhodopsin	80
5.3	PAPER 3: Picosecond Quantum-classical Dynamics reveals that the Coexistence of Light-induced Microbial and Animal Chromophore Rotary Motion Modulates the Isomerization Quantum Yield of Heliorhodopsin	86
6	Appendix	101

Acronyms

$E_{S_1}^f$ excited state photoisomerization energy barrier.	GFP green fluorescent proteins.
E_r reorganization energy.	GPCR G-protein-coupled receptors.
$\Delta E_{S_1-S_0}^a$ vertical excitation energy.	HBN hydrogen bond network.
f_{osc} oscillator strength.	HeR heliorhodopsin.
λ_{max}^a maximum absorption wavelength.	HLA hydrogen link atom.
λ_{max}^e maximum emission wavelength.	LA light-adapted.
ϕ_{fluo} fluorescence quantum yield.	MAE mean absolute error.
ϕ_{iso} photoisomerization quantum yield.	MD molecular dynamics.
ARM automatic rhodopsin modelling.	MM molecular mechanics.
ASR anabaena sensory rhodopsin.	MO molecular orbital.
BLA bond length alternation.	NeoR neorhodopsin.
BOA Born-Oppenheimer approximation.	PES potential energy surface.
BR bacteriorhodopsin.	QM quantum mechanics.
CAS complete active space.	RASSCF restricted active space self consistent field.
CASSCF complete active space self consistent field.	RGC rhodopsin-guanlyl cyclase.
CCW counterclockwise.	rPSB retinal protonated Schiff base.
ChR channelrhodopsin.	RR resonance Raman.
CI configuration interaction.	SCF self consistent field.
CoIn conical intersection.	SCMD semiclassical molecular dynamics.
CT charge transfer.	SD Slater determinant.
CW clockwise.	TaHeR <i>thermoplasmatales archaeon</i> rhodopsin.
DA dark-adapted.	TAS transient absorption spectroscopy.
FC Franck-Condon.	
FCI full configurational interaction.	
FS fluorescent state.	

List of Figures

1.1	General aspects of microbial and animal rhodopsins. A) Overview of, left, the bacteriorhodopsin structure (PDB ID:6G7H) and, right, the bovine rhodopsin structure (PDB ID:6G7H). rPSB chromophores in black. B) rPSB photoisomerization diversity between animal and microbial rhodopsins. Atomic numbers are given for the isomerizing double bonds. Red arrows indicate the directionality of the isomerization, counterclockwise (CCW) in animal and clockwise (CW) in microbial.	14
1.2	Figure adapted from [1] illustrating a pictorial overview of the phylogeny of known microbial rhodopsins.	15
1.3	Color tuning exemplified by the two charge model with respect to the all- <i>trans</i> rPSB chromophore. (A) rPSB photon excitation when a negatively charged counterion is in proximity of the Schiff moiety. (B) rPSB photon excitation in the absence of a negatively charged counterion. (C) rPSB photon excitation when the rPSB chromophore counterion is located in a region halfway the polyene chain.(D) Effects of A-C to the vertical excitation energy ($\Delta E_{S_1-S_0}^a$), yielding a spectral blue-shift (case A) or red-shift (case C).	16
1.4	Illustration of the rPSB photoisomerization reaction in (A) animal and (B) microbial rhodopsins. In animal rhodopsins the isomerization occurs in ultrafast timescales through a barrier-less energy profile, contrasting with most microbial rhodopsins where longer excited state lifetimes suggest the presence of an isomerization barrier on S_1 ($E_{S_1}^f$). Vertical arrows denote the photon absorption initiating the photoisomerization. Conical intersection (CoIn) represent degenerate points between S_0 and S_1 states. From the CoIn, two reaction paths can be followed: a reactive one leading to photoproduct formation (red arrow) and an unreactive one leading to reactant-back formation, the latter also called internal conversion (blue arrow). (C-D) Origin of the fluorescence in microbial rhodopsin due to the presence of a S_1 barrier impairing the photoisomerization. (C) A shallow S_1 surface or a barrier may be generated via S_1/S_2 mixing (see reference [2]). (D) Alternatively, a barrier can be generated via coupling of the S_1 and S_0 state, as proposed in this thesis for NeoR. rPSB resonance formulas are also given.	19

1.5	S_1 isomerization channels in TaHeR and Rh. A) rPSB chromophore in TaHeR (left) and Rh (right). The moiety at the bottom represents the reactive part of the two chromophores. The curly arrows indicate the S_1 reaction coordinates represented by the α and α' coordinates associated with the rotations around C11=C12 and C13=C14 double bonds, respectively. B) Schematic representation of the photoisomerization channels documented in TaHeR (left) and Rh (right). The arrows in the FC region indicate the S_1 dynamics that, in TaHeR, but not in Rh, leads to an ultrafast breaking (at BF) of the photoexcited population in two subpopulations reaching two different CoIn ₁₁ and CoIn ₁₃ isomerization decay channels. As documented in Section 3.2.4 , TaHeR CoIn ₁₁ channel was found completely unreactive since leading to internal conversion exclusively.	20
1.6	Comparison between GFP (left) and NeoR (right) fluorescence emission brightness. Adapted from [3].	23
1.7	Schematic representation of membrane topologies of microbial, animal and heliorhodopsins. The presentation of microbial rhodopsin is based on BR (PDB ID:6G7H), and the animal rhodopsin is based on Rh (PDB ID: 1U19). C and N amino termini facing the extracellular side (EC) are also reported to show the inverted orientation of heliorhodopsins. α -helices are denoted using A to G cylinders.	25
2.1	From left to right: type of orbitals in CAS/RAS partition, CAS partition, RAS partitions	32
2.2	Simple illustration of the different contributions to the total MM energy . . .	34
2.3	Overview of the a -ARM model of Neorhodopsin (NeoR) . In gray, atoms fixed at the homology model position. In red, relaxed (MM level) atoms composing the rPSB chromophore binding cavity. In blue, relaxed (MM level) atoms of the lysine linker residue (Lys 266) covalently bound to the rPSB chromophore. In green, relaxed atoms (QM level) of the rPSB chromophore. The frontier between QM and MM partitions is also shown.	36

- 3.1 Geometrical and electronic structural changes in NeoR photoisomerization. A) Schematic representation of the hypothetical S_0 and S_1 energy changes occurring along the S_1 relaxation that involves the bond length alternation (BLA, quantified by the difference between the average of the double-bond lengths and the average of the single-bond lengths of the conjugated chain) and isomerization (α) coordinates. The rPSB resonance formula show a delocalized positive charge at the S_0 and S_1 energy minima corresponding to the dark adapted (DA) and fluorescent (FS) states, respectively. The symbol “ $\delta+$ ” gives a qualitative measure of the amount of positive charge located along the rPSB conjugated chain. B) Representation of the bond length alternation (BLA) mode and the α mode along the C13=C14 double bond. BLA_{PSB} is the -C14-C15 and C15-N bond lengths difference. 41
- 3.2 Choice of the NeoR chromophore counterion and model validation. A) Overview of the structure of the all-*trans* rPSB chromophore (orange) and its four potential residue counterions (in green). The lysine residue (in green) bounded to the rPSB chromophore is also displayed. B) Computed (CASPT2 level) λ_{max}^a , λ_{max}^e and E^r of NeoR with varying counterion choices. C) Correlation between experimental (Obs. $\Delta E_{S_0-S_1}$) and computed (Comp. $\Delta E_{S_0-S_1}$) values of vertical excitation energies defining λ_{max}^a in the wild type (indicated as WT) and a set of mutants. D) Superimposition of experimental and computed (dotted line, protocol adopted for band simulation given in **Section 2.4.1**) absorption band of wild type NeoR. The experimental band has been digitalized from the corresponding ref. [4] 44
- 3.3 Electronic character variation along the S_1 relaxation of the rPSB chromophore in NeoR. A) Representation of the two limiting resonance formulas adopted to describe the electronic character of the rPSB chromophore. B) Electron density variation ($\delta\rho_{abs}$) characterizing the vertical $S_0 \rightarrow S_1$ transition from the dark-adapted (DA). Blue and red clouds correspond to electron density decrease and increase respectively. Isovalue set to 0.002 a.u. The associated resonance formulas correspond to resonance hybrids also anticipated in **Figure 3.1**. Blue bubbles represent the QM positive charge (in e unites). Only absolute values > 0.05 e are reported. As indicated by the red box, the total charge residing in the -C14-C15-N-C ϵ - rPSB fragment is also given. C) Same data for the $S_1 \rightarrow S_0$ emission from the fluorescent state (FS). 46

- 3.4 NeoR geometrical character variation along the S_1 relaxation of the rPSB chromophore. A) Bond lengths of the polyene chain of the rPSB chromophore. The open triangles and circles refer to the DA and FS structures, respectively. Bottom, the bond length differences $\Delta d = d_{S_{DA}} - d_{S_{FS}}$ (green) representing the displacements of the stretching coordinates found in the FS. The red and blue colors refer to double and single bonds, respectively. B) Same as A) but using a molecular view. The arrows indicate the dominant geometrical change corresponding, clearly, to a variation in the bond length alternation (BLA, see definition in the caption of **Section 3.1**) in a region of the conjugated chain distant from the Schiff base moiety. The relevant bond lengths are given in A. 47
- 3.5 Energy, charge and BLA profiles along the photoisomerization of, from left to right, C13=C14, C11=C12, C9=C10 and C7=C8 double bonds in NeoR. A) CASPT2 S_1 energy profiles computed in presence of the full protein environment (orange squares), after setting to zero the charges of the entire protein (gold squares), after setting to zero only the charges of the E141 and after removing the whole protein in a full QM calculation (i.e., in vacuo, empty squares). B) Evolution of S_0 and S_1 Charge_{PSB} (the charge residing in the C14-C15-N-C ϵ rPSB moiety) and evolution of BLA_{PSB} (C14-C15 and C15-N bond lengths difference, see **Figure 3.1B**). 48
- 3.6 Origin of the $E_{f_{S_1}}$ barrier in NeoR. A) Representation of the increase in the distance between the negative and positive charge centroids due to the positive charge confinement along the C13=C14 and C9=C10 photoisomerization paths. Comparison between electrostatic potential (ESP) maps indicates that along the C9=C10 coordinate the extent of the confinement is less pronounced being the charge at the CoIn spread on a longer rPSB chromophore moiety (i.e., C10-C11-C12- C13-C14-C15-N). B) Proposed origin of the isomerization barrier in terms of COV energy (HCOV, dashed line) destabilization due to the charge confinement resulting from the mixed [CT] - c'[COV] to the pure [COV] electronic structure change along the S_1 adiabatic energy profile (in orange). The left and right panels display the shape of the S_1 and S_0 adiabatic potential energy curves along the BLA_{PSB} coordinate (C14-C15 and C15-N bond lengths difference, see **Figure 3.1B**) at FS (left) and CoIn (right). The BLA_{PSB} and α coordinates are substantially orthogonal. 51

- 3.7 Calculated and experimental resonance Raman (RR) spectrum in H₂O. The calculated spectrum was generated using a gaussian band shapes of the vibrational transition derived from the QM partition of the NeoR QM/MM model (green trace). The spectrum also includes the line presentation of the vibrational transitions. Frequencies were scaled by a factor of 0.9. Modes with major C=C and C-C stretching characters (factoring to >50% of the contribution of a coordinate) are indicated by black and gray letters, respectively. The notation follows the numbering of the structural formula on top of the figure. Details of the calculated rPSB modes are listed in **Appendix**. 53
- 3.8 TaHeR and Rh models validation. A) Comparison between computed (in orange) and observed (in blue) λ_{max}^a values, calculated as $\Delta E_{S_0-S_1}$ for TaHeR and a set of mutants. B) Same as point A for Rh and related mutants. Comparison between computed (in orange) and observed (in blue) absorption bands (in arbitrary units) for C) TaHeR and D) Rh. The experimental bands were digitalized from literature data [5, 6]. 58
- 3.9 Overview of the S₁ dynamics of A) TaHeR and B) Rh, simulated with 200 quantum-classical trajectories. Unreactive S₁ trajectories are marked by blue symbols placed at the time of the S₁/S₀ decay (hopping time). Similarly, reactive trajectories are marked by red symbols. As described in the main text an explicated in **Figure 1.5**, a trajectory is termed as reactive when, after the hopping, continue to propagate on S₀ in the direction of the photoproduct. In contrast, unreactive trajectories revert their direction and propagate in the direction of the reactant. As shown in **Figure 1.5** and here in the molecular sketches, α and α' indicate the dihedral angle associated to the isomerizing double bond, C11=C12 and C13=C14 in TaHeR and C11=C12 in Rh. 60
- 3.10 Minimum energy path computed for TaHeR describing A) the clockwise (CW) photoisomerization around the C13=C14 double bond (α' coordinate) and B) the counterclockwise (CCW) photoisomerization around the C11=C12 double bond (α' coordinate). Each profile shows S₀, S₁ and S₂ energy profiles computed at the 3 root state average CASPT2/CASSCF(12,12)/6-31G*/AMBER94 energy profiles. On top, view of the two located minimum energy CoIn's displaying the relevant torsional deformations. To notice how both reaction path are associated to shallow S₁ profiles. 61
- 3.11 Time evolution of the aborted bicycle-pedal photoisomerization coordinates (ABP) of TaHeR. A) Representation of the evolution of C12-C13=C14-C15 (i.e. α' , in blue) and C14-C15=N-Ce (in gray) dihedral angles for trajectories along the CoIn₁₃ channel aligned at the S₁ → S₀ decay time (t = 0 fs). B) Representation of the evolution of C10-C11=C12-C13 (i.e. α , in blue) and C8-C9=C10-C11 (in gray) dihedral angles for trajectories along the CoIn₁₁ channel aligned at the S₁ → S₀ decay time (t = 0 fs). 62

- 3.12 Schematic representation of α , β and τ for the C11=C12 moiety of rPSB chromophore. The same representations apply to α' and β' and the corresponding τ' parameter for the C13=C14 bond moiety. τ and τ' are geometrical parameter proportional to the π -overlap (roughly, the dihedral between the axes a and a' of the p-orbitals forming the π -bond) of the corresponding double bond. 63
- 3.13 Reactivity analysis of TaHeR trajectory sets. A) Time evolution of the overlap index $d\tau'/dt$ towards the CoIn₁₃ channel. The top panel shows the correlation between the distribution of reactive decays (red circles) and the $d\tau'/dt$ sign. The bottom panel displays the same data but with trajectories aligned at the $S_1 \rightarrow S_0$ decay time ($t = 0$ fs) and starting 40 fs before the decay. Gray data refers to the S_1 while red and blue data refers to S_0 depending on whether the trajectory is reactive (red) or unreactive (blue). The yellow curve represents the average value before the $S_1 \rightarrow S_0$ decay. The displayed numerical value refers to the percentage of trajectories decaying with $d\tau'/dt > 0$. B) Same as A for the CCW motion leading to the CoIn₁₁ channel. Notice that now the displayed numerical value refers to $d\tau'/dt < 0$. C) Pictorial illustration of the τ' velocity in terms of decreasing overlap along the CW progression towards the CoIn₁₃ channel. The curly arrow represents the rotational velocity of the orbital axes. D) Same as C for the CoIn₁₁ channel where, in contrast, the velocity of the CCW overlap decrease is decelerating during the progression towards the decay channel. 66
- 3.14 Quantum efficiency quenching mechanism in TaHeR. A) Population time evolution (trajectories aligned at the $S_1 \rightarrow S_0$ decay time, $t = 0$ fs) of the closest the distances between the rPSB chromophore atoms characterizing the CW rotation of the bicycle-pedal plane of the CoIn₁₁ channel with neighboring amino acids indicated below. The vertical set of black dots indicate the distribution of distances at the $S_1 \rightarrow S_0$ hopping points. Below, the definition and color coding of the distances monitored in the top. The rotating bicycle pedal plane is also defined pictorially and within a snapshot of the reacting rPSB region at the $S_1 \rightarrow S_0$ decay time ($t = 0$ fs). B) Same as A for the CoIn₁₃ channel. 67

Chapter 1

Introduction

1.1 Objective of the Research

The research presented and discussed in this thesis aims to enhance our understanding of how the spectroscopy and photochemical reactivity of conjugated double-bonds can be controlled at the supramolecular level (i.e. by incorporating the substrate in a specific macromolecular environment). This is achieved through the study of two recently discovered natural light-responsive proteins within the rhodopsin family. In fact, rhodopsins, together with phytochromes and xanthopsins, utilize an "elementary" light-induced double-bond isomerization reaction to catch, store, and subsequently leverage, the energy of light to trigger a multitude of biological functions across all the kingdoms of life [7]. The research objective is, therefore, the mechanistic investigation of the photoisomerization taking place inside two different protein environments. This objective represents a meaningful goal for two main reasons. Firstly, it could provide novel insights into how nature exploits the chemistry of photoisomerization to incorporate light-sensitivity in proteins. Secondly, it could serve as a source of inspiration for bio-inspired technological applications, as the ability, invariably linked to this reaction, of these proteins to sense and respond to light has broad applications in various technological fields, such as optogenetics [8] or synthetic biology [9].

More specifically, this thesis reports on the computational investigation of two rhodopsins called neorhodopsin (NeoR) and *thermoplasmatales archaeon* rhodopsin (TaHeR). NeoR was discovered in 2020 [4] in the fungi spores produced by certain members of the phylum *chytridiomycota* and, as shown in **Section 1.4**, exhibits a set of extreme spectral properties that are particularly desirable in optogenetic tools. On the other hand, as described in **Section 1.5**, TaHeR is a rhodopsin expressed in the archaea *thermoplasmatales archaeon* and belongs to the novel and elusive heliorhodopsin (HeR) family, only discovered in 2018 [10, 6]. For both proteins, an *in silico* study of the photoisomerization event was carried out through the construction of two analogous hybrid quantum mechanics / molecular mechanics (QM/MM) models.

1.2 A Brief Rhodopsin Survey

As anticipated, rhodopsins, which are also called retinal proteins, represent a class of transmembrane proteins constituted by 7 α helices that are capable to sense and respond to light. The sensitivity to light owes to the presence of a retinal protonated Schiff base (rPSB) chromophore, an aldehyde of vitamin A, covalently linked to the protein via an iminium bond with the terminal amine group of a lysine residue, commonly referred as "lysine linker" (**Figure 1.1**).

Across the approximately 500.000 known rhodopsin members, the rPSB chromophore is encased in cavities with different amino acidic compositions and shapes but, despite such chemical environmental diversity, it undergoes the same kind of process in response to light stimulus: energy transduction into an intramolecular motion culminating in the isomerization of a specific π -bond [11]. This primary event triggers the cyclic formation of a series of protein intermediates, the photocycle, that enables repetitive photoactivations of the protein. Each photocycle mediates various biological functions, including bright and dim-light vision in vertebrates [1], light-sensing, phototaxis, photoprotection, ion-pumping in eubacteria, archaea and viruses [1, 12, 13]; light-oriented growth in plants [14]; ion-gating in unicellular algae [15] and catalytic activity in fungi, green algae and chanoflagellates [3, 13], just to cite a few.

The ubiquitous natural distribution of rhodopsins attests their great efficiency and versatility. Over the years, they have served as a endless source of interest, as well as inspiration, for scientists working in different fields such as photochemistry, photobiology, and biophysics. One notable biophysical application discussed in this thesis is exemplified by the optogenetics (see **Section 1.4**). Nevertheless, our comprehension of the photoactivation process critically depend on the availability of reliable structural information before, in the unphotolyzed DA state, and after, in the light-adapted (LA), the reaction of photoisomerization. In this regard, several effort have been made to develop experimental techniques for elucidating the crystallographic structure of rhodopsins. However, the Protein Data Bank (PDB) [16] still contains a restricted number of X-ray structures for rhodopsins from different organisms since crystallizing membrane proteins presents significant challenges compared to soluble proteins [17]. In order to overcome the problem and to generate resolved structure of rhodopsins that are not yet attainable with the current experimental technologies, it is mandatory to leverage computational methods allowing their *in silico* design through comparative modelling. This involves using rhodopsins with similar sequences and available X-ray structures as templates for homology modelling predictions. The case of NeoR, presented in this thesis, illustrates this situation. Given the absence of a crystallographic structure, a comparative homology modeling approach was employed to predict the structure of DA, used for constructing the QM/MM model and simulating the photoisomerization event.

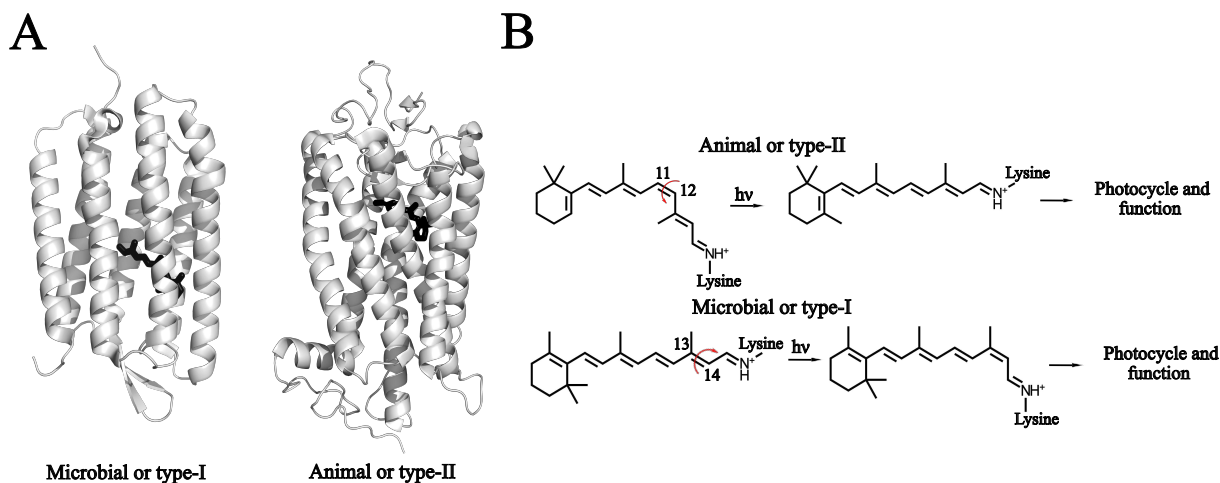


Figure 1.1: General aspects of microbial and animal rhodopsins. A) Overview of, left, the bacteriorhodopsin structure (PDB ID:6G7H) and, right, the bovine rhodopsin structure (PDB ID:6G7H). rPSB chromophores in black. B) rPSB photoisomerization diversity between animal and microbial rhodopsins. Atomic numbers are given for the isomerizing double bonds. Red arrows indicate the directionality of the isomerization, counterclockwise (CCW) in animal and clockwise (CW) in microbial.

1.2.1 Microbial and Animal Rhodopsins

Before the discovery of HeRs in 2018 (see **Section 1.5**) [10], rhodopsins were generally classified in two superfamilies; type-I, or microbial rhodopsins and type-II, or animal rhodopsins [1]. Notably, despite sharing the same protein architecture (see **Figure 1.1A**), animal and microbial rhodopsins feature no detectable sequence identity. This peculiarity has been attributed to either a deep divergence between two evolutionarily related families of transmembrane proteins or to convergent evolution [18, 19]. One difference lies in the configuration of the rPSB chromophore in the unphotolyzed DA states. This difference is also manifested in different photoisomerization stereo- and regiochemistry; while the photocycle in microbial rhodopsins is activated by the clockwise (CW) all-*trans* to 13-*cis* conversion, in animal rhodopsin it is activated by the counterclockwise (CCW) 11-*cis* to all-*trans* conversion [11].

Microbial rhodopsins were initially discovered in the archaea *halobacterium salinarum* and were thus initially epitomized as haloarchaeal rhodopsins. More specifically, *halobacterium salinarum* was found to express up to four retinal proteins; a light-driven outward H^+ pump [20], a light-driven inward Cl^- [21] and sensory rhodopsins I and II (SRI and SRII), acting as positive and negative phototaxis sensors [22, 23], respectively. However, over the past 20 years, since thousands of related retinal proteins with similar or different functions have been identified in all domains of cellular life of microorganisms and even viruses, the term “microbial rhodopsins” has become predominant (see **Figure 1.2**).

One notable group of microbial rhodopsins is called channelrhodopsin (ChR), discovered in green algae, where they function as light-gated cation channels capable to depolarize the

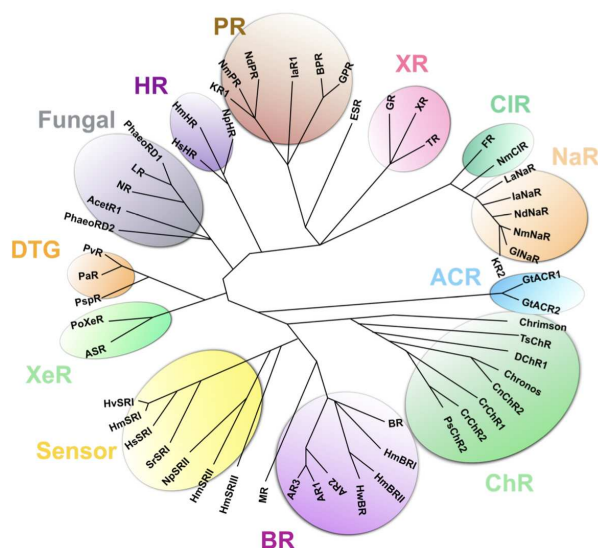


Figure 1.2: Figure adapted from [1] illustrating a pictorial overview of the phylogeny of known microbial rhodopsins.

plasma membrane upon light absorption [24, 25]. The discovery of ChR led to the emergence of a new field, the optogenetics (**Section 1.4**), wherein light-gated ion channels and light-driven ion pumps are used to depolarize and hyperpolarize the axons of cell circuits [8].

While microbial rhodopsins are employed in microorganisms for light-dependent metabolic functions exclusively, animal rhodopsins serve as G-protein-coupled receptors (GPCR) in the eyes and other organs in both vertebrates and invertebrates where they perform both daylight and nocturnal vision, as well as non-visual functions, such as body color change, seasonal reproduction, circadian rhythms, pupillary constriction [1].

1.3 Photochemistry of Rhodopsins

The diversity of the rhodopsin functions is now evident. This diversity is mirrored in distinct amino acid compositions of the binding cavities that host the rPSB photoisomerization reaction that, in fact, displays significant variations in photophysical and photochemical properties [11]. These variations can be considered as the by-product of the action of natural evolution that has shaped the rhodopsin sequences to achieve adaptation in different biological contexts. One example is the observed increase in photoisomerization quantum yield (ϕ_{iso}) found in certain rhodopsin cavities. In solution, for instance, the all-*trans* to 13-*cis* and the 11-*cis* to all-*trans* photo-conversions is slow (2-10 ps and 4 ps) and poorly efficient (0.1-0.3 and 0.2 ϕ_{iso} values) [26, 27, 28, 29]. In contrast, they become extremely fast (0.5 ps and 0.1 ps) and efficient (0.6 and 0.7 ϕ_{iso} values) when occurring inside the binding cavities of bacteriorhodopsin (BR) [30] and bovine rhodopsin [5], respectively. This underscores the crucial role of the chemical environment provided by the protein cavity in facilitating

the isomerization process, leading to improved speed and efficiency and ultimately achieving enhanced photo-sensitivity by enabling and "easier access" to rhodopsin photocycle.

In the paragraphs below, a general overview will be provided for the key concepts beyond two of the most studied properties of the rPSB chromophore inside rhodopsins: the maximum absorption wavelength (λ_{max}^a) (Section 1.3.1) and ϕ_{iso} (Section 1.3.3).

1.3.1 Color Tuning

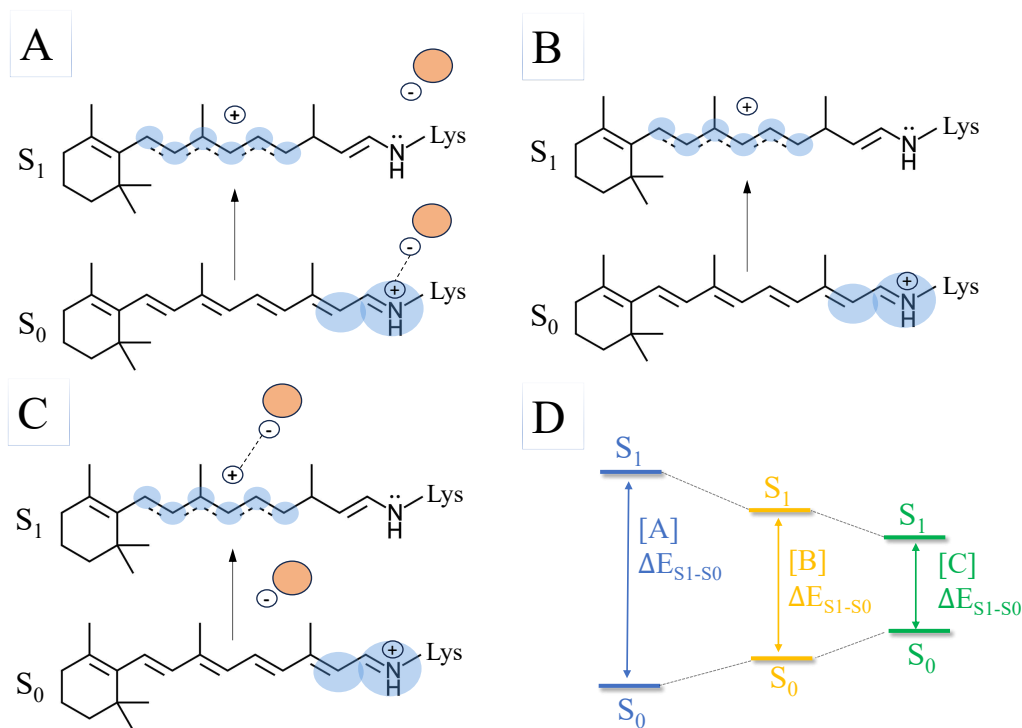


Figure 1.3: Color tuning exemplified by the two charge model with respect to the all-*trans* rPSB chromophore. (A) rPSB photon excitation when a negatively charged counterion is in proximity of the Schiff moiety. (B) rPSB photon excitation in the absence of a negatively charged counterion. (C) rPSB photon excitation when the rPSB chromophore counterion is located in a region halfway the polyene chain. (D) Effects of A-C to the vertical excitation energy ($\Delta E_{S_1-S_0}^a$), yielding a spectral blue-shift (case A) or red-shift (case C).

As described in the preceding paragraph, the rPSB photoisomerization serves as gateway step for the activation of the rhodopsin photocycle and the corresponding function. As the term implies, this reaction requires a light stimulus to occur. In the dark, rhodopsins generally remain inactivated since the thermal isomerization has an high activation energy barrier, exceeding 20 kcal mol^{-1} [31, 32], for instance, in animal rhodopsins. Upon illumination,

however, the rPSB chromophore can be promoted to the excited state, and the reaction, as described in **Section 1.3.3**, follows a different, generally energy barrier-less, reaction path.

Hence, the "spark" is represented by the absorption of a photon, and the maximum likelihood for this event occurs when the wavelength of the incoming photon corresponds to the λ_{max}^a of the rhodopsin. This last aspect is critical since organisms expressing rhodopsins occupy different ecological niches with different illumination levels and spectral windows. The way in which rhodopsins tune the λ_{max}^a is normally referred as spectral or color tuning. In the animal family, a remarkable example of color tuning is provided by the rhodopsins expressed in the retina of the cottoid fishes of lake Baikal (Siberia) where, for increasing habitat depths, rhodopsins have accumulated point mutations for tuning the λ_{max}^a toward increasing blue-shift values [33, 31]. In haloarchaea, the chloride-pumps rhodopsins possess λ_{max}^a values in the yellow, optimal for capturing the photons of light which can penetrate the surface of salt ponds where they live. Overall, the observed rhodopsin spectral range of the λ_{max}^a of the rPSB chromophore varies from the 447 nm (most blue-shifted) found in the deep-sea fish *silver spinyfin* [34] to the 690 nm (most red-shifted) found in NeoR [4].

In general, however, rhodopsins feature a red-shifted λ_{max}^a with respect to the rPSB in solution ($\lambda_{max}^a=440$ nm), as a consequence of the interactions between the cavity and the rPSB chromophore. The phenomenon is called "opsin shift" and depends on how cavity environment modulates the vertical excitation energy ($\Delta E_{S_1-S_0}^a$) difference between the ground (i.e. S_0) and excited (i.e. S_1) electronic states involved in the electronic transition. The most important factors influencing $\Delta E_{S_1-S_0}^a$ are i) the distortion of the polyene chain of rPSB, and ii) the electrostatic interactions between rPSB and the surrounding amino acids [35].

Concerning i), a more planar polyene chain results in an higher degree of conjugation and π delocalization, resulting in a red shifted λ_{max}^a . Conversely, twisting along the single bond decreases the conjugation, leading to a blue shifted λ_{max}^a [36]. Regarding ii), a crucial factor involves the interaction with the rPSB chromophore counterions [37, 38, 39]. In fact, the stability of the positively charged rPSB chromophore in the hydrophobic environment provided by the cavity is ensured by the presence in close proximity of an negatively charged residue (aspartate or glutamate). The most important contribution for the comprehension of the counterion effect to the opsin shift is provided by the external two-point charge model proposed by Nakanishi and coworkers [40]. In order to understand the model, one has to consider that in S_0 most of the positive charge of the rPSB chromophore resides in the Schiff base moiety. The absorption of a photon triggers an instantaneous charge transfer (CT) process, causing the rPSB chromophore to display a more delocalized charge distribution toward the β -ionone ring. Within this picture, a the negative charge placed near the Schiff base would result in S_0 stabilization (see **Figure 1.3A**) and an increase in $\Delta E_{S_1-S_0}^a$ (blue-shift). Instead, a negative charge placed halfway the polyene chain (see **Figure 1.3B**) would result in a greater S_1 stabilization with respect to S_0 , and therefore a decrease in $\Delta E_{S_1-S_0}^a$ (red-shift).

1.3.2 Color Tuning in the Studied Systems

While this two-point charge model is intuitive, it is rather simplistic. In many real systems, including those studied in this work, the electrostatic interaction projected by the protein cavity onto the rPSB chromophore is much more intricate. This complexity arises from factors such as variable and possibly larger number of charged residues, including multiple rPSB counterions and positively charged residues, and also complex hydrogen bond network (HBN)s providing complex dipole orientations toward rPSB.

This complexity is exemplified in the rhodopsin systems studied in this work. In NeoR, an at least 4 different counterion residues, namely E136, E141, E262 and D140, are found closely interacting with the chromophore. Moreover, as discussed in **Section 3.1**, the simplicity of the two charge models also affects the treatment of the electronic character of S_0 , which is conventionally described by a covalent-like resonance formula with most of the positive charge located on the Schiff base. However, it will be shown that this description depends on the electrostatic environment, as different counterion configurations yield different charge distributions in S_0 . Furthermore, it will also be shown that the NeoR models were capable to replicate the experimentally observed behaviour only when the counterion (i.e. E141) was set far from the Schiff base, and not closely interacting with it. This scenario closely resembles an alternative spectral tuning model proposed by the group of Borhan and Geiger, featuring in a rhodopsin model with no counterions in proximity of the Schiff base moiety, but an evenly distributed electrostatic potential embedding [35, 41]. The absence of a rPSB counterion yields a conjugated system that is more responsive to subtle changes in its environment due to residue dipoles.

In the TaHeR model discussed in **Section 3.2**, the protonation state of two adjacent histidines residues, namely H23 and H82, close to the Schiff base, exerted a great impact on the computation of the spectral properties. In fact, this amino acid can display a charge of +1 when both the δ -nitrogen and ϵ -nitrogen of the imidazole ring are protonated (Hip, [HN δ -His-N ϵ H] configuration), while can be neutral when either the δ -nitrogen (HID, [HN δ -His-N ϵ] configuration), or the ϵ -nitrogen (Hie, [N δ His-N ϵ H] configuration) are deprotonated.

1.3.3 Photoisomerization Efficiency

Figures 1.4A and **1.4B** show a simplified two-dimensional potential energy surface (PES) diagram for the rPSB photoisomerization in animal and microbial rhodopsins.

Light absorption promotes the rPSB in the all-*trans* (microbial) or 11-*cis* (animal) configurations, from the ground electronic state (S_0) to the Franck-Condon (FC) region of the first excited (S_1) state. This vertical transition, as explained in the previous paragraph, involves an instantaneous change in the electronic character of rPSB, described as a CT process from the Schiff base toward the opposite β -ionone region of the molecule. This generates an inverted single and double bond character pattern, triggering nuclear relaxation along the bond length alternation (BLA) of the chromophore. If a new and stable configuration is not found, such relaxation directly steers the rPSB chromophore toward the region of the S_1/S_0 conical

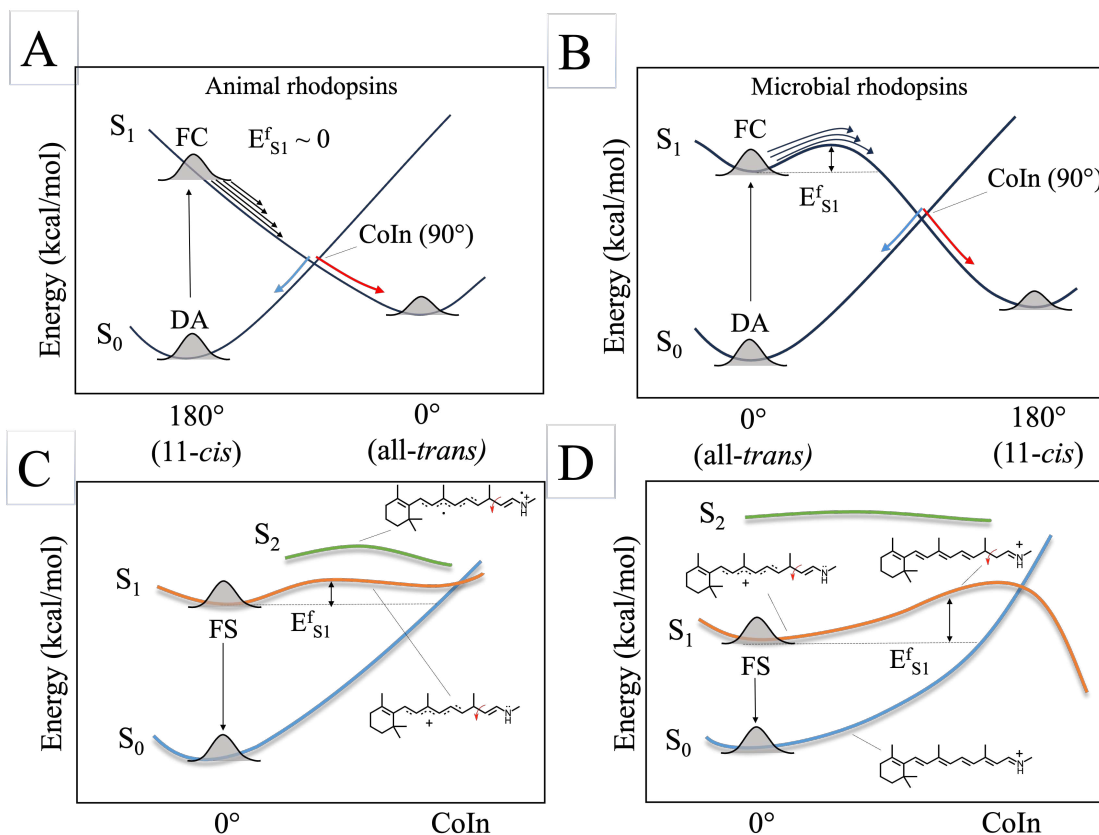


Figure 1.4: Illustration of the rPSB photoisomerization reaction in (A) animal and (B) microbial rhodopsins. In animal rhodopsins the isomerization occurs in ultrafast timescales through a barrier-less energy profile, contrasting with most microbial rhodopsins where longer excited state lifetimes suggest the presence of an isomerization barrier on S_1 ($E_{S_1}^f$). Vertical arrows denote the photon absorption initiating the photoisomerization. Conical intersection (CoIn) represent degenerate points between S_0 and S_1 states. From the CoIn, two reaction paths can be followed: a reactive one leading to photoproduct formation (red arrow) and an unreactive one leading to reactant-back formation, the latter also called internal conversion (blue arrow). (C-D) Origin of the fluorescence in microbial rhodopsin due to the presence of a S_1 barrier impairing the photoisomerization. (C) A shallow S_1 surface or a barrier may be generated via S_1/S_2 mixing (see reference [2]). (D) Alternatively, a barrier can be generated via coupling of the S_1 and S_0 state, as proposed in this thesis for NeoR. rPSB resonance formulas are also given.

intersection (CoIn), featuring a ca. 90° twisted double-bond. The timescale of this process defines the excited state lifetime, while the isomerizing π bond the reaction regio-selectivity. From the CoIn, two S_0 relaxation paths can be then followed: one leading to the successful decay and the corresponding formation of the 13-*cis* or all-*trans* photoproducts (see red ar-

row in **Figure 1.4**), and one to the unsuccessful decay resulting in the backward formation of the original reactants, a process called internal conversion (see blue arrow in **Figure 1.4**).

The combination of high reaction speed and high number of successful decay events is at the core of high ϕ_{iso} and, therefore, high photo-sensitivity. For instance, in animal rhodopsins where the reaction triggers the photocycle implicated in vision, it occurs in ca. 0.1 ± 0.05 ps and with exceptionally high ϕ_{iso} 0.66 ± 0.03 values [5, 42]. Another example of high photo-sensitivity is found in the microbial BR, which displays an excited state lifetime of ca. 0.5 ± 0.1 and an high ϕ_{iso} value of 0.60 ± 0.05 [30].

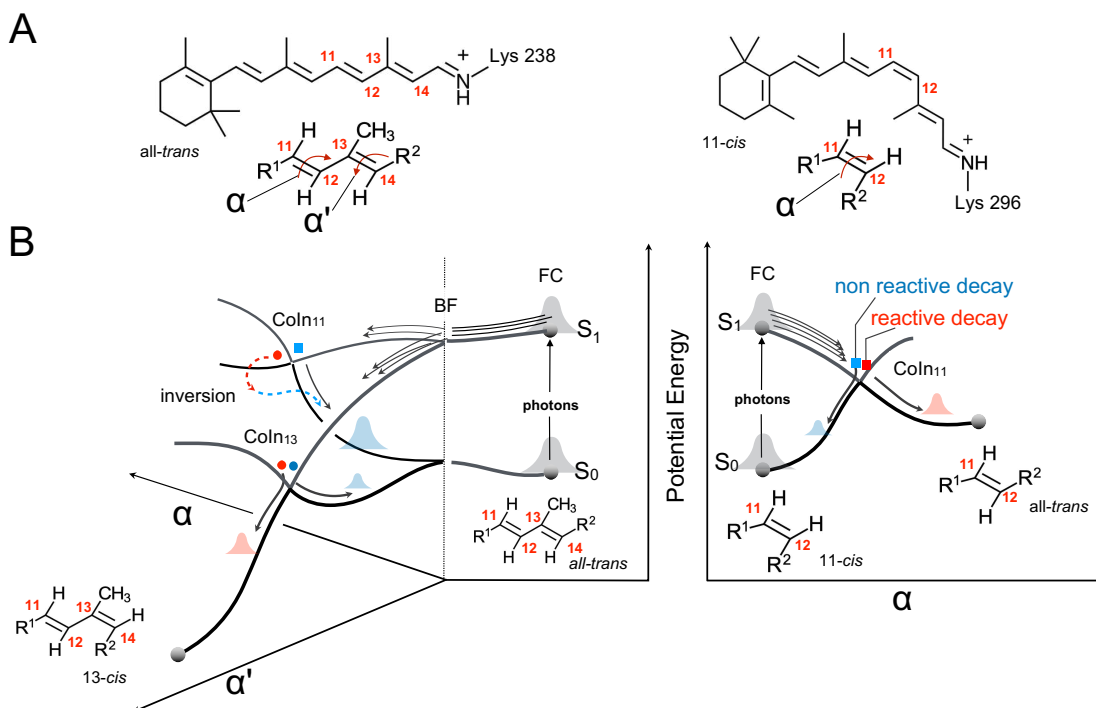


Figure 1.5: S_1 isomerization channels in TaHeR and Rh. A) rPSB chromophore in TaHeR (left) and Rh (right). The moiety at the bottom represents the reactive part of the two chromophores. The curly arrows indicate the S_1 reaction coordinates represented by the α and α' coordinates associated with the rotations around C11=C12 and C13=C14 double bonds, respectively. B) Schematic representation of the photoisomerization channels documented in TaHeR (left) and Rh (right). The arrows in the FC region indicate the S_1 dynamics that, in TaHeR, but not in Rh, leads to an ultrafast breaking (at BF) of the photoexcited population in two subpopulations reaching two different CoIn₁₁ and CoIn₁₃ isomerization decay channels. As documented in **Section 3.2.4**, TaHeR CoIn₁₁ channel was found completely unreactive since leading to internal conversion exclusively.

Nevertheless, most microbial rhodopsins exhibit reduced reaction speeds (> 1 ps) and ϕ_{iso} (< 0.3), as result of the presence of transient, but stable (i.e. associated to a minimum

energy point on S_1), excited state configurations where the rPSB accommodates after S_1 relaxation from the FC (see **Figure 1.4C** and **Figure 1.4D**). As shown in **Section 3.1.5**, this S_1 relaxation is associated to an energy parameter, namely reorganization energy (E_r), whose value can be used to assess the stability of these configurations. The presence of these transient configurations is ensured by sizable excited state photoisomerization energy barrier ($E_{S_1}^f$) that impair the isomerization process. In fact, achieving enhanced photosensitivity relies on the height of $E_{S_1}^f$, as it not only influence the excited state lifetime (i.e. reaction speed) but also the probability for the absorbed photon to be re-emitted through fluorescence, which is detrimental for the rhodopsin function. In summary, for ϕ_{iso} to be very high, the fluorescence quantum yield (ϕ_{fluo}), defined as the ration between absorbed and emitted photons, has to be very low, and this is ensured by null $E_{S_1}^f$ values.

The results documented in **Section 3.2**, report on an original control mechanism for reduced ϕ_{iso} found in TaHeR, a member of the HeR family. This mechanism is not based on a low ϕ_{iso} but, rather, on poor reaction regioselectivity due to the existence of parallel photoisomerization channels. These channels, ensured by similar and almost null $E_{S_1}^f$ values, coexist and compete during the entire excited state deactivation. This picture contrasts with the highly regioselective reaction found in the visual animal rhodopsin expressed in *bos taurus* (Rh), which has served as comparison. **Figure 1.5** illustrates this mechanism. Unlike Rh, which propagates exclusively along one isomerization coordinate associated with the C10-C11=C12-C13 dihedral angle (α , see right panel of **Figure 1.5B**), the reaction in TaHeR lacks in regioselectivity. The excited state population bifurcates in a branching region (BF in **Figure 1.5B**, left panel), initiating parallel isomerization motions along C10-C11=C12-C13 (α) and C12-C13=C14-C15 (α') coordinates, leading to two distinct CoIn's (CoIn₁₁ and CoIn₁₃, respectively). Such lack in selectivity, caused by competing intramolecular isomerization paths associated to similar $E_{S_1}^f$ values, reduces the photoisomerization efficiency of the canonical C13=C14 channel. In fact, the computed photoisomerization efficiency of the all-*trans* to 13-*cis* conversion is 0.17, which would be theorized to be 0.63 if the competing C11=C12 channel were suppressed.

1.3.4 Origin of the Fluorescence in Microbial Rhodopsin

Over the year, the lack of known fluorescent rhodopsins made the mechanism responsible for the origin of the S_1 $E_{S_1}^f$ largely unexplored. As discussed above, this was expected, if one considers fluorescence as an energy wastage process, and thus a process detrimental for the fast and efficient photocycle activation. However, as discussed in **Section 1.4**, the need of modern optogenetic tools based on microbial rhodopsins has fueled the search of artificial variants with enhanced fluorescence. In this regard, one of the most significant contribution was provided by work of Arnold and coworkers [43] on the Arch3 rhodopsin family from *halorubrum sodomenses*, which culminated in the identification of one variant, called Arch7, with a ϕ_{fluo} value of 0.012. Subsequently, in 2020, the discovery of the natural emission fluorescence of NeoR (ϕ_{fluo} value 0.2) offered an unprecedented opportunity to study the

origin of the $E_{S_1}^f$ in microbial rhodopsins, and a large part of this thesis, reported in **Section 3.1**, will be dedicated at this scope. Below, the most relevant notions and findings for the molecular level understanding of the fluorescence in microbial rhodopsins are provided.

The first three electronic singlet states (S_0 , S_1 , and S_2) of the rPSB chromophore in the all-*trans* configuration in its S_0 equilibrium geometry can be classified as having $1A_g$, $1B_u$ and $2A_g$ symmetry, respectively. The $1A_g$ character is associated to a closed-shell covalent electronic structure. The $1B_u$ state is dominated by a singly excited electronic configuration leading to a CT character, while the $2A_g$ state is dominated by a double excitation and has a diradical character. Moreover, the selection rules for electronic transitions dictate that only the $1A_g$ to $1B_u$ and $1B_u$ to $2A_g$ transitions have nonzero oscillator strength [11].

Thus far, the only study providing mechanistic insights into the origin of $E_{S_1}^f$ at the electronic structural level was conducted by Marin et al. on the anabaena sensory rhodopsin (ASR) and two of its mutants [2]. The main findings are summarized in **Figure 1.4C**. According to this mechanism, a barrier can be generated by the electronic state mixing between the reactive S_1 and the partially reactive S_2 states. In fact, the diradical character of $2A_g$ associated to S_2 features a semi-locked C13=C14 bond, opposed to the elongated C13=C14 bond characterizing the $1B_u$ character of the S_1 state. Such mixing would create a small $E_{S_1}^f$ (ca. 2-3 kcal/mol) or, possibly, "shallow" S_1 regions that slow down the evolution toward the CoIn, increasing the probability of the system to re-emit the absorbed photon through fluorescence.

As discussed in **Section 3.1**, the results obtained from the simulation of the rPSB photoisomerization event in NeoR pointed to a novel and different mechanism. It will be shown that the natural fluorescence of NeoR is ensured by an exceptionally high $E_{S_1}^f$ (> 15 kcal/mol) that originates from what we have called *retinal chromophore charge delocalization and confinement process* mechanism [44]. This mechanism does not involve the mixing with the S_2 state, which is found very high in energy, but considers a mixing with the S_0 state that determines a change in the electronic character of the S_1 state during the photoisomerization process. More specifically, it was observed that, from the S_1 minimum energy geometry (or fluorescent state, FS) responsible for the emission, and as soon as the isomerization starts, the S_1 shifts from a CT character, typical of $1B_u$, characterized by a vastly diffused positive charge, to a pure covalent character in the CoIn region, typical of $1A_g$, characterized by a full localized positive charge. This shift pairs with a proportional increase in S_1 energy-destabilization, explaining why in NeoR the canonical all-*trans* to 13-*cis* photoisomerization, typical of microbial rhodopsins, was completely lost, in favor of a brighter fluorescence emission.

1.4 Neorhodospin, an Unusual Fungal Fluorescent Rhodopsin with Optogenetic Potential

Light sensing in fungal life is extremely complex and variegated [45]. Fungi employ different kind of photoreceptors to adapt, escape and conquer new habitats. These include, for instance, flavonoid derivatives like cryptochromes and BLUF proteins for UV and blue light sensing, rhodopsins, and phytochromes for green and red light sensing. Indeed, in certain fungal zoospores (i.e. spores with a motile flagellum) sensing to green light is achieved by employing rhodopsin-guanylyl cyclase (RGC), consisting of an N-terminal rhodopsin coupled to a guanylyl cyclase within one polypeptide chain [45]. In 2020, a more complex RGC system was discovered in the zoospores of the fungus *rhizoclostridium globosum*, from the phylum chytridiomycota [4]. This system was shown to function as a heterodimer, consisting of two sub units with distinct spectral sensitivities. One subunit has a blue/green sensitivity while the other, named as neorhodopsin (NeoR), displays an all-*trans* rPSB chromophore with sensitivity to the far-red, with a $\lambda_{max}^a = 690$ nm. Before this discovery, no other natural rhodopsins had shown a λ_{max}^a greater than 610 nm [46], making NeoR the most red-shifted ever. Interestingly, the zoospores of *rhizoclostridium globosum* do not contain any phytochrome-based photoreceptors, and this is highly suggestive for their possible replacement with NeoR for red light sensing [47]. Successive metagenomic searches have identified additional RGC's across several members of the chytridiomycota phyla, all of which contain NeoR-like sub units with variable sensitivity to the far-red region, with a 640 nm to 721 nm λ_{max}^a range [3].

In addition to the red-shifted absorption, NeoR possesses very unusual spectral properties for being a rhodopsin. In fact, the absorption band was found mirrored with a remarkably intense emission band centered at 707 nm, which yields a minimum Stokes shift value of only 17 nm (350 cm^{-1}). The brightness of the band is quantified by the ϕ_{fluo} value of 0.2 and the extinction coefficient value of $129,000$ $\text{M}^{-1} \text{cm}^{-1}$. As discussed in **Section 1.3.3** and further investigated in **Section 3.1**, such natural bright emission must be ensured by an high E_{S1}^f impairing the rPSB photoisomerization event, and this is also translated in a longer excited state lifetime, found indeed greater than 1 ns [4]. Therefore, NeoR must had witnessed a presently unknown adaptation process that had decreased the efficiency of the photocycle activation, in favor of a brighter fluorescence emission. In fact, such high ϕ_{fluo} , only ca. four times weaker than that of the green fluorescent proteins (GFP) (see **Figure 1.6** for GFP vs. NeoR fluorescence brightness comparison), represents an absolute

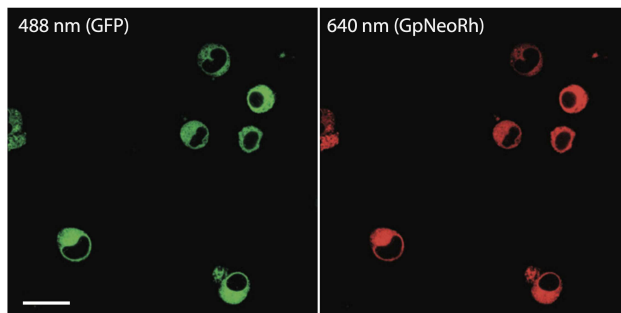


Figure 1.6: Comparison between GFP (left) and NeoR (right) fluorescence emission brightness. Adapted from [3].

anomaly across the family, with rhodopsins commonly exhibiting ϕ_{fluo} values spanning the 0.000001–0.0001 range [48, 49, 50, 51]. Even rhodopsin variants resulting from bio engineering efforts had shown only modest increases, reaching a maximum of 0.012 [43, 52].

Apart from its potential utility as a fluorescent marker protein, NeoR’s unique absorption and emission properties make it an attractive system for applications in optogenetics. Optogenetics involves using light to interfere with the activity of cells in living tissues which have been genetically modified to express light-sensitive proteins, such as rhodopsins [8, 53]. A fascinating application is in neurons. In this regard, one of the challenges faced by optogenetic scientists is the development of methods capable to excite, inhibit and visualize the action potential. Two classes of molecules are adopted for the scope; the actuators and sensors. Actuators are optically controllable molecules which can influence the transmembrane voltage acting as ion channels or pumps. In other words, action potentials may be triggered or suppressed by using light of a specific wavelength. In this regards, ChR’s are widely used for neuronal stimulation with unprecedented spatial and temporal precision. On the other hand, sensors are employed for their voltage-dependent emission properties that can be used to visualize the passage of the action potential along the axons [54].

Nevertheless, despite the significant impact of the optogenetic approach, a notable constraint of this technique is the limited penetration depth of visual light in biological tissues. In contradistinction to all known rhodopsins, NeoR exhibits both absorption and emission properties within the far-red spectral window. This distinctive feature positions NeoR as a promising starting point for developing optogenetic tools excited by near-infrared light, thus achieving unprecedented penetration depths.

1.5 The Discovery of Heliorhopsins

In 2018, in order to extend the search for uncharacterized rhodopsins, Bejà and coworkers employed functional metagenomics to screen a fosmid library from the freshwater habitat of Lake Kinneret (Israel) in the presence of all-*trans* retinal in combination with a search for coloured *Escherichia coli* clones [10]. One clone, the fosmid KIN48C12 yielded reddish *Escherichia coli* cells only when all-*trans* retinal was added. Further analysis pinpointed an area suspected to contain an open reading frame that codes for a hypothetical protein with no detectable sequence similarity to known rhodopsins, but which contains seven predicted transmembrane helices with a lysine residue in the seventh helix. This led to the hypothesis that these transmembrane proteins might represent a previously uncharacterized family of rhodopsin homologues. These proteins were proposed to form a distinct clade in the phylogenetic tree of rhodopsins, confidently separated from type-1 rhodopsins by the < 15% of sequence identity. Indeed, the membrane orientation of these proteins was predicted to be opposite to that of microbial and animal rhodopsins, with the N terminus in the cytoplasm (see **Figure 1.7**). These findings collectively indicated that the identified family of predicted seven transmembrane proteins represented a distinct major group within the rhodopsin family, termed as heliorhodopsins (derived from "helios", meaning "sun", plus

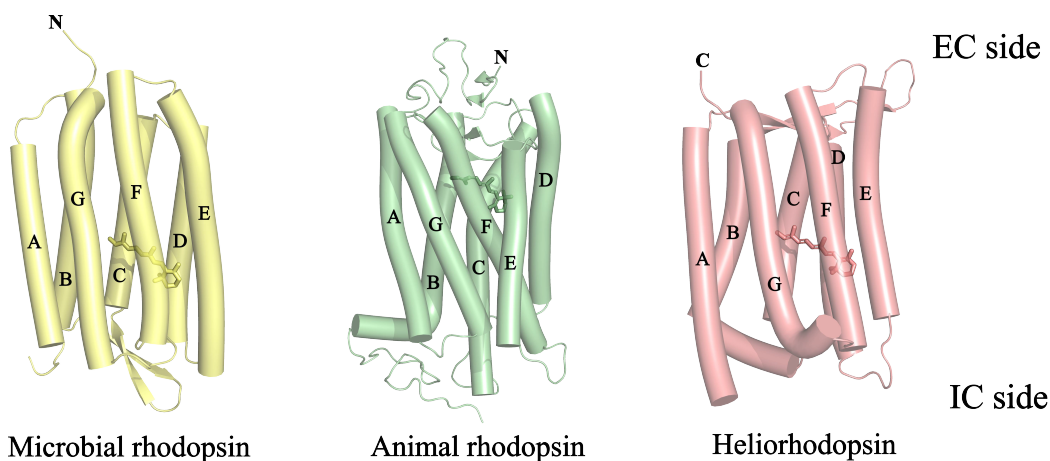


Figure 1.7: Schematic representation of membrane topologies of microbial, animal and heliorhodopsins. The presentation of microbial rhodopsin is based on BR (PDB ID:6G7H), and the animal rhodopsin is based on Rh (PDB ID: 1U19). C and N amino termini facing the extracellular side (EC) are also reported to show the inverted orientation of heliorhodopsins. α -helices are denoted using A to G cylinders.

”rhodopsin”, abbreviated as HeR) [10].

HeRs were identified in bacteria, archaea, algae, and even in certain algal viruses. They are highly abundant and have a global presence in both freshwater and saltwater aquatic environments, with detections in the the Mediterranean Sea, Red Sea, Indian Ocean, Atlantic Ocean, and Pacific Ocean, underscoring their broad ecological distribution [10].

Although the slow ($> 1s$) photocycle was suggestive for a light sensory activity [10], the function of HeRs is still debated. Initially, the lack of a pH change in the external solvent upon illumination of *E. coli* cells expressing HeRs suggested that HeRs may not function as ion transporters or channels [10]. However, more recently, certain viral HeRs of the algae *emiliana huxleyi* functioning as outward proton pumps were discovered, opening up the possibility of using HeRs as novel templates for the development of optogenetic tools [55]. Over the years, various other functions have been proposed with different HeR members, including DNA protection against light-induced oxidative damage [56] or light-driven transport of amphiphilic molecules in diderms [57]. Finally, the growing interest on HeRs is also motivated by their possible exploitation as light-driven redox enzymes possibly capable of carbon fixation [58].

The first crystallographic structure deposited in the PDB databank of a representative of the HeR family belongs to the uncultured *thermoplasmatales archaeon* (TaHeR), a class of acidophiles single-cells organism of the euryarchaeota phylum (PDB ID: 6IS6) [6]. The structure of TaHeR confirmed that HeRs are embedded in the membrane with an inverted orientation relative to other rhodopsins [6]. Overall, TaHeR shares a similar fold with that of type-1 rhodopsins. Interestingly, the crystallographic structure displayed an unexpected lateral fenestration between TM5 and TM6 helices (E and F helices in **Figure 1.7**) through

which the rPSB is exposed to the membrane environment. The fenestration is not observed in the structure of any other rhodopsins and so this observation suggests that this lateral fenestration is a common structural feature among HeRs. In fact, since TaHeR and other HeR possessing species do not have a retinal biosynthetic pathway, it was proposed that HeRs efficiently capture exogenous retinal from the lipid environment [6].

TaHeR bounds a rPSB chromophore in the all-*trans* configuration and exhibits a $\lambda_{max}^a = 542$ nm. HPLC analysis performed before and after illumination showed that, analogously to microbial rhodopsins, most of the rPSB chromophore converts to the 13-*cis* configuration upon illumination. This primary event was shown to trigger an internal proton transfer [10] event (i.e. the proton is never released from the protein) from the rPSB to the counterion (Glu108), and then, according to a recent computational study [55] based on HeR 48C12 (43% of sequence identity with TaHeR), toward a nearby histidines pair, namely His23 and His82, found conserved across all members of the HeR family [6, 10]. Time-resolved spectroscopy assigned to ca. 0.20 the ϕ_{iso} and to 0.22 ps the corresponding time constant. However, the temporal traces describing excited state relaxation were extinguished only in a time range of some picoseconds, and this was highly suggestive for a long excited lifetime, similarly to most microbial rhodopsin.

Chapter 2

Methods

In this thesis the computational study of the light-sensitivity of two novel and recently discovered rhodopsins is presented. The study was carried out through the construction of hybrid quantum mechanics / molecular mechanics (QM/MM) models that has allowed the simulation of the rPSB photoisomerization event in two different protein cavities. QM/MM simulations were used to address target reaction properties, such as maximum absorption wavelength (λ_{max}^a), maximum emission wavelength (λ_{max}^e) or photoisomerization quantum yield (ϕ_{iso}). The aim of this chapter is therefore to provide a brief description of the underlying principles on which these simulations rely, as well as to outline the protocol workflow adopted for the construction of the QM/MM models. If not specified, most of the academic material described below comes from the references [59] and [60].

2.1 Quantum Mechanics methods

2.1.1 The Hamiltonian operator

In quantum chemistry, in order to describe the time-dependent (t) state of a system composed by N-particles, a function ψ of all particle coordinates $\mathbf{R}=(\mathbf{r}_1, \mathbf{r}_1 \dots \mathbf{r}_N)$ called wave function is used:

$$\psi = \psi(\mathbf{R}, t) \tag{2.1}$$

ψ embeds all the information that can be determined about the state of system. Any observable (like position, momentum and energy) of ψ are associated to a proper operator. For instance, let \hat{X} be the monodimensional position operator and $\psi(\mathbf{r}_1, t)$ the wavefunction describing the state of one particle, the so-called *eigenvalue problem* is defined as:

$$\hat{X}\psi(\mathbf{r}_1, t) = x\psi(\mathbf{r}_1, t) \tag{2.2}$$

Where:

- x is the eigenvalue (*characteristic*) of the operator. It corresponds to the value of the x-component of the vector \mathbf{r}_1 describing the position of the particle at a certain time.
- $\psi(\mathbf{r}_1, t)$ is the eigenfunction of \hat{X}

Therefore, in quantum chemistry, one can retrieve properties of the wave function by holding the proper operator and solving the corresponding eigenvalue problem. Now, moving to a molecule with N -nuclei and n -electron, the associated wave function ψ depends on both nuclear $\mathbf{R}=(\mathbf{R}_1, \mathbf{R}_2 \dots \mathbf{R}_N)$ and electronic $\mathbf{r}=(\mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_n)$ coordinates:

$$\psi = \psi(\mathbf{R}, \mathbf{r}, t) \quad (2.3)$$

Again, provided a known wave function, setting the proper eigenvalue problem allows the study of the system. The time-dependent Schrödinger equation describes how to predict the time evolution of the state of the system described by the wave function in **Eq. 2.3**:

$$i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{R}, \mathbf{r}, t) = \hat{H} \Psi(\mathbf{R}, \mathbf{r}, t) \quad (2.4)$$

But, if one is interested in time-invariant wave functions the dependency to the time can be neglected ($\psi(\mathbf{R}, \mathbf{r})$). This leads to the more accessible time independent Schrödinger equation:

$$\hat{H} \psi(\mathbf{R}, \mathbf{r}) = E_{tot} \psi(\mathbf{R}, \mathbf{r}) \quad (2.5)$$

By the comparison of **Eq. 2.2** and **Eq. 2.5** one can notice that solving the time independent Schrödinger equation can be translated in solving is an eigenvalue problem having the operator \hat{H} , the so-called Hamiltonian operator, associated to the total energy of the molecule since consists on all the kinetic (K) and potential (V) energy interactions:

$$\hat{H} = \hat{T} + \hat{V} = \hat{T}_e + \hat{T}_N + \hat{V}_{e-e} + \hat{V}_{N-N} + \hat{V}_{e-N} \quad (2.6)$$

Where \hat{T}_e , \hat{T}_N are respectively the kinetic energy operators of electrons and nuclei, and \hat{V}_{e-e} , \hat{V}_{N-N} and \hat{V}_{e-N} are the potential energy operators of electron-electron interaction, nuclei-nuclei interaction and nuclei-electron interaction, respectively.

Each of these terms can be explicitated in the full functional form using atomic units:

$$\hat{T} = \hat{T}_e + \hat{T}_N = \sum_{i=1}^n \frac{1}{2} \nabla_i^2 + \sum_{k=1}^N \frac{1}{2m_k} \nabla_k^2 \quad (2.7)$$

$$\hat{V} = \hat{V}_{e-e} + \hat{V}_{N-N} + \hat{V}_{e-N} = \frac{1}{4\pi\epsilon_0} \left(\sum_i^n \sum_{j=i+1}^n \frac{1}{r_{ij}} + \sum_k^N \sum_{l=k+1}^N \frac{Z_k Z_l}{r_{kl}} - \sum_i^n \sum_k^N \frac{Z_k}{r_{ik}} \right) \quad (2.8)$$

Where r_{ij} , r_{kl} and r_{ik} indicate, respectively, the electron-electron distance, nuclei-nuclei distance and electron-nuclei distance.

2.1.2 The Born-Oppenheimer approximation

Given the large difference in mass between electrons and nuclei and so in their time scale motion, the Born-Oppenheimer approximation (BOA) allows to consider the motion of the electrons independent from that of the nuclei. This straight-forward assumption makes possible to split the wave function $\psi(\mathbf{R}, \mathbf{r})$:

$$\psi(\mathbf{R}, \mathbf{r}) = \psi_N(\mathbf{R})\psi_e(\mathbf{R}, \mathbf{r}) \quad (2.9)$$

Where $\psi_N(\mathbf{R})$ denotes the nuclear wave function and $\psi_e(\mathbf{R}, \mathbf{r})$ the electronic wave function. Merging together **Eq. 2.9**, **2.5** and **2.6**, one obtains:

$$[\hat{T}_e + \hat{T}_N + \hat{V}_{e-e} + \hat{V}_{N-N} + \hat{V}_{e-N}]\psi_N(\mathbf{R})\psi_e(\mathbf{R}, \mathbf{r}) = E_{tot}\psi_N(\mathbf{R})\psi_e(\mathbf{R}, \mathbf{r}) \quad (2.10)$$

Since \hat{T}_e and \hat{V}_{e-e} are operators associated to the electronic wave function $\psi_e(\mathbf{R}, \mathbf{r})$ while \hat{T}_N and \hat{V}_{N-N} are operators associated to the nuclear wave function $\psi_N(\mathbf{R})$ exclusively, the problem of solving the **Eq. 2.10** can be separated in two tasks, one electronic and one nuclear:

1. Solving the electronic problem:

$$\hat{H}_e\psi_e(\mathbf{r}, \mathbf{R}) = [\hat{T}_e(\mathbf{r}) + \hat{V}_{e-e}(\mathbf{r}) + \hat{V}_{e-N}(\mathbf{r}, \mathbf{R})]\psi_e(\mathbf{r}, \mathbf{R}) = E_{ele}\psi_e(\mathbf{r}, \mathbf{R}) \quad (2.11)$$

This is the electronic Schrödinger equation. \hat{H}_e identifies the electronic Hamiltonian and embeds operators associated to the electronic wave function ψ_e only. Here the dependency to the nuclear coordinates \mathbf{R} is merely parametric. This means that for a given nuclear geometry and a known electronic wave function ψ_e , solving the above electronic problem provides the electronic energy E_{ele} for that geometry.

2. Solving the nuclear problem:

$$[\hat{T}_N(\mathbf{R}) + \hat{V}_{N-N}(\mathbf{R}) + E_{ele}]\psi_N(\mathbf{R}) = E_{tot}\psi_N(\mathbf{R}) \quad (2.12)$$

This is the nuclear Schrödinger equation. For a given nuclear displacement, the electronic energy (E_{ele}) is inserted in the nuclear Schrödinger equation in order to obtain the total energy of the system.

Since most of the times the E_{ele} is enough for solving the computational chemistry tasks, for a given ψ_e , one can repeat the calculations of $E_{ele}(\mathbf{R})$ for each nuclear geometry, or at least only those more significant for a given chemical reaction. This leads to characterization of the PES.

The BOA is usually a good approximation, and remains so until PES of different ψ_e are well energetically spaced. In fact, the approximation breaks down when two (or more) solution of the electronic Schrödinger equation, applied to different ψ_e , come closer energetically,

or in other words, when two different ψ_e becomes degenerate for a certain nuclear disposition, or again in PES terms, when two different PES intersect. These topographic regions of the PES are called CoIn's and play a crucial role in photochemical reactions, such as the photoisomerization of the rPSB.

2.1.3 The Hartree-Fock method

Above it has been illustrated the rationale beyond the calculation of the E_{ele} for a given ψ_e of a molecule with n -electrons. This is possible by solving the electronic Schrödinger equation in **Eq. 2.11** if either the electronic Hamiltonian operator \hat{H}_e or the electronic wave function ψ_e are known. Below a more detailed discussion on how to obtain both the former and the latter term is provided.

To correctly describe an electron, besides its spatial distribution, it is necessary to specify its spin. For this reason, at one electron can be assigned one spin-orbital wave function $\chi(\mathbf{x}_i)$, with \mathbf{x}_i denoting the spin-orbital coordinates:

$$\chi_i(\mathbf{x}_i) = \begin{cases} \phi(\mathbf{r})\alpha(i) \\ \phi(\mathbf{r})\beta(i) \end{cases} \quad (2.13)$$

Here, $\phi(\mathbf{r})$ is the molecular orbital (MO) and describes the spatial distribution of the electron and depends on the adopted choice of set of basis functions ξ_i :

$$\phi = \sum_i^N c_i \xi_i \quad (2.14)$$

While $\alpha(i)$ $\beta(i)$ are the *Spin-up* (\uparrow) and *Spin-down* (\downarrow) spin functions, respectively. The ψ_e becomes:

$$\psi_e(\mathbf{x}_1, \mathbf{x}_2 \dots \mathbf{x}_n) = \chi_1(\mathbf{x}_1)\chi_2(\mathbf{x}_2)\dots\chi_n(\mathbf{x}_n) \quad (2.15)$$

and, in order to make it asymmetric (as electron are *fermion* particles), it can be written as the determinant of an $n \times n$ square matrix, also called Slater determinant (SD):

$$\psi_e(\chi_1, \chi_2 \dots \chi_n) = \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \chi_3(\mathbf{x}_1) & \dots & \chi_k(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \chi_3(\mathbf{x}_2) & \dots & \chi_k(\mathbf{x}_2) \\ \vdots & \vdots & \vdots & & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \chi_3(\mathbf{x}_N) & \dots & \chi_k(\mathbf{x}_N) \end{vmatrix} \quad (2.16)$$

Concerning \hat{H}_e , this can be expressed as :

$$\hat{H}_e = \sum_i^N h_i + \sum_i^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (2.17)$$

Where h_i is the single (*monoelectronic*) Hamiltonian associated to each electron while the second term is V_{e-e} (see **Eq. 2.11**) and accounts for each electron-electron interaction. Solving the Schroedinger equations using this \hat{H}_e represents a prohibitively high-demand computational task. One solution is provided by the Fock operator \hat{f}_i :

$$\hat{f}_i = \sum_i^N [h_i + \hat{v}_i^{HF}] \quad (2.18)$$

The Fock operator is a single electronic operator and lacks of terms that describe each single electron-electron repulsion. However, when acting on one electron, it accounts (via the \hat{v}_i^{HF} term) for the potential energy of that electron in an average mean field created by all the other electrons in the system. Using \hat{f}_i and, accordingly to the variational principle, one can find the set of the MO that minimize the E_{ele} by solving the Fock equations:

$$\hat{f}_i \phi_i = \epsilon_i \phi_i \quad (2.19)$$

2.1.4 The CASSCF method

The basic idea of the configuration interaction (CI) methods is to express the wave function ψ_{CI} as a linear combination of N SDs with the set of MOs provided by the HF method after a self consistent field (SCF) procedure:

$$\psi_{CI} = \sum_{i=0}^n C_i \psi_i \quad \text{with} \quad \psi_0 = \psi_{HF} \quad (2.20)$$

Where each SD represents a different electronic configuration:

$$\psi_{HF} = \begin{pmatrix} \vdots \\ - \\ - \\ - \\ \uparrow \\ \uparrow \\ \uparrow \\ \vdots \end{pmatrix} \quad \psi_1 = \begin{pmatrix} \vdots \\ - \\ - \\ \uparrow \\ \uparrow \\ \uparrow \\ \vdots \end{pmatrix} \quad \psi_2 = \begin{pmatrix} \vdots \\ - \\ - \\ \uparrow \\ \uparrow \\ \uparrow \\ \vdots \end{pmatrix} \quad \dots \quad \psi_N = \begin{pmatrix} \vdots \\ - \\ \uparrow \\ \uparrow \\ \uparrow \\ \uparrow \\ \vdots \end{pmatrix} \quad (2.21)$$

When considering all the possible electronic configurations, one obtains the full configurational interaction (FCI) wave function that can be associated to the exact electronic energy of the molecule. Since even for small molecules and moderately sized basis set the number of SDs is enormous, computing the FCI wave function becomes hardly feasible already for molecules with a few tens of atoms and for most modern computers [61]. One common approach to deal with this problem is to truncate the FCI wavefunctions of some SDs. Examples of truncated FCI wave functions are called complete active space self consistent field

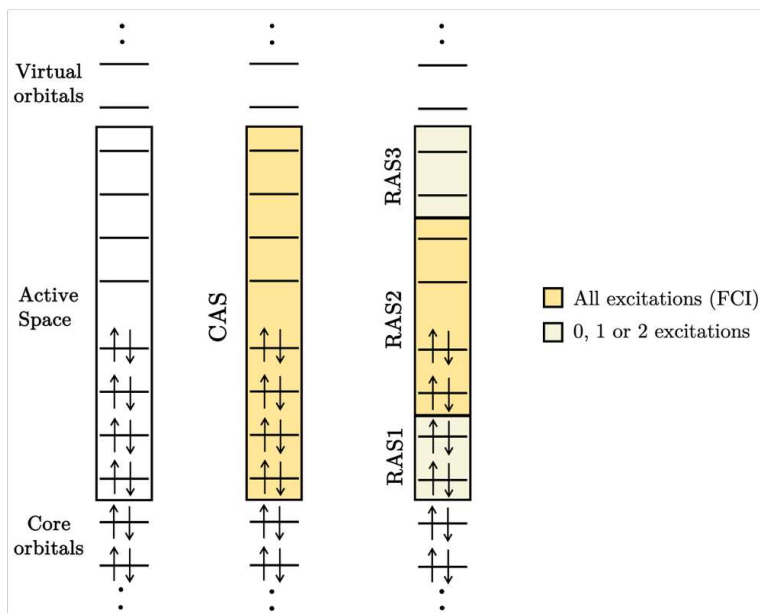


Figure 2.1: From left to right: type of orbitals in CAS/RAS partition, CAS partition, RAS partitions

(CASSCF) and restricted active space self consistent field (RASSCF) wavefunctions [62, 63]. The "complete active space" corresponds to the set of MOs that are chemically significant for the full characterization (excitation process and bond-breaking) of the chemical process under investigation. For instance, concerning the case study provided in this thesis, the correct characterization of the *cis-trans* photoisomerization of the rPSB chromophore required an active space formed by the 6π and $6\pi^*$ orbitals. To summarize, the CASSCF wave function can be defined as the FCI wave function for the set of orbitals contained in the complete active space.

$$\psi_{CAS}(N, n) = \sum_i \psi_i(FCI) \quad (2.22)$$

Where (N, n) is a common notation for complete active space (CAS) wave functions which indicates that N electrons are distributed in all possible ways in n MOs (FCI expansion). Therefore, the notation $\psi_{CAS}(12, 12)$ would indicate a complete active space formed by 12 electrons and 12 MOs.

The "restricted active space" splits the complete active space in three subspaces: RAS1, RAS2 and RAS3. While electrons within RAS2 can undergo to every kind of excitation, electrons in RAS1 can be only doubly excited and MOs in RAS3 can accept only two electrons. **Figure 2.1** illustrates the complete active space and restricted active space orbital partitions.

2.1.5 The CASPT2 method

Given a proper choice of the active space, the CASSCF method can be extremely accurate for exploring the shape of the PES's both in ground state and excited states. However, it is unable to describe the electron correlation arising from the MO's excluded from the active space. This missing description could be incorporated by adding additional SD's to the CASSCF wavefunction. An alternative and cheaper approach is represented by the perturbative CASPT2 method [64, 65, 66] which is widely adopted in computational protocols [67, 61], such as those employed in this thesis. Below is a brief description of the CASPT2 method provided.

The total Hamiltonian is expressed as a sum of an approximate Hamiltonian (\hat{H}_0) and small "perturbations" (\hat{V}_i) multiplied by a constant λ_i :

$$\hat{H} = \hat{H}_0 + \lambda_1 \hat{V}_1 + \lambda_2 \hat{V}_2 \dots + \lambda_n \hat{V}_n \quad (2.23)$$

By solving the Schrödinger equation, the eigenfunctions and eigenvalues of each of these Hamiltonians are:

$$\psi = \psi_0 + \lambda_1 \psi_1 + \lambda_2 \psi_2 \dots + \lambda_n \psi_n \quad (2.24)$$

$$E = E_0 + \lambda_1 E_1 + \lambda_2 E_2 \dots + \lambda_n E_n \quad (2.25)$$

E_0 is the energy associated to the approximate wave function ψ_0 free of any perturbation and is called the zero-order energy. The addition of n-th perturbation terms are used to improve E_0 but the parameter λ decreases in magnitude along the expansion ($\lambda_1 > \lambda_2 > \dots > \lambda_n$). Hypothetically, an infinite expansion would lead to the exact Energy (E). However, it is usually sufficient to truncate the expansion at the second order to ensure an accurate solution. In this thesis, the CASPT2/CASSCF notation is the conventional way to indicate that the CASSCF wave function is corrected to the second-order according to the CASPT2 perturbative method.

2.2 Molecular Mechanics (MM) methods

In the molecular mechanics (MM) approach, molecules are treated as simple "ball and springs" models where the building blocks are the atom types (balls) and the type of bonds between them (springs). The atom type depends on the atomic number and on the chemical environment (*e.g.* aliphatic carbon sp^2 and aromatic carbon sp^2 are different atom types) and the same atom type can be transferred many times inside the molecule. Since electrons are not included, all the motions are treated classically (no quantum effects). As a consequence, many chemical processes, like bond formation/breaking, cannot be simulated. This restricts the field of applications of MM methods to the conformational analysis only.

Given a certain geometry (\mathbf{R}), the MM energy ($E_{MM}(\mathbf{R})$) is written as a sum of terms, each describing the energy associated to a deviation from an equilibrium value:

$$E(\mathbf{R})_{MM} = E_{Stretching} + E_{Bending} + E_{Torsional} + E_{Non-bonding} + E_{Cross} \quad (2.26)$$

Where $E_{Stretching}$ is the energy bonding energy between 2 atoms, $E_{Bending}$ is the energy associated to the angle described by 3 atoms, $E_{Torsional}$ is the energy associated to dihedral angle, $E_{Non-bonding}$ describes the non-bonding interactions between atom A and atom N, and which is composed by Van der Waals (E_{VV}) and Electrostatic (E_{EEL}) interactions, and, finally, E_{Cross} measures the energy coupling between stretching, bending and torsional terms. **Figure 2.2** gives an illustration of the various terms. Each term of **Eq. 2.26** has its own functional form, and it is characterized by parameters which take into account the characteristic of the interaction (*e.g.* l_0 is a parameter contained in $E_{Stretching}$ describing the equilibrium bond length for two atoms). The set of parameters is usually obtained by quantum mechanics (QM) calculations, or alternatively, derived from experimental data (*e.g.* IR spectroscopy). The full form of **Eq. 2.26** constitutes a force field. In this thesis, the AMBER94 has been used for performing MM calculations [68].

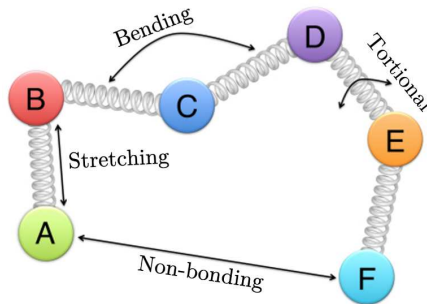


Figure 2.2: Simple illustration of the different contributions to the total MM energy

2.3 Quantum Mechanics / Molecular Mechanics Methods

MM methods, relying on a sum of readily available terms (*i.e.* parameters), are computationally more affordable compared with QM methods, which involve the solution of an eigenvalue problem. However, as stated above, MM methods are not capable of determining the electronic structure of a molecule and thus to describe bond breaking and bond formation. Therefore, in the context of the rPSB photoisomerization a quantum mechanical description is obviously required. The objective of the present research is the investigation of such reaction within a large system, *i.e.* the rhodopsin. Treating the whole system at the QM level, however, proves to be impractical. One solution lies in the application of the *hybrid* QM/MM methods, particularly suitable for studying big molecular systems that contain a small reactive center. The family of rhodopsins represents a perfect field of application; proteins where, as shown in **Section 1.2**, the photo-active rPSB chromophore is encased in a small cavity that is not directly involved in the photochemical reaction.

In QM/MM methods the characterization of the reactive center is left to QM methods while the surrounding environment to MM methods [69, 70]:

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \quad (2.27)$$

Where \hat{H}_{QM} accounts for the QM subsystem considered in vacuo, \hat{H}_{MM} the MM subsystem and $\hat{H}_{QM/MM}$ the interactions between the QM and MM subsystems. The corresponding energies are:

$$E = E_{QM} + E_{MM} + E_{QM/MM} \quad (2.28)$$

A general classification of QM/MM methods is based on the treatment of the frontier between the two QM and MM subsystems and, more specifically, how the $\hat{H}_{QM/MM}$ term is explicitated. In the QM/MM methods adopted in this thesis the LA scheme has been used: a monovalent atom (hydrogen in this case, and so HLA scheme) is placed along the frontier order to saturate the valence of the QM part [71]. Concerning $\hat{H}_{QM/MM}$, the electrostatic embedding method [70] was adopted.

2.3.1 The a-ARM protocol

One example of QM/MM modelling is provided by the automatic rhodopsin modelling (ARM), a protocol that was used for the construction of the QM/MM models of rhodopsins reported in this thesis. Below, the theoretical background needed to follow the steps involved in its workflow is provided.

The ARM protocol is a tool that allows for the automatic, fast and standardized construction of gas phase QM/MM models of wild-type rhodopsin systems and their mutant [72]. An updated version of ARM (*a*-ARM) featuring an increased level of automation has been proposed [73]. The protocol has been validated by calculating the ΔE_{S1-S0}^a for a set of more 30 phylogenetically distant rhodopsins, showing a mean absolute error (MAE) of c.a. +3.0 kcal/mol with respect to the experimental data. Over the years, ARM models have been shown to yield congruous (i.e., built by employing exactly the same protocol) animal and microbial rhodopsin models that correctly reproduce trends in λ_{max}^a value [48, 72, 73, 74, 75, 76, 77].

As shown in **Figure 2.3**, one *a*-ARM model consists on:

- *QM atoms*: Those atoms composing the rPSB chromophore, the linker lysine C_ε atoms and the hydrogen link atom (HLA) atom. These atoms are treated at the QM level.
- *MD atoms*: Those atoms that are free to relax during the molecular dynamics. These include the QM atoms and the side chain atoms of the residues surrounding the rPSB chromophore (cavity residues)
- *MM frozen atoms*: Those atoms whose coordinates are always kept fixed at the original (i.e. crystallographic or homology modelling position). Therefore, these are all the atoms remaining out of the MD partition.

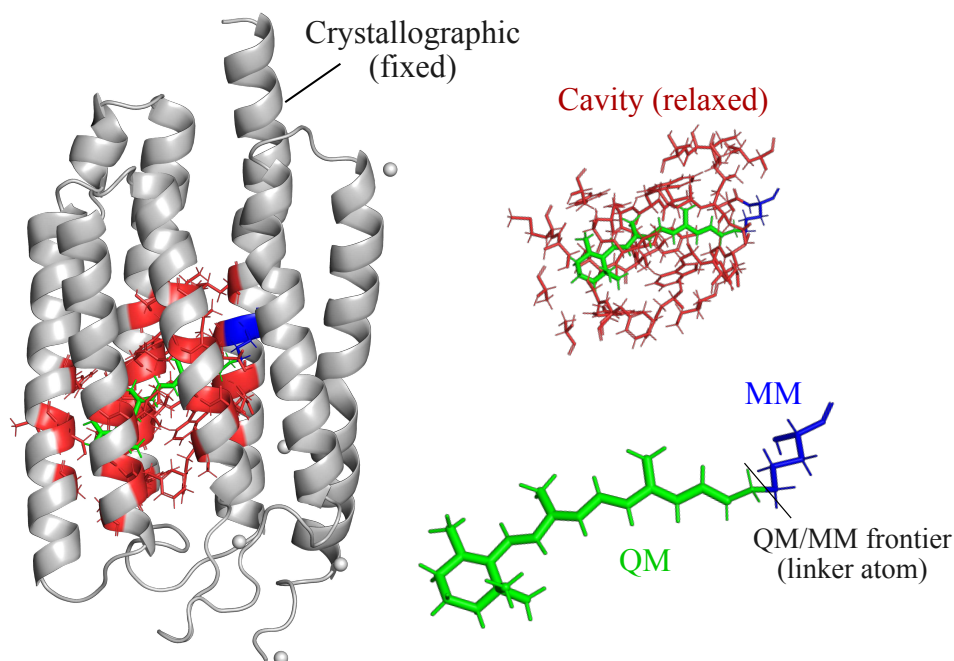


Figure 2.3: Overview of the *a*-ARM model of Neorhodopsin (NeoR). In gray, atoms fixed at the homology model position. In red, relaxed (MM level) atoms composing the rPSB chromophore binding cavity. In blue, relaxed (MM level) atoms of the lysine linker residue (Lys 266) covalently bound to the rPSB chromophore. In green, relaxed atoms (QM level) of the rPSB chromophore. The frontier between QM and MM partitions is also shown.

Technically, the *a*-ARM protocol consists of a collection of bash scripts (controlled by a python-based server) which links a series of publicly available computational packages. The protocol workflow starts with an input file in pdb format that can be directly retrieved from

the protein data bank or from a homology modelling structure. In the initial setup, the rPSB chromophore, the lysine linker and the rPSB counterions are identified automatically without calls to any external program while the protein cavity is identified with fpocket [78]. Then, Propka3.1 [79] is invoked to predict the ionization states of charged amino acids and total charge of the model is neutralized adding Cl^- and Na^+ ions in specific positions by independently neutralize the inner and outer (i.e. intra-cellular and extra-cellular) protein surfaces. Eventual mutations are carried out by MODELLER [80], a program for predicting sidechains conformations for a given protein backbone, using a backbone-dependent rotamer library.

After this initial setup, the hydrogen atoms, usually omitted in crystallographic structures and poorly treated by the homology modeling calculations, are added by Dowser [81]. The system is then optimized at the MM level and 10 independent molecular dynamics room-temperature relaxations of 1 ns are performed at the MM level using GROMACS [82]. Each MD consists of 50 ps of heating phase, 150 ps of equilibration phase and 800 ps of production phase. The final structures of the MD are then prepared of the QM/MM calculations performed by Molcas[83]/Tinker[84] interface. The HLA is added across the $\text{C}_\delta\text{-C}_\epsilon$ bond (the QM/MM frontier, see **Figure 2.3**) and the QM atoms and MM atoms are readily identified. The 10 models are optimized in parallel through different steps featuring an increased level of theory: from HF/3-21G/AMBER94 to single root CASSCF(12,12)/6-31G*/AMBER94 to obtain the final geometries. These geometries are finally used to calculate $10 \Delta E_{S_1-S_0}^a$ using a 3-root CASPT2/CASSCF(12,12)/6-31G*/AMBER94 calculation to correct the energy of the CASSCF(12,12)/6-31G*/AMBER94 wave function.

2.4 Semiclassical Molecular Dynamics

The aim of molecular dynamics (MD) is to study the time evolution of molecular systems. Using this method, it is possible to derive time dependent properties, like excited state lifetimes or reaction rates, statistical properties, like ϕ_{iso} , and, even more importantly, insights into the dynamics factors governing reaction mechanisms. In this thesis we employed one of the possible approaches, called semiclassical molecular dynamics (SCMD), chosen because it can deal, with great efficiency and reasonable computational expense, with the photochemistry of rhodopsins [11, 61]. In fact, for the description of photochemical reactions, it is mandatory to use a methodology that is able to describe multiple PESs and the radiationless transitions among them. For this reason, classical MD approaches are not informative since they are capable to sample the ground state PES exclusively.

In a MD simulation, nuclei can be treated in good approximation as classical particles and so the time evolution of their coordinates can be simulated by solving the Newton's second equation of motion:

$$-\frac{dV}{d\mathbf{r}} = m \frac{d^2\mathbf{r}}{dt^2} \quad (2.29)$$

Here, the V term is the potential energy at position \mathbf{r} . The left-hand side is the negative of the energy gradient, also called the force (\mathbf{F}) on the particle(s). In semiclassical calculations (e.g. SCMD), the energy gradients are derived from QM (or QM/MM) calculations. The method that is used in this thesis to propagate the nuclear coordinates is called *Verlet algorithm*. It is a numerical method that can be used in classical representations of any particle system to propagate the coordinates over the time. In the pure Velocity Verlet algorithm (adopted in this thesis) the geometry and the velocity of the system at the next time step ($\mathbf{x}_{t+\Delta t}$ and $\mathbf{x}'_{t+\Delta t}$, respectively) are calculated using geometry, force and velocity from the current step (\mathbf{x}_t , \mathbf{F}_t and \mathbf{x}'_t), using the following equations, obtained with a second order Taylor expansion of the nuclear coordinates:

$$\mathbf{x}'_{t+\Delta t/2} = \mathbf{x}'_t + \frac{\mathbf{F}_t}{2m} \Delta t \quad (2.30)$$

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \mathbf{x}'_{t+\Delta t/2} \Delta t \quad (2.31)$$

$$\mathbf{x}_{t+\Delta t} \rightarrow \mathbf{F}_{t+\Delta t} \quad (2.32)$$

$$\mathbf{x}'_{t+\Delta t} = \mathbf{x}'_{t+\Delta t/2} + \frac{\mathbf{F}_{t+\Delta t}}{2m} \Delta t \quad (2.33)$$

As a result, the molecular system is propagated using forces (energy gradients) calculated on a single PES, also called relax root. In order to describe electronic transition, *ad-hoc* surface hopping algorithms are used. Particularly, they control the root of relaxation at each step, deciding whether the system jumps or not to another PES. In this thesis the stochastic Tully surface-hop method was employed [85].

Summarizing, we refer to SCMD for a molecular simulation in which:

1. *Nuclei* are treated as simple classical particles and propagated using classical algorithms (i.e. *Verlet*)
2. Forces are calculated with QM methods
3. Quantum effects within classical approximations are simulated using *ad-hoc*'s algorithms (e.g. Tully's)

and we define the *initial condition* of a SCMD:

- a *geometry*, a vector of the initial nuclei positions
- a *velocity*, a vector of the initial nuclei velocities (i.e. forces)
- a *relax root*, the initial PES (i.e. electronic state) of relaxation

2.4.1 Room Temperature Simulation Protocol

Here, it is provided the description of the protocol adopted to generate the ground state room temperature Boltzmann-like distribution of the rhodopsins investigated in this thesis. The described protocol was taken from references [42, 86] where the rPSB photoisomerization of the rhodopsin expressed in *bos taurus* was recently studied. These room temperature distributions are represented by a set of initial conditions (a set of geometries plus velocities, see above) that are descriptive and coherent with a thermodynamic ensemble. Therefore, in any case, before exciting and propagating the computed population on the desired excited state, it is important to assess the quality of the distributions. In the study-cases reported in this work (**Sections 3.1 3.2**) this quality check was always carried out by computing the absorption band and comparing it with the experimentally observed one. Only when a good agreement in both shape and maximum absorption wavelength is shown, the set of initial conditions were considered suitable for the excited state propagation. In this thesis, the absorption bands were computed from the Gaussian convolution of the stick spectra generated by computing $\Delta E_{S_1-S_0}^a$ values at the 3 root state average CASPT2/CASSCF(12,12)/6-31G*/AMBER94 level of theory from the geometries of the initial conditions and using the oscillator strengths ($f_{S_0S_1}$) to weight the intensity of each $S_0 \rightarrow S_1$ transition. Clearly, this procedure cannot consider and take into account band line shapes due to vibronic transitions. This limitation is clearly exemplified by the case of the NeoR absorption band (see **Section 3.1.3** which displays, indeed, a vibrationally resolved structure. Below, it is provided a schematic description of the algorithm adopted to produce the initial conditions. This is composed of a sequential series of steps:

1. Starting from the ground state (S_0) equilibrium structure of the QM/MM model (e.g. *a*-ARM model), a room temperature ground state MD at the MM level of theory and at the temperature of 298 K is runned using AMBER94 force field.
2. Snapshots (geometries and velocities) are extracted from the above MD at time intervals of 100 ps. For instance, if 50 initial conditions are needed, one needs to run 50 x 100 ps = 5 ns of MD in the 1st step.
3. For each snapshot, a 200 fs trajectory with a step size of 1 fs at HF/6-31G*/Amber94 level of theory is performed.
4. Then, each trajectory is continued for 50 fs at the 2-roots state average CASSCF(12,12)/6-31G*/Amber94 level of theory with the energy gradient computed on the ground state (i.e. relaxed root = S_0). Finally, the final set of geometries and velocities represents the initial conditions.

Chapter 3

Results

3.1 Red-shift and Fluorescence Enhancement of Neorhodopsin

The results presented in this section, drawing upon the findings from **Paper 1** and **Paper 2**, aim to unveil the molecular mechanism underlying the extreme and, not canonical for being a microbial rhodopsin, spectral properties of neorhodopsin (NeoR). These properties include the significant bathochromic shift, high emission brightness, small Stokes shift, and distinctive Resonance Raman (RR) spectral profile. The study involved the construction of a QM/MM model for NeoR, which was firstly shown to accurately reproduce the experimentally observed spectral behavior, and then used to investigate the change in the geometrical and electronic characters experienced by the rPSB chromophore during the photoisomerization event. Specifically, the QM/MM simulations revealed that, due to an unconventional choice of the counterion configuration, the canonical microbial-like photoisomerization reaction around the C13=C14 bond became highly disfavoured, having an energy barrier estimated greater than 20 kcal/mol. This led to a prolonged excited state lifetime and to an enhanced emission brightness. Furthermore, being the isomerization on the C13=C14 blocked, the calculations pointed to the opening of alternative, typically unreactive in microbial rhodopsin, reaction pathways involving the C9=C10 and C7=C8 double bonds.

3.1.1 Abstract

Understanding how the rhodopsin sequence can be modified to precisely modulate the spectroscopic properties of its retinal protonated Schiff base chromophore is a prerequisite for the rational design of more effective optogenetic tools. A key challenge is in establishing the rules necessary for achieving highly fluorescent rhodopsins with near-infrared absorption. In this study multi-configurational quantum chemistry is utilized to create a computer model of a recently discovered natural rhodopsin, called Neorhodopsin, which exhibits precisely such properties. The model, shown able to successfully replicate relevant experimental observables,

reveals a geometrical and electronic structure of the chromophore with a highly diffuse charge distribution along its conjugated chain. Furthermore, the model indicates that a charge confinement process occurring along the chromophore's excited state isomerization coordinate is the primary cause of the observed fluorescence enhancement.

3.1.2 Introduction

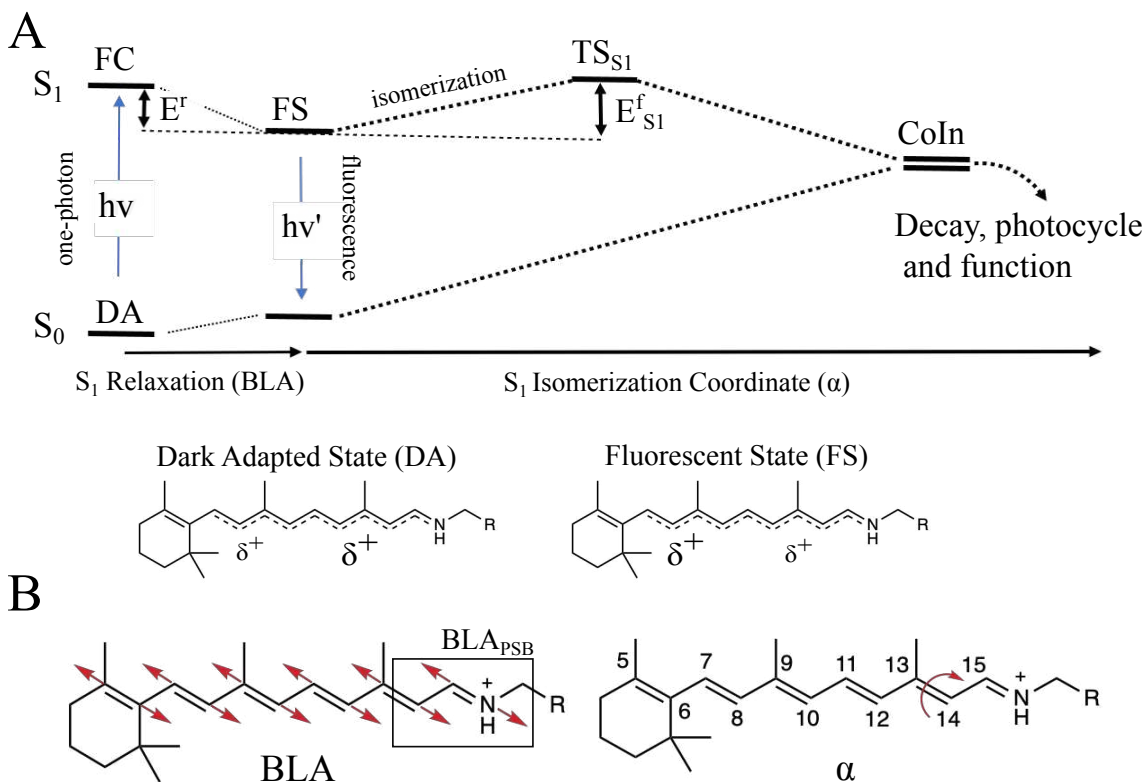


Figure 3.1: Geometrical and electronic structural changes in NeoR photoisomerization. A) Schematic representation of the hypothetical S_0 and S_1 energy changes occurring along the S_1 relaxation that involves the bond length alternation (BLA, quantified by the difference between the average of the double-bond lengths and the average of the single-bond lengths of the conjugated chain) and isomerization (α) coordinates. The rPSB resonance formula show a delocalized positive charge at the S_0 and S_1 energy minima corresponding to the dark adapted (DA) and fluorescent (FS) states, respectively. The symbol “ $\delta+$ ” gives a qualitative measure of the amount of positive charge located along the rPSB conjugated chain. B) Representation of the bond length alternation (BLA) mode and the α mode along the C13=C14 double bond. BLA_{PSB} is the -C14-C15 and C15-N bond lengths difference.

As it was outlined in **Section 1.4**, modern optogenetics require membrane-localized signaling tools that could emit intense fluorescence upon irradiation with red light. However,

until recently, the available tools, based on engineered microbial rhodopsins, could only generate weak fluorescence signals that impair their performance.

Since the optical properties of microbial rhodopsins owe to the presence of the covalently bounded all-*trans* retinal protonated Schiff base (rPSB) chromophore and its interaction with the surrounding protein environment, a deep molecular comprehension of the factor dictating such properties is highly desirable. In this regard, few studies [87, 88, 89] have formulated rules for tailoring the absorption and emission properties of the retinal chromophore based on the effects of homogeneous electrostatic fields acting on isolated chromophores or via chromophore chemical modifications. However, it is expected that a simple electrostatic picture might not be sufficient to explain the origin of these properties in the complex environment offered by the protein cavity. Other factors, such as non-homogeneous electrostatic fields or chromophore-cavity steric effects, could also play an important role.

In 2020 the discovery of NeoR [4] offered an unprecedented case study that could potentially expand our comprehension of red-shifted and highly fluorescent rhodopsins. As shown in **Section 1.4**, NeoR belongs to the family of rhodopsin guanylyl-cyclase (RGC) expressed in the *rhizoclostridium globosum* from Chytridiomycota, the only phylum of fungi producing motile and flagellated spores (zoospores) [47, 90], and expresses, by far, a rPSB chromophore with the most red-shifted λ_{max}^a and the brightest fluorescence among all known microbial rhodopsins.

Deciphering how natural evolution in NeoR has tuned these extreme spectroscopic properties of the rPSB chromophore could expand our ability to design optogenetic tools with augmented functionality. Therefore, NeoR modelling has represented a new promising learning opportunity that could be also used to assess the transferability of the rules mentioned above. In particular, NeoR offered the opportunity to disclose the molecular-level mechanism controlling the branching between fluorescence emission and photoisomerization. Such branching, which is schematically illustrated in **Figure 3.1** for all-*trans* rPSB, has been shown to dominate the fluorescence modulation in a set of GFP-like protein variants [91, 92]. More specifically, in these systems, the ϕ_{iso} appears to be directly proportional to the energy barrier ($E_{S_1}^f$) controlling access to a CoIn located along the first singlet excited state (S_1) isomerization coordinate and controlling the decay to the ground state (S_0). In this study it was assumed that the same mechanism operates in NeoR that was then used as a “laboratory” model for proposing a mechanism capable to connect sequence variation and rPSB emission. To do so, it was also assumed, in line with the evidence coming from a set of Arch3 variants displaying enhanced fluorescence [93, 94], that the NeoR emission was a one-photon process and that, therefore, originates directly from its DA.

In order to pursue these objectives, a QM/MM model of NeoR based on multiconfigurational quantum chemistry was constructed and used to simulate the photoisomerization event. Given the lack of any crystallographic structure, the QM/MM model was built from a previously reported homology model structure [4], predicted by MODELLER [95]. While the absence of a crystallographic structure may have limited, in principle, the fidelity of the environment description with respect to that found in nature, the target was to generate an

atomistic model capable to replicate all relevant spectroscopic and photochemical observables and to use it to explain the high ϕ_{fluor} of NeoR in terms of geometrical, electrostatic and steric effects.

Accordingly, after the validation, the QM/MM model was firstly used to investigate the electronic structures of the NeoR DA and fluorescent state (FS) and, secondly, was used to investigate the NeoR photoisomerization with the target of documenting the magnitude and origin of $E_{S_1}^f$. It was shown that the confinement of the delocalized positive charge on the Schiff base moiety of the rPSB backbone can explain the existence of large $E_{S_1}^f$ values and, in turn, the high ϕ_{iso} of NeoR.

3.1.3 Model Construction and Validation

An initial QM/MM model of NeoR was constructed using the *a*-ARM technology (**Section 2.3.1**) [96, 73, 72] starting from the homology modelling predicted structure of MODELLER mentioned above. The model showed that the all-*trans* rPSB chromophore of NeoR is embedded in a cavity featuring a peculiar amino acid composition with two glutamic (E136 and E262) and one aspartic (D140) acidic residues located in the vicinity of the Schiff base moiety. However, due to the lack of experimental information on the residue protonation state, the chromophore counterion assignment remained ambiguous [67]. For this reason, a set of customized *a*-ARM models featuring different protonation states for the E136, E262, D140 plus E141, a residue located halfway along the rPSB conjugated chain (see **Figure 3.2A**), were built and ranked by computing the λ_{max}^a and maximum emission wavelength (λ_{max}^e) maxima as well as the E_r defined by the basic mechanism illustrated in **Figure 3.1A**. The λ_{max}^a and λ_{max}^e values were computed in terms of $\Delta E_{S_1-S_0}^a$ between S_0 and S_1 at the DA and FS equilibrium geometries, respectively. E_r was instead computed as the energy difference between the FC point and FS state and was used to quantify the energy decrease associated with S_1 relaxation. The results are collected in **Figure 3.1B** which displays the λ_{max}^a , λ_{max}^e , and E_r values for models where the ‘‘counterion tetrad’’ defined above have total charges of 0, -1, -2, -3. For completeness, the scenario with a total charge -3 was reported even if the two transitions displayed by these models were not allowed, being the oscillator strength (f_{osc}) close to zero.

Figure 3.2 shows that the model with a deprotonated E141 and neutral E136, D140, and E262 (from now simply *a*-ARM_{E141}) residues was the most accurate. In fact, *a*-ARM_{E141} yielded λ_{max}^a and λ_{max}^e values only -46 and -14 nm blue-shifted with respect to the experimental values, as well as the smallest difference (49 nm, translated in 1098 cm⁻¹) between those values consistent with the experimentally observed tiny Stokes shift. All other assessed protonation states yielded a poor comparison with the observed quantities. For instance, although the model with all four residues protonated (i.e., with no counterion) produced a λ_{max}^a close to the experimental one, it lacked of a stable FS structure since no energy barrier could be located preventing access to the CoIn along α . Also, consistently with the high intensity of the observed absorption and emission bands [4, 3], the computed f_{osc} were found very high: 1.71 and 1.90. Such values were confirmed via multistate XMCQDPT2 calculations:

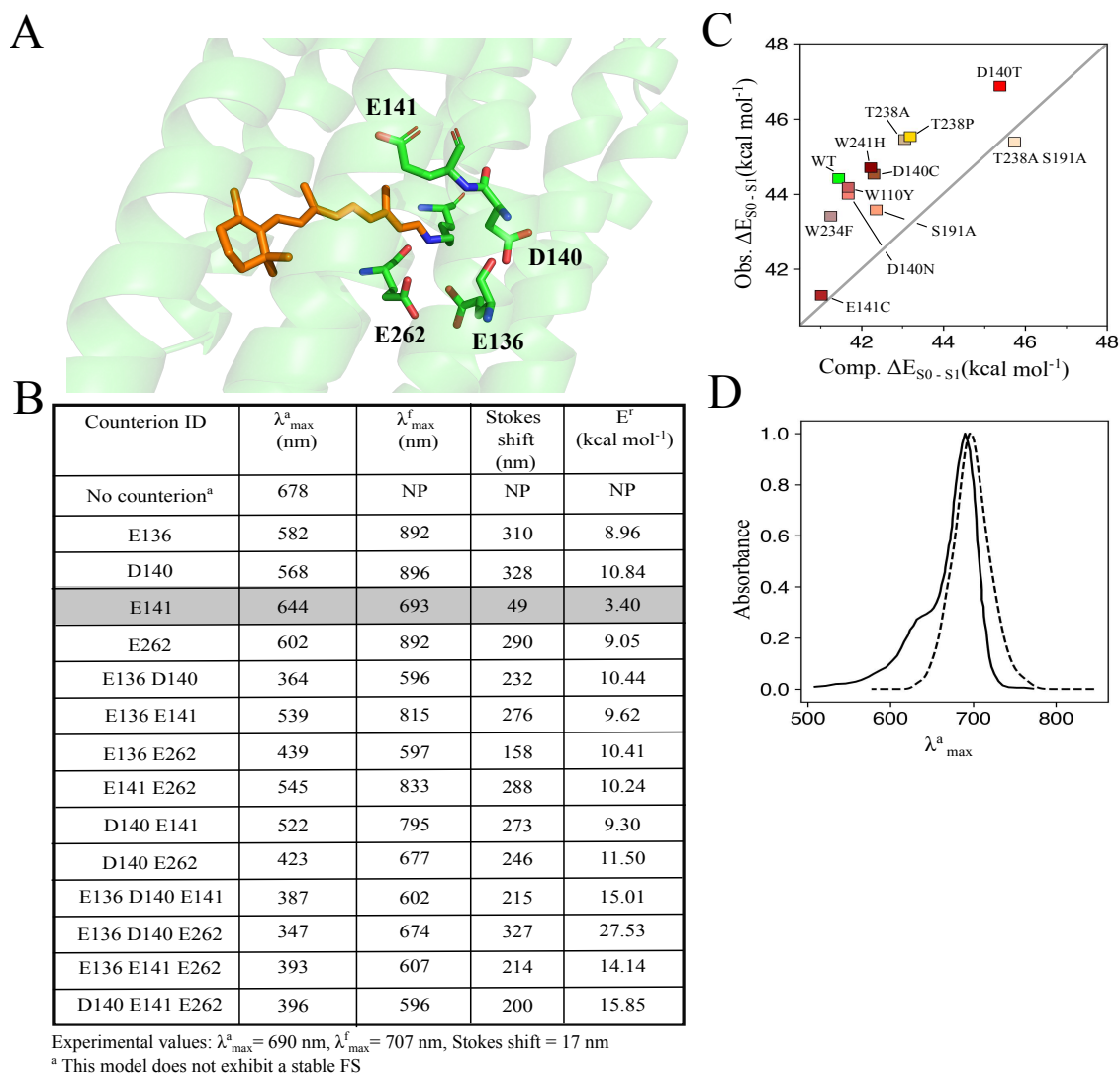


Figure 3.2: Choice of the NeoR chromophore counterion and model validation. A) Overview of the structure of the all-*trans* rPSB chromophore (orange) and its four potential residue counterions (in green). The lysine residue (in green) bounded to the rPSB chromophore is also displayed. B) Computed (CASPT2 level) λ_{\max}^a , λ_{\max}^e and E^r of NeoR with varying counterion choices. C) Correlation between experimental (Obs. ΔE_{S0-S1}) and computed (Comp. ΔE_{S0-S1}) values of vertical excitation energies defining λ_{\max}^a in the wild type (indicated as WT) and a set of mutants. D) Superimposition of experimental and computed (dotted line, protocol adopted for band simulation given in **Section 2.4.1**) absorption band of wild type NeoR. The experimental band has been digitalized from the corresponding ref. [4]

1.66 and 1.80.

In order to further assess the quality of $\alpha\text{-ARM}_{E141}$, there were also constructed the models for a set of NeoR variants whose λ_{\max}^a have been experimentally measured. As shown

in **Figure 3.2C**, the models reproduced the experimentally observed trend indicating that a-ARM_{E141} described, qualitatively, the effects of cavity residue replacements. To notice that the trend was reproduced with a systematic blue shift, which is typical of a-ARM models [73, 96, 97]. a-ARM_{E141} was also used to simulate the wild type NeoR absorption band at room temperature by computing $\Delta E_{S_1-S_0}^a$ values and S_0 to S_1 transition probabilities for 200 initial conditions representing the Boltzmann distribution. Comparison between the simulated and observed data in **Figure 3.2D** showed that the center of the computed band (703 nm) was only +13 nm red-shifted with respect to the experimental λ_{max}^a value. a-ARM_{E141} only appeared to miss a shoulder at 640 nm that is, likely, of vibronic origin and therefore not captured by our simulation based on electronic transitions exclusively (see **Section 2.4.1**).

3.1.4 Electronic Character of the Absorption and Emission

The agreement with the experimental data has allowed to use a-ARM_{E141} to investigate the large bathochromic shift, negligible Stokes shift, and intense fluorescence of NeoR. In this regard, the aim of **Figure 3.3** is to document the variation in electronic character upon light absorption and emission by examining the vertical electron density changes at DA and FS ($\delta\rho_{abs}$ and $\delta\rho_{emi}$, respectively). The results supported the hypothesis that both S_1 and S_0 electronic characters were combinations of putative covalent (COV) and charge transfer (CT) states loosely associated with the limiting resonance formulas of **Figure 3.3A**. In particular, the results revealed that, the DA structure displayed an unusually large CT weight in S_0 yielding a positive charge spread along the carbon atoms of the chromophore with the limited +0.29e fraction of charge residing in the C14-C15-N-C ϵ moiety (from now on, the charge residing on such moiety will be called Charge_{PSB}). The vertical transition to S_1 only slightly modified such charge distribution. For instance, when taking the mid C13=C14 as a reference one can see only a small 0.03e translocation towards the β -ionone ring. The same behavior was seen at FS when looking at the vertical emissive transition for which one finds Charge_{PSB} values of +0.27e and +0.24e for S_1 and S_0 , respectively. In conclusion, as illustrated in **Figure 3.3B, C** the DA and FS transitions can be both qualitatively interpreted as transitions between adiabatic states (i.e., S_0 and S_1) corresponding to in-phase and out-of-phase mixing of two diabatals (or resonance formulas) close in energy. Such interpretation appears to be related to the one proposed for explaining the observed absorption and emission trends of GFP-like proteins [91].

The description above was not in line with the consensus electronic structure of the rPSB chromophore [98, 1] of rhodopsins. In fact, the DA $S_0 \rightarrow S_1$ transition, is usually described as a transition starting from a COV-dominated state featuring a positive charge localized on the -C15=N- moiety and not a delocalized charge spread on the -C9=C10-C11=C12-C13=C14-C15=N- chain as seen in **Figure 3.3B**. In the past, starting from such a state, an at least three times larger charge translocation has been computed upon DA $S_0 \rightarrow S_1$ excitation for bovine rhodopsin [98] and BR [99].

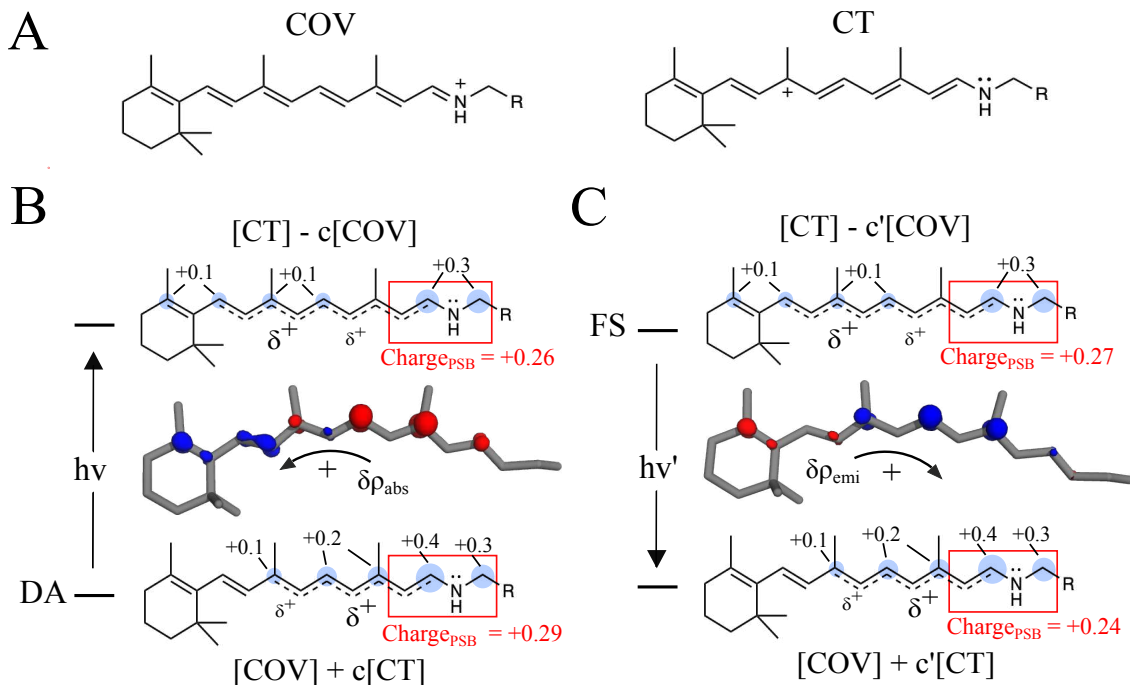


Figure 3.3: Electronic character variation along the S_1 relaxation of the rPSB chromophore in Neor. A) Representation of the two limiting resonance formulas adopted to describe the electronic character of the rPSB chromophore. B) Electron density variation ($\delta\rho_{abs}$) characterizing the vertical $S_0 \rightarrow S_1$ transition from the DA. Blue and red clouds correspond to electron density decrease and increase respectively. Isovalue set to 0.002 a.u. The associated resonance formulas correspond to resonance hybrids also anticipated in **Figure 3.1**. Blue bubbles represent the QM positive charge (in e unites). Only absolute values > 0.05 e are reported. As indicated by the red box, the total charge residing in the -C14-C15-N-C ϵ - rPSB fragment is also given. C) Same data for the $S_1 \rightarrow S_0$ emission from the FS.

3.1.5 S_1 relaxation

Consistently with the computed negligible (0.01e) difference in Charge_{PSB} value between FC point and the FS state, the S_1 electronic relaxation of a-ARM $_{E141}$ can be interpreted as a relatively minor change in the weights of the COV and CT diabatic states. The geometrical variation accompanying such a process is documented in **Section 3.4** and corresponds to a minor progression along the BLA coordinate (this is defined as the difference between the average single-bond length and the average double-bond length of a conjugated chain, see **Figure 3.1**) of the chromophore, leading to a E_r 3.5 kcal/mol value (see **Figure 3.2**). Such assignment is based on the hypothesis of a negligible contribution from the surrounding protein environment, in line with the results reported for GFP's [91]. Interestingly, the BLA-driven nature of the S_1 relaxation is a known feature among GFPs where the only geometrical difference between the DA and FS states is attributed to a different BLA displacement [91].

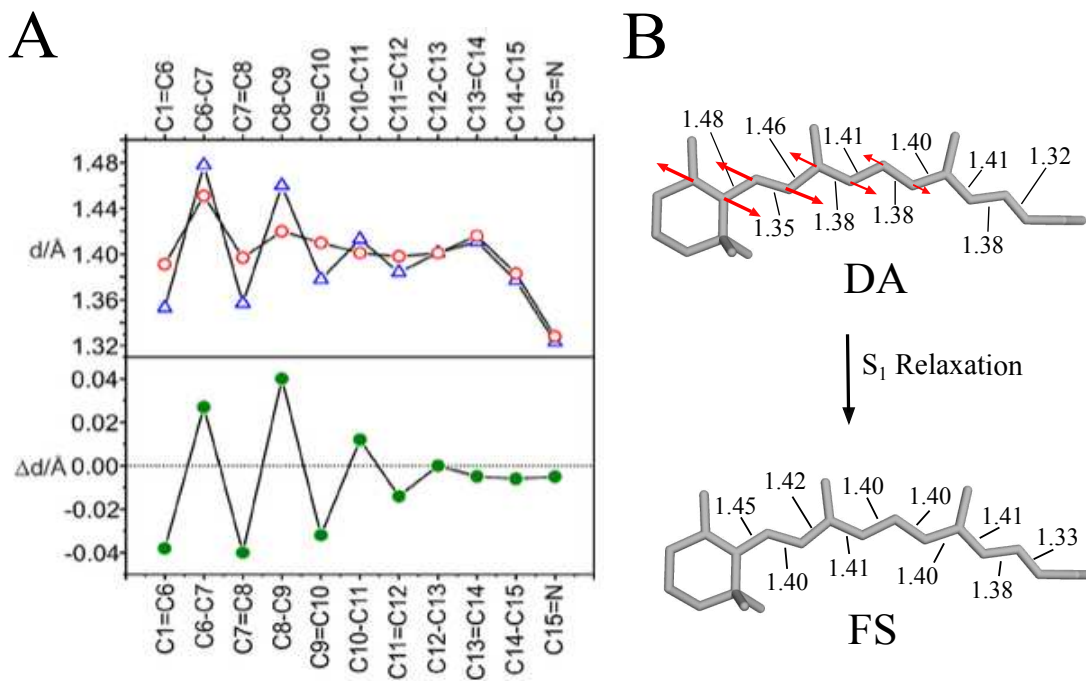


Figure 3.4: NeoR eometrical character variation along the S_1 relaxation of the rPSB chromophore. A) Bond lengths of the polyene chain of the rPSB chromophore. The open triangles and circles refer to the DA and FS structures, respectively. Bottom, the bond length differences $\Delta d = d_{S_{DA}} - d_{S_{FS}}$ (green) representing the displacements of the stretching coordinates found in the FS. The red and blue colors refer to double and single bonds, respectively. B) Same as A) but using a molecular view. The arrows indicate the dominant geometrical change corresponding, clearly, to a variation in the bond length alternation (BLA, see definition in the caption of **Section 3.1**) in a region of the conjugated chain distant from the Schiff base moiety. The relevant bond lengths are given in A.

In NeoR, the BLA change was mostly described by a contraction and elongation of C6-C7/C8-C9/C10-C11 and the C7=C8/C9=C10/C11=C12 bonds, respectively, while the BLA of the C14- C15-N Schiff base moiety (from now on simply BLA_{PSB} , see **Figure 3.1B**) was minimally involved and only changed of 0.01 Å.

3.1.6 NeoR photoisomerization

As anticipated above and discussed in **Section 1.3.3** and **1.3.4**, below we assumed that the branching (see **Figure 3.1**) between the canonical rPSB double bond photoisomerization of microbial rhodopsin and emission determined the fluorescence brightness of NeoR. Accordingly, α -ARM_{E141} was used to compute the approximate S_1 minimum energy path (MEP) describing the clockwise (CW) torsional deformation along α , namely the C12-C13-C14-C15

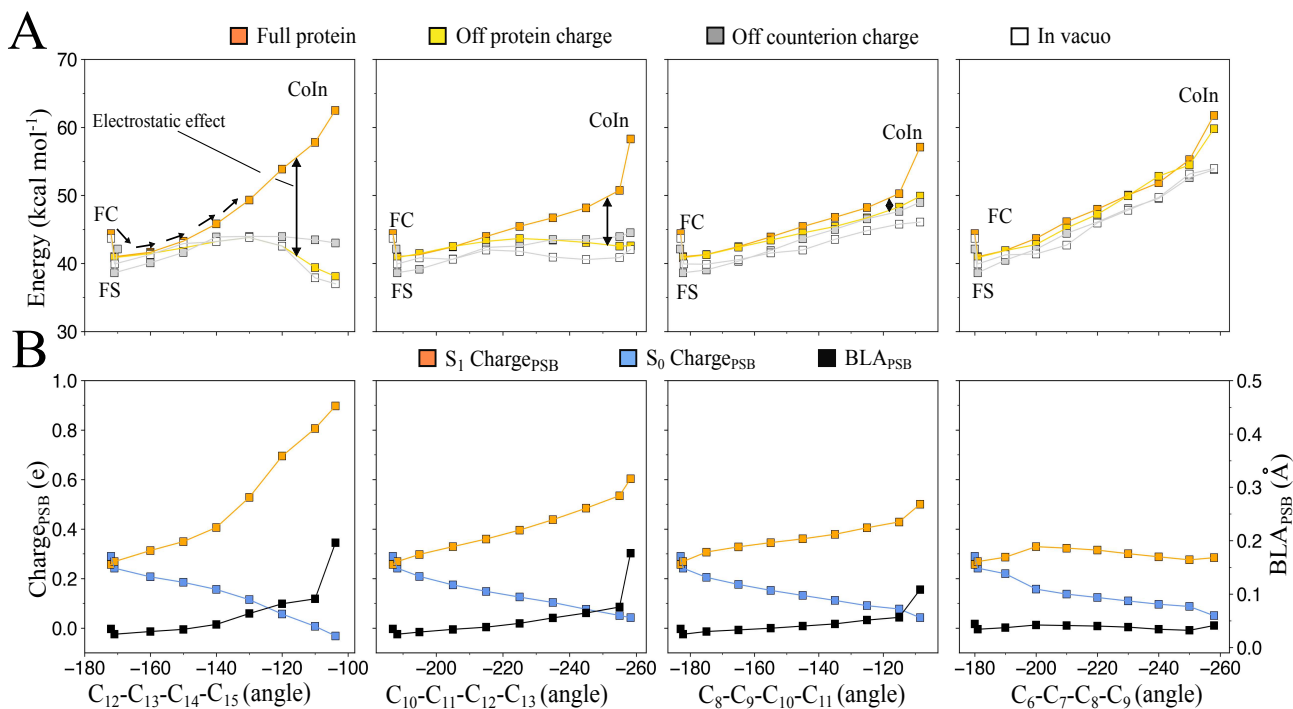


Figure 3.5: Energy, charge and BLA profiles along the photoisomerization of, from left to right, C13=C14, C11=C12, C9=C10 and C7=C8 double bonds in NeoR. A) CASPT2 S_1 energy profiles computed in presence of the full protein environment (orange squares), after setting to zero the charges of the entire protein (gold squares), after setting to zero only the charges of the E141 and after removing the whole protein in a full QM calculation (i.e., in vacuo, empty squares). B) Evolution of S_0 and S_1 Charge_{PSB} (the charge residing in the C14-C15-N-C ϵ rPSB moiety) and evolution of BLA_{PSB} (C14-C15 and C15-N bond lengths difference, see **Figure 3.1B**).

dihedral angle describing the C13=C14 double bond torsion connecting FS to the ca. 90° twisted CoIn giving access to S_0 . The energy maximum located along the MEP energy profile must therefore reflect the barrier height (i.e., $E_{S_1}^f$). Three additional MEPs, corresponding to the counterclockwise (CCW) isomerization of the C11=C12 and C7=C8 double bond and to the CW isomerization of the C9=C10 double bond, were also computed. The choice of the CW/CCW pattern for adjacent double bonds conforms to the well-known aborted bicycle-pedal motion [11], the archetypal space-saving reaction coordinate for the rPSB chromophore isomerization.

Figure 3.5 displays the CASPT2 S_1 energy profiles (top panels) together with the evolution of the S_0 and S_1 Charge_{PSB} and the BLA_{PSB} coordinates (bottom panels). As discussed in **Section 3.1.4** such charges were used as indicators of the weights of the COV and CT diabatic states in the adiabatic S_0 and S_1 energy profiles. All four S_1 isomerization paths pointed to the presence of a barrier (see orange energy profiles). In fact, in all cases the energy

monotonically increased from the FS state, with the CoIn corresponding to the highest point along the MEP. Therefore, $E_{S_1}^f$ magnitude was estimated as the energy difference between the CoIn and the FS states. $E_{S_1}^f$ is found of 21 (25), 17 (22), 16 (16), 20 (16) kcal/mol at the CASPT2 (XMCQDPT2) level for respectively the C13=C14, C11=C12, C9=C10 and C7=C8 isomerizations. Since the lowest energy barrier was still relatively large, the model supported a barrier-controlled mechanism for FS emission. Thus, NeoR would be an analog of the GFP-like fluorescent reporters [92, 11] as, in contrast with microbial rhodopsins such as bacteriorhodopsin [11], its relatively high barrier would induce slow internal conversion kinetics. This conclusion is consistent with the nanosecond excited state lifetime of NeoR while, most known fluorescent rhodopsins feature an excited state lifetime that does not exceed the picosecond range [52, 11, 100, 101, 2].

The calculations predicted lower energy barriers with respect to the canonical microbial C13=C14 isomerization, which appeared highly disfavored. Consequently, alternative photoisomerization pathways might originate, as it was recently corroborated by Sugiura et al. in the NeoR from *obelidium mucronatum*, sharing 78% of sequence identity with the NeoR studied in this work [102]. The result appeared also consistent with earlier findings of Cembran et al. [87], aimed at investigating the relationship between the position of an acetate counterion and the reactivity of a nearly isolated protonated polyene chain. They found that placing the counterion above the polyene favors the photoisomerization of the double bonds closest to the counterion; this result can be loosely associated with the favored C9=C10 and C7=C8 isomerization in the a-ARM_{E141} model.

The evolution of the Charge_{PSB} along the C13=C14 isomerization coordinate (see the orange curve in the corresponding panel of **Figure 3.5**) revealed that the S_1 weight of the COV diabatic state increased monotonically along the S_1 MEP until it dominates the region approaching the CoIn. This corresponds to confinement (or localization) of the charge in the small Schiff base moiety that, along the terminal part of the MEP (i.e., near the 90° twisted conformation), hosts a π -system orthogonal to the one residing along the rest of the rPSB conjugated chain. This was not a general behavior since depended on the isomerizing double bond. In fact, the C7=C8 MEP in **Figure 3.5** featured, along the entire S_1 profile, a steady mixed COV/CT character. These results pointed to a change in the origin of the critical $E_{S_1}^f$ barrier along different isomerization coordinates. More specifically, it was expected that the electrostatic effect imposed by the NeoR cavity may have different effects along different MEPs with a maximal effect on the canonical C13=C14 energy profile and a minimal effect on the C7=C8 energy profile. It was thus necessary to also evaluate steric effects.

In the a-ARM_{E141} model the electrostatic effect is due to the protein point charges including those describing the negatively charged E141 counterion. In order to assess the impact of such an effect on the isomerization energy profiles, these were reevaluated after setting to zero all protein point charges while keeping the geometrical progression unchanged (see gold energy profiles). Consistently with a dominant role played by the protein electrostatics, the slope in the S_1 profile associated with the C13=C14 coordinate was strongly decreased and even inverted from positive to negative in its last part. This effect was found gradually

reduced along the C11=C12 and C9=C10 coordinates (compare the vertical double arrows) and disappeared completely along the C7=C8 coordinate. Then, in order to disentangle the electrostatic and steric contributions to the computed $E_{S_1}^f$ value, the same energy profiles have been re-evaluated in the absence of a whole protein environment (see the energy profiles marked with empty squares). The results demonstrated that while in C11=C12 and C9=C10 MEPs the S_1 profile became completely flat, in C7=C8 MEP the S_1 energy barrier was only reduced but persisted, indicating a destabilization that originated from the rPSB geometrical progression. Notice that such progression was due, in all cases, to indirect electrostatic and steric effects determining the DA, FS, and CoIn geometries (i.e., determining the isomerization coordinate) and included the effect of the polarization of the rPSB π -electron density due to the counterion.

3.1.7 Fluorescence Enhancement Mechanism

In microbial rhodopsins the canonical S_1 isomerization produces the 13-*cis* rPSB chromophore. In general, as explained in **Section 1.3**, this is an ultrafast (sub-picosecond) reaction only allowing a negligible fluorescence emission from the FS state. In contrast, the results shown above were used to formulate a mechanistic theory for the fluorescence enhancement explaining how the NeoR electrostatics generates the high C13=C14 isomerization barrier of **Figure 3.5**.

Such theory takes the progressive confinement of the initially delocalized rPSB charge described above as the key event blocking the C13=C14 isomerization. Accordingly, Charge_{PSB} and BLA_{PSB} were used to follow the geometrical and electronic changes of the rPSB chromophore during the photoisomerization event. The first index displayed a monotonic charge increase from +0.27 to +0.90e (at FS and CoIn respectively), consistently with a monotonic increase of the positive charge on the Schiff base moiety. The second index pointed to a 0.03 to 0.20 Å change consistently with the reconstitution of a C=N double bond along the path and full localization of the charge on such a bond. Such progressive charge confinement was directly proportional to the increase in the electrostatic effect along the MEP of C13=C14 of **Figure 3.5** (i.e., the one indicated by the double-headed vertical arrow) and, therefore, to the energy increase leading to the large computed $E_{S_1}^f$ value. Consequently, it was proposed that the molecular mechanism driving the energy increase was the progressive increase in distance between the negative E141 counterion charge and the centroid of the confining charge. As illustrated in **Figure 3.6** the progressive positive charge confinement shifts the centroid of the positive charge away from the E141 residue, unavoidably leading to destabilization. This mechanism was supported by the computed decrease in electrostatic effect (i.e., again, the destabilization indicated by the vertical arrow) along the C13=C14, C11=C12 and C7=C8 MEPs of **Figure 3.10**. As an example, in **Figure 3.6A** it was also shown that the C9=C10 isomerization could not lead to the same electrostatic effect as, in this case, the charge did not get confined far from E141 but remained delocalized along the extended C10-C11-C12-C13-C14-C15-N moiety. This caused only a limited change in the counterion-chromophore interaction consistently with the computed decrease in electrostatic

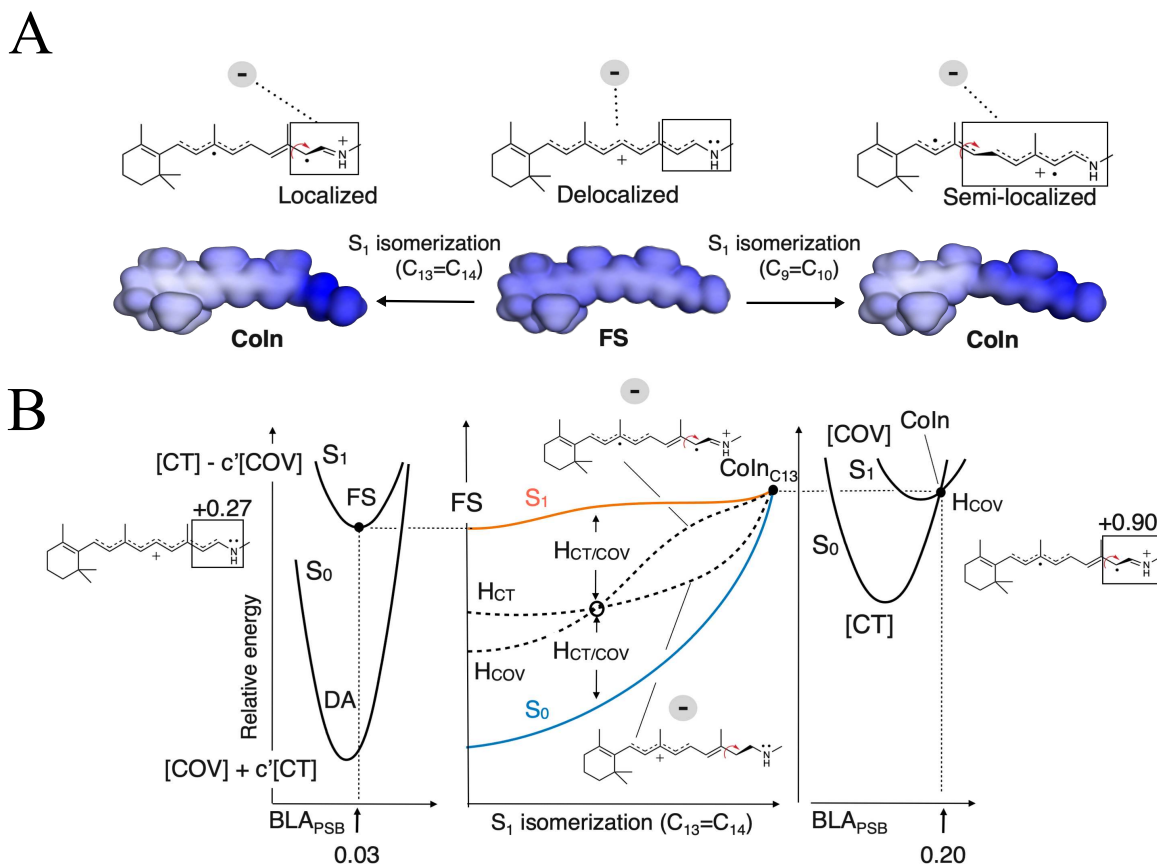


Figure 3.6: Origin of the $E_{f_{S_1}}$ barrier in NeoR. A) Representation of the increase in the distance between the negative and positive charge centroids due to the positive charge confinement along the C13=C14 and C9=C10 photoisomerization paths. Comparison between electrostatic potential (ESP) maps indicates that along the C9=C10 coordinate the extent of the confinement is less pronounced being the charge at the CoIn spread on a longer rPSB chromophore moiety (i.e., C10-C11-C12- C13-C14-C15-N). B) Proposed origin of the isomerization barrier in terms of COV energy (HCOV, dashed line) destabilization due to the charge confinement resulting from the mixed $[CT] - c'[COV]$ to the pure $[COV]$ electronic structure change along the S_1 adiabatic energy profile (in orange). The left and right panels display the shape of the S_1 and S_0 adiabatic potential energy curves along the BLA_{PSB} coordinate (C14-C15 and C15-N bond lengths difference, see **Figure 3.1B**) at FS (left) and CoIn (right). The BLA_{PSB} and α coordinates are substantially orthogonal.

stabilization. The charge confinement on the Schiff base moiety was thus critical.

The mechanism described above can be reinterpreted in terms of changes in the energy of the COV diabatic state (see HCOV in **Figure 3.6**) featuring a positive charge permanently located on the Schiff base moiety (see the resonance formula in **Figure 3.3A**). The highly

delocalized electronic structure of the FS pointed to close HCOV and HCT values consistent with the small Charge_{PSB} value of S_1 . On the other hand, the large S_1 Charge_{PSB} computed in the CoIn region pointed to an adiabatic state dominated by the COV diabatic. Thus, the charge confinement effect described above could be translated into an increase in the weight of the COV diabatic state along the isomerization coordinate. This justifies the steep S_1 energy increase in the CoIn region that would originate from the simultaneous increase in the COV weight and COV destabilization. Such a delocalization-confinement mechanism (i.e., without an initially delocalized charge there is no progressive confinement) also suggested a lesser sensitivity of the S_0 state to the progression along the isomerization coordinate. In fact, while S_0 became progressively dominated by the CT diabatic state, in CT the rPSB charge remained relatively unconfined along a long segment (-C7-C8-C9-C10-C11-C12- C13-) of the rPSB backbone even in the CoIn region.

3.1.8 Resonance Raman Simulation

This section focuses on resonance Raman (RR) spectroscopy as a critical benchmark for the characterization of the electronic structure of NeoR. RR spectroscopy has proven to be a powerful tool for investigating the rPSB chromophore geometrical and electronic structure in both microbial [103, 104, 105, 106, 107, 108] and animal [109, 110] rhodopsins. In fact, the main observables of this technique, band frequencies and their intensities, are shaped by the details of the S_0 and S_1 characters of rPSB.

Below, both the experimental and simulated RR spectra of NeoR are provided and discussed. This comparison served for two primary purposes. Firstly, it was instrumental for a more robust RR band assignment, specifically identifying the nuclear modes responsible for each experimentally observed RR band, and, secondly, it has served as an additional validation of the QM/MM model of NeoR discussed above.

The RR of NeoR spectra measured at 80K with 1064 nm of excitation to achieve sufficiently strong resonance enhancement, such that the spectra exclusively display the RR bands of the rPSB without any interference from the protein, is shown in **Figure 3.7** (red curve). This exhibited a complex band pattern between 900 and 1600 cm^{-1} characterized by ca. 20 partially overlapping bands at strong and medium intensities. Two dominant bands were observed in the ethylenic band region at 1553 and 1586 cm^{-1} , with the latter having the highest intensity among all the spectra. Notably, no band attributable to the Schiff base (C15=N) stretching were detected. The simulated spectra has shown a remarkable band profile agreement (**Figure 3.7**, green curve), having two intense band at 1569 and 1594 cm^{-1} attributable to the experimental one at 1586 cm^{-1} and one additional intense band at 1554 cm^{-1} attributable to that at 1553 cm^{-1} . Additionally, the simulation predicted null intensity for the band assigned to the Schiff base stretching at 1646 cm^{-1} (for the list of the frequencies, intensities and internal coordinate descriptions of the normal modes see **Appendix**). Such agreement has allowed for the assignment of the normal modes. It resulted that the intense experimental band at 1553 and 1586 cm^{-1} were due to normal modes largely characterized by the C1=C6, C7=C8 and C9=C10 double bonds and C6-C7, C8-C9 and C10-C11 single

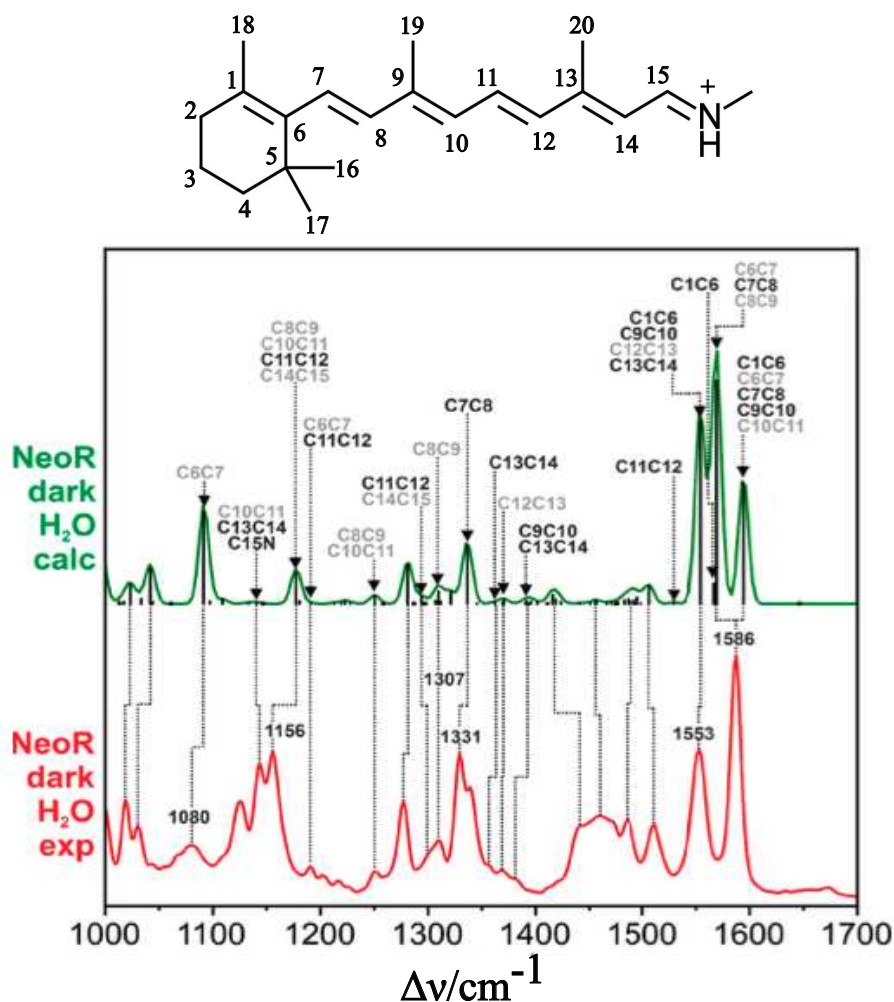


Figure 3.7: Calculated and experimental resonance Raman (RR) spectrum in H₂O. The calculated spectrum was generated using a gaussian band shapes of the vibrational transition derived from the QM partition of the NeoR QM/MM model (green trace). The spectrum also includes the line presentation of the vibrational transitions. Frequencies were scaled by a factor of 0.9. Modes with major C=C and C-C stretching characters (factoring to >50% of the contribution of a coordinate) are indicated by black and gray letters, respectively. The notation follows the numbering of the structural formula on top of the figure. Details of the calculated rPSB modes are listed in **Appendix**.

bonds stretching modes. Furthermore, it was noticed that such stretching modes describe the normal modes associated to the bands with the RR highest intensity. On the contrary, normal modes described by to stretching modes closer to Schiff base were predicted with a lower and almost null intensity.

This behaviour can be expected on the basis of the Albrecht's mechanism [111, 112] for RR spectroscopy that associate the highest intensity to the stretching coordinates that display

the highest excited-state displacements. In this regard, **Figure 3.4** shows indeed that the main geometrical displacements between the DA and FS were exactly located along the bond lengths in the rPSB region close to the β -ionone ring. On the contrary, these geometrical differences became null when approaching to the Schiff base moiety, culminating with a bond length difference between DA and FS (Δ_d) nearly null for the C15=N double bond, whose stretching mode was indeed predicted with a null RR intensity (see above).

In summary, this study added further experimental validation for the unusual rPSB electronic structure of NeoR predicted by the QM/MM model. The fact that the model was able to reproduce the observed RR spectral profile implies that the chromophore S_0 and S_1 electronic structures were well described. Thus, in contrast to canonical microbial rhodopsins, the NeoR model seems to exhibit a rPSB chromophore having a S_0 state with strong CT character, resulting into an unusual alternating BLA pattern only in the region far from the Schiff base (i.e. from C1=C6 to C9=C10).

3.1.9 Conclusions

Above we have shown that, the conventional C13=C14 photoisomerization event of microbial rhodopsin, can be blocked or slowed down by a suitable change in the electrostatic environment of the all-*trans* rPSB chromophore. More specifically, the presented a-ARM_{E141} model, indicated that NeoR could be the product of an evolutionary process driven by the translocation of a negatively charged residue away from the chromophore Schiff base region. The main local effect of this process was the generation of a DA state featuring a delocalized rPSB positive charge. The rest of the described properties, including spectral properties such as the large bathochromic shift, small Stokes shift, and sizable S_1 isomerization energy barrier along the canonical C13=C14 torsional coordinate, were a consequence of such a change. This behavior was recently documented by El-Tahawy et al. in isolated rPSB chromophores subject to homogeneous, strongly negative red-shifting electric fields.

The effect associated with the repositioning of the rPSB counterion was previously proposed to explain the λ_{max}^a changes observed in a set of rhodopsin mimics based on the human cellular retinol binding protein II (hCRBPII) [41]. Similar to this NeoR model, the members of the set displaying a large red shift were found to be associated with a counterion located far from the Schiff base moiety. This conclusion was reached through X-ray crystallographic analysis supporting the hypothesis that a red-shifted λ_{max}^a must be associated with an even distribution of the iminium charge along the chromophore π -conjugated chain. Such delocalization can be associated with the rPSB delocalized charge seen in the a-ARM_{E141} calculation and interpreted as a COV/CT resonance hybrid (see **Figure 3.3A**) or, in a different language, to a near cyanine limit situation [113].

a-ARM_{E141} leads to a possible general principle for the engineering of other highly fluorescent rhodopsins that we call “delocalization-confinement”. Such principle establishes that the electrostatic field generated by the cavity, for instance via a specific counterion localization, must yield a vastly delocalized geometrical and electronic structure of the rPSB conjugated chain in both the DA and the FS state of the protein. In this condition, an electrostatically

induced high reaction barrier can be generated via a rPSB charge confinement process occurring, unavoidably, along the canonical C13=C14 isomerization path in the region entering the corresponding CoIn channel.

Finally, the reported results provided evidence that the spectroscopy of retinal proteins is regulated by the same principles regulating GFP-like fluorescence. More specifically, it was proposed that the GFP variants achieving maximal π -electron delocalization (called the cyanine limit) were the ones where a COV and CT configurations of the protein chromophore have exactly the same weight, thus pushing the λ_{max}^a value to the extreme red and culminating in a null Stokes shift.

3.2 Photoisomerization Quantum Yield Modulation in the Heliorhodopsin Family

This section, based on the findings of **Paper 3**, is focused on the study of the ϕ_{iso} in one member of the recently discovered HeR family, expressed in the archaea *thermoplasmatales archaeon*. With the exception of a few insights from transient absorption spectroscopy (TAS) studies, the exact mechanism of the photoisomerization in the diverse heliorhodopsin (HeR) family remains largely unknown. For this reason, and in order to provide an experimentally consistent atomistic interpretation of the HeR photoisomerization, this chapter presents the result of hundreds of picosecond quantum-classical trajectory computations capable of reproducing the experimentally observed dynamics outcome and timescale. By doing this it will be shown a diverse and previously unreported photoisomerization mechanism where a microbial-like and animal-like regiochemically different photoisomerization motions coexist and compete. Significantly, the simulations reveal a previously undescribed ϕ_{iso} control mechanism based on the existence of an unreactive animal-like excited state decay channel that subtracts molecules to the experimentally observed 13-*cis* to all-*trans* channel, ultimately reducing its efficiency.

3.2.1 Abstract

*Rhodopsins represent a large group of light-responsive proteins counting approximately 500.000 members divided into two vast and evolutionary distinct superfamilies. In all members the function is invariably triggered by the photoisomerization of a single retinal chromophore. In 2018, a third and widespread family, named as heliorhodopsins, was discovered using functional metagenomics in lake Kinneret (Israel) and readily detected globally in both freshwater and saltwater environments. Heliorhodopsins, with their markedly different structural features with respect to the animal and microbial superfamilies, has offered a new opportunity to study how evolution has exploited the reaction of photoisomerization to achieve adaptation. One important question is related to the mechanism of such a reaction and how it differs from that of animal and microbial rhodopsin homologues. In order to address the problem, hundreds of quantum-classical trajectories were used to simulate the spectroscopically documented picosecond light-induced dynamics of an heliorhodopsin member from the archaea thermoplasmatales archaeon (TaHeR). It will be shown that, consistently with the observations, the trajectories has revealed the existence of two excited state decay channels. However, inconsistently with previous hypotheses, only one channel is associated to the -C13=C14- isomerization typical of microbial rhodopsins while the second channel is characterized by the -C11=C12- isomerization typical of animal rhodopsins. The fact that such -C11=C12- motion is found aborted upon decay and ground state relaxation, explains why illumination of TaHeR only produces the 13-*cis* isomer with a low quantum efficiency. It is argued that the documented lack of regioselectivity in double-bond excited state twisting motion is the result of an “adaptation” that could be completely lost via specific residue substitutions modulating the steric hindrance*

experienced along the isomerization motion.

3.2.2 Introduction

As discussed in **Section 1.3**, rhodopsins employ the photoisomerization of the rPSB chromophore to start specific photocycles associated with various biological functions across all the kingdoms of life. The study of the diversity of the rPSB isomerization and its efficiency has represented therefore an unavoidable milestone along the path to exploit rhodopsin functions. More specifically, it has become apparent that the investigation of the molecular-level factors controlling ϕ_{iso} [42, 86] in diverse rhodopsins would have an impact on the design of effective optogenetic tools [114, 96] for neurosciences [115, 116, 25] and of novel synthetic biology circuits allowing, for instance, for artificial CO₂ fixation [117, 9]. More globally, the investigation of such factors is also needed for understanding why in photic zones such as the upper ocean [118], specific microbial rhodopsins massively contribute to the solar energy capture [119]. Nevertheless, despite the above facts, only a limited number of studies have discussed or focused on the rPSB isomerization diversity in both animal [120, 121, 122] and microbial [123, 124, 125, 126, 127, 128] rhodopsins.

The discovery of heliorhodopsins (HeRs) in lake Kinneret (Israel) and their subsequent detection in various organisms across different aquatic environments worldwide, as discussed in **Section 1.5**, significantly expanded the known rhodopsin diversity [10]. This discovery has offered a new chance to study of how organisms has exploited the rPSB photoisomerization for targeting specific biological functions and adaptation to diverse environments. Notably, the ϕ_{iso} of HeRs was found not exceeding 0.20 [129], in striking contrast to the 0.64 value for the canonical proton-pump bacteriorhodopsin [30] as well as the 0.67 value for dim-light visual pigments [5].

These differences have suggested a distinct photoisomerization mechanism and, in order to shed light on it, the results of quantum-classical population dynamics performed on one member of the HeRs, namely TaHeR, are presented below. More specifically, the reaction was studied by constructing, starting from the crystallographic structure (PDB ID: 6IS6) [6], an hybrid quantum mechanics / molecular mechanics (QM/MM) model of the protein and using it to propagate hundreds of Tully’s quantum-classical trajectories for over two picoseconds. The results were compared with those of rPSB photoisomerization reaction occurring in the dim-light visual rhodopsin expressed in the vertebrate *bos taurus* (Rh) as an example of highly efficient 11-*cis* to all-*trans* photoisomerization of a canonical animal rhodopsin.

As pictorially shown in **Figure 1.5**, these calculations have revealed that while in Rh the rPSB photoisomerization is accomplished through a single coordinate describing the counter-clockwise (CCW) π -bond breaking of the C11=C12 double-bond typical of animal rhodopsins, in TaHeR the isomerization motion follows simultaneously two coordinates corresponding to the clockwise (CW) and CCW π -bond breaking of the C13=C14 and C11=C12 double-bonds, respectively. As shown in the following sections this result has two consequences; firstly, it provides an atomistic reinterpretation of recent TAS measurements [129], which pointed to the existence of two ultrafast decay channels, but both assigned to C13=C14 isomerization.

Secondly, it reveals a previously unreported and effective way in which HeRs control ϕ_{iso} . The same quantum-classical simulations have showed that the absence of the C11=C12 isomerization product in TaHeR is explained by a ground state mechanism that selectively inverts the animal-like isomerization motion initiated in the excited state, leading to reconstitution of the original all-*trans* rPSB reactant (see **Figure 1.5B**, left panel).

3.2.3 Model Construction and Validation

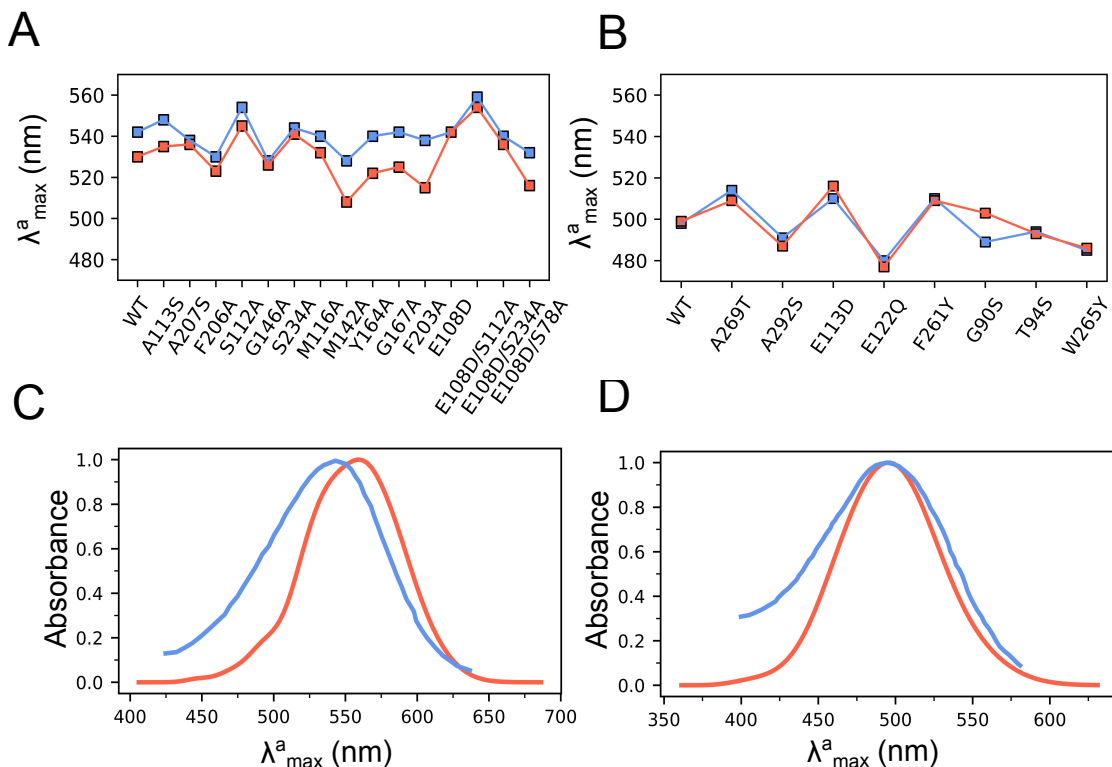


Figure 3.8: TaHeR and Rh models validation. A) Comparison between computed (in orange) and observed (in blue) λ_{max}^a values, calculated as $\Delta E_{S_0-S_1}$ for TaHeR and a set of mutants. B) Same as point A for Rh and related mutants. Comparison between computed (in orange) and observed (in blue) absorption bands (in arbitrary units) for C) TaHeR and D) Rh. The experimental bands were digitalized from literature data [5, 6].

As a first step toward the investigation of the TaHeR and Rh photoisomerization events, the focus was on the validation of the two QM/MM models generated using the *a*-ARM protocol discussed in **Section 2.3.1**. To do so, it was checked whether the QM/MM models were able to reproduce the experimentally observed differences in λ_{max}^a . Moreover, as shown in **Section 3.2.4**, a further validation was also provided by the agreement between observed and computed ϕ values.

3.2. PHOTOISOMERIZATION QUANTUM YIELD MODULATION IN THE HELIORHODOPSIN FAMILY

As shown in **Figures 3.8A and 3.8B**, the constructed TaHeR and Rh models yielded $\Delta E_{S_1-S_0}^a$, expressed in terms of λ_{max}^a , only +12 nm blue shifted (530 nm) and -1 nm red shifted (499 nm) with respect to the experimentally observed values, respectively. Furthermore, by using the wild type QM/MM models as starting point for constructing a set of *a*-ARM mutant models, it was also possible to reproduce the experimental trend in λ_{max}^a displayed by a set of single mutants of Rh, and a set of single and double mutants of TaHeR (again, like occurred in NeoR in **Section 3.2**, in most cases with a blue-shifted error typical of *a*-ARM models). Notably, the glutamic-to-aspartic acid mutation at the 108 counterion position of TaHeR yielded identical λ_{max}^a value with respect to the wild-type, again in line with the observation [130]. Also, it is worthwhile mentioning that in the TaHeR model the protonation state of the two peculiar His23 and His82 residues is coherent with its proton release pathway (see **Section 1.3.2**). The agreement between computed and experimentally observed λ_{max}^a supported the use of the two QM/MM models for simulating the corresponding room-temperature Boltzmann like distributions, and the subsequent simulation of the absorption bands (**Section 2.4.1** gives a description of the adopted protocol). The computed absorption bands, shown in **Figures 3.8C and 3.8D**, confirmed the quality of the distribution represented by 200 geometries and velocities.

3.2.4 Excited State Dynamics and TAS Assignment

In this section, as schematically introduced in **Figure 1.5**, the dramatically different time progression of the two photoexcited room-temperature populations, is reported. This appeared ballistic and coherent in Rh but slower and more complex in TaHeR. This is shown in **Figures 3.9A and 3.9B** where the time evolution of α and α' , the isomerizing reaction coordinates, are displayed for each trajectory. From figures comparison, it becomes apparent that while the S_1 depopulation of TaHeR was not completed after 1.5 ps (only 93 out of 200 hopped to S_0), Rh exhibited a full S_1 depopulation in less than 0.2 ps. This difference is already evident at < 0.05 ps timescales; when most Rh trajectories exhibited a large CCW C11=C12 twist (ca. -70° on average) while no TaHeR trajectories displayed a rPSB double bond twist larger than $\pm 20^\circ$. After this initial phase, all 200 Rh trajectories decayed to S_0 at a single CoIn₁₁ channel and reconstituted the C11=C12 π -bond in either the CW or the CCW directions. This did not occur in TaHeR where two CoIn₁₁ and CoIn₁₃ channels, featuring a distinct regiochemistry and stereochemistry, were populated and coexisted for the entire simulation time. CoIn₁₃ (see red and blue circles in **Figure 3.9A** respectively) was dominated by the CW twisting of the C13=C14 double bond. CoIn₁₁ (see blue squares in **Figure 3.9A**), on the other hand, was dominated by the CCW twisting of the C11=C12 double bond, in full analogy with Rh. Notably, the two TaHeR competing channels yielded nearly identical decay counts (52 and 41 decays, respectively, after 1.5 ps of simulation time).

The computed timescales were found qualitatively consistent with the observed TAS dynamics for both TaHeR [129] and Rh [131], thus providing a further validation of the constructed QM/MM models. For TaHeR, TAS had shown that the $S_1 \rightarrow S_0$ decay is completed after several ps (> 8.5 ps) and exhibits a multi-exponential decay including fast (< 1 ps)

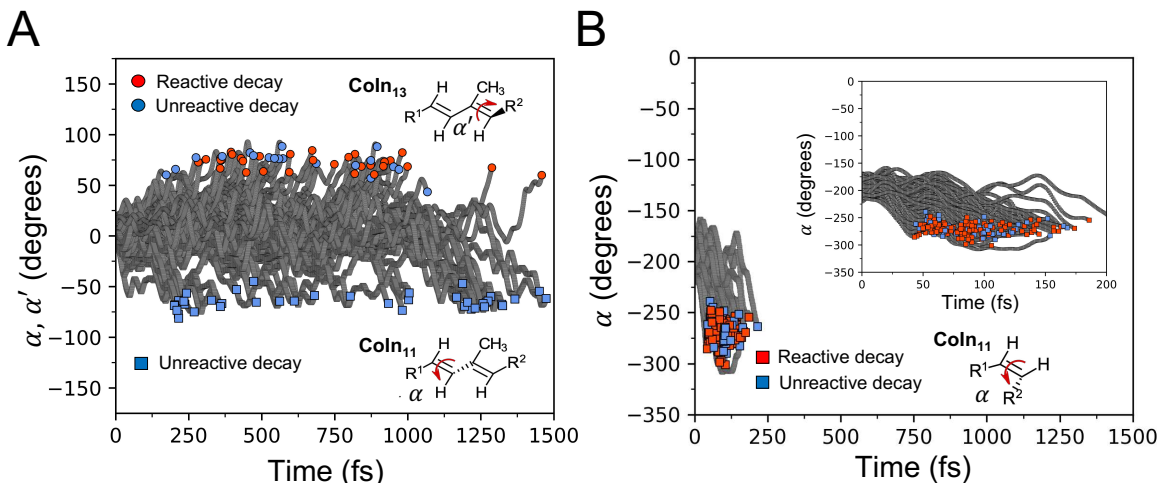


Figure 3.9: Overview of the S₁ dynamics of A) TaHeR and B) Rh, simulated with 200 quantum-classical trajectories. Unreactive S₁ trajectories are marked by blue symbols placed at the time of the S₁/S₀ decay (hopping time). Similarly, reactive trajectories are marked by red symbols. As described in the main text and expanded in **Figure 1.5**, a trajectory is termed as reactive when, after the hopping, continues to propagate on S₀ in the direction of the photoproduct. In contrast, unreactive trajectories revert their direction and propagate in the direction of the reactant. As shown in **Figure 1.5** and here in the molecular sketches, α and α' indicate the dihedral angle associated to the isomerizing double bond, C11=C12 and C13=C14 in TaHeR and C11=C12 in Rh.

and slow (> 1 ps) components. However, these observations were interpreted by assuming the C13=C14 rotation as the sole excited state process. Therefore, the fast as well as slower components were associated to a multi-exponential process where the subpicosecond formation of the 13-*cis* photoproduct begins ca. 0.2 ps after photoexcitation, while slower decay components were attributed to “unreactive processes” generating less or no photoproduct. Instead, our TaHeR simulation indicated that the observed fast and slower decay components are the result of the coexistence of the two S₁ isomerization channels described above. Our calculations associated the observed fast and slow excited state traces of TaHeR with the presence of S₁ isomerizations with distinct dynamics, regioselectivity, stereoselectivity and efficiencies. Indeed, although in our trajectories (**Figure 3.9A**) earlier decays were observed approximately after 0.2 ps, the majority of CoIn₁₃ decays took place within 1 ps (only 3 out of 52 trajectories decayed after 1 ps), while along CoIn₁₁ the trajectories decayed more sparsely on a longer timescale. In striking contrast, the simulated Rh dynamics that fully complies with TAS measurements [131, 132], has shown that the rPSB photoisomerization occurs, substantially, through a single mono-exponential decay (disregarding the very inefficient formation of isorhodopsin featuring a C9=C10 rotation [133]). Thus, in TaHeR the low photoisomerization ϕ value would be the result of the lack of isomerization regioselectivity

3.2. PHOTOISOMERIZATION QUANTUM YIELD MODULATION IN THE HELIORHODOPSIN FAMILY

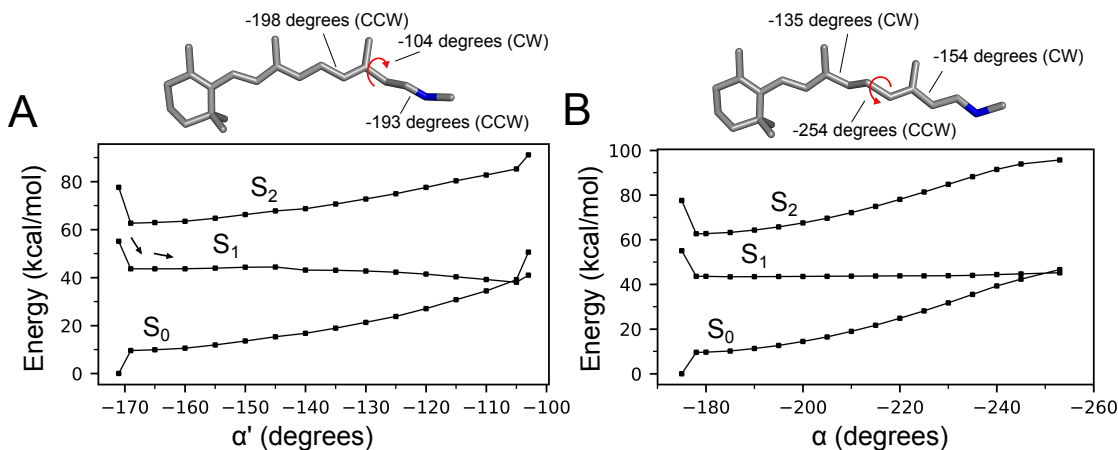


Figure 3.10: Minimum energy path computed for TaHeR describing A) the clockwise (CW) photoisomerization around the C13=C14 double bond (α' coordinate) and B) the counterclockwise (CCW) photoisomerization around the C11=C12 double bond (α' coordinate). Each profile shows S_0 , S_1 and S_2 energy profiles computed at the 3 root state average CASPT2/CASSCF(12,12)/6-31G*/AMBER94 energy profiles. On top, view of the two located minimum energy CoIn's displaying the relevant torsional deformations. To notice how both reaction path are associated to shallow S_1 profiles.

during S_1 progression combined with an unreactive animal-like C11=C12 decay at CoIn₁₁. In fact, the 41 trajectories decaying through the CoIn₁₁ channel have led to reconstitution of the original all-*trans* rPSB exclusively (see blue squares, **Figure 3.9A**). Conversely, the CoIn₁₃ channel was found highly “reactive”, as out of the 52 trajectories evolving along this path, 33 led to photoproduct formation (see red circles, **Figure 3.9A**) with 13-*cis*,15-*trans* (32 trajectories) or 13-*cis*,15-*cis* configurations (1 trajectory), as observed experimentally [6], and the remaining 19 relaxed back to the all-*trans* rPSB configuration (see blue circles, **Figure 3.9A**). Based on the assumption that the trajectories left on S_1 never decay productively (see above), a ϕ value of 0.17 was computed (33 reactive trajectories over 200) of TaHeR, which is close to the experimentally observed value of 0.19 [129]. Concerning Rh, all the trajectories displayed $S_1 \rightarrow S_0$ decays within ca. 0.1 ps (first 49 fs last 218 fs). The computed ϕ value was of 0.70, again in line with the observed 0.67 [5].

The computation of the S_1 minimum energy paths (MEPs) associated to the TaHeR channels, as illustrated in **Figure 3.10**, has also revealed that the two reaction paths have comparable, and nearly null, excited state photoisomerization energy barriers ($E_{S_1}^f$). More specifically, the two paths were achieved through reaction coordinates spanning very flat S_1 PES regions from -180° to -220° along α and from -170° to -130° along α' . These MEPs are very different from the steep MEP reported for Rh [134] but consistent with the longer S_1 dynamics experimentally observed for TaHeR [129].

3.2.5 Aborted Bicycle-Pedal Isomerization Motion

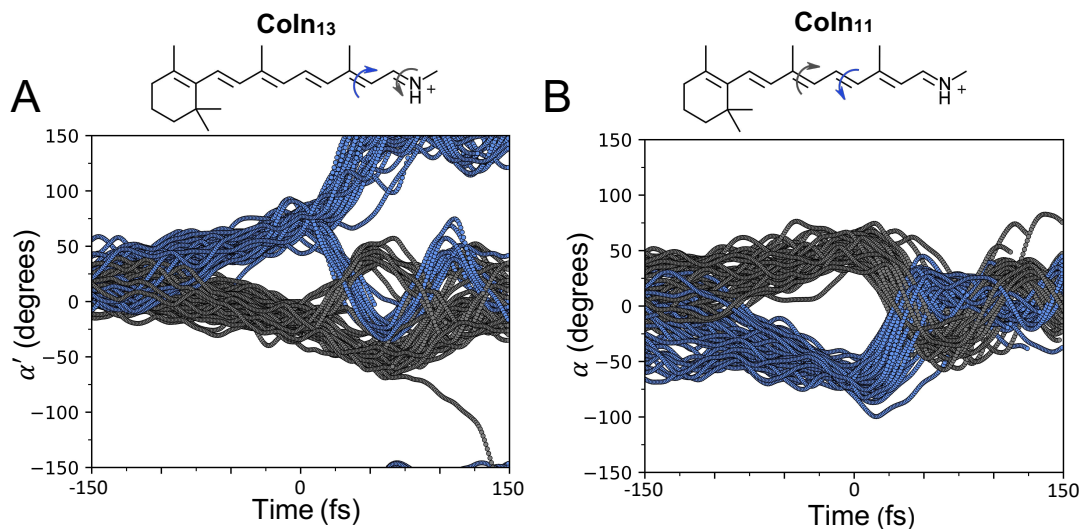


Figure 3.11: Time evolution of the aborted bicycle-pedal photoisomerization coordinates (ABP) of TaHeR. A) Representation of the evolution of C12-C13=C14-C15 (i.e. α' , in blue) and C14-C15=N-Ce (in gray) dihedral angles for trajectories along the CoIn₁₃ channel aligned at the $S_1 \rightarrow S_0$ decay time ($t = 0$ fs). B) Representation of the evolution of C10-C11=C12-C13 (i.e. α , in blue) and C8-C9=C10-C11 (in gray) dihedral angles for trajectories along the CoIn₁₁ channel aligned at the $S_1 \rightarrow S_0$ decay time ($t = 0$ fs).

This section aims to illustrate the conservation in TaHeR (and thus inside the HeR family) of a variant of the bicycle-pedal isomerization mechanism, originally proposed by Warshel for Rh [135], known as aborted bicycle-pedal (ABP) motion. This motion has been computationally demonstrated to be active in both animal and microbial rhodopsins [134, 133, 136, 99], and more recently experimentally confirmed through time-resolved XFEL studies on Rh [137]. The ABP coordinate is clearly displayed at the population level in **Figure 3.11**, reporting the evolution of critical dihedral angles along the two TaHeR channels. Accordingly, in CoIn₁₃, the CW isomerization of the C13=C14 bond was found coupled with the opposite CCW rotation of the C15=N bond. Upon decay to S_0 , however, this coupled bicycle-pedal motion was interrupted as C15=N reverted its rotation, resulting in the isomerization at the C13=C14 exclusively. To notice, however, the presence of one trajectory that exhibited a full bicycle-pedal mechanism, yielding the 13-*cis*, 15-*cis* photoproduct. The right diagram is related to the CoIn₁₁ channel and demonstrates that, in contrast with CoIn₁₃, both C9=C10 and C11=C12 rotations inverted directions immediately after decay to S_0 such that no net isomerization took place.

3.2.6 Quantum Yield Modulation

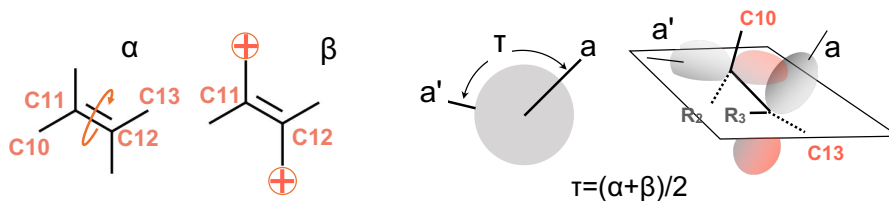


Figure 3.12: Schematic representation of α , β and τ for the C11=C12 moiety of rPSB chromophore. The same representations apply to α' and β' and the corresponding τ' parameter for the C13=C14 bond moiety. τ and τ' are geometrical parameter proportional to the π -overlap (roughly, the dihedral between the axes a and a' of the p-orbitals forming the π -bond) of the corresponding double bond.

In this section the mechanism responsible for the experimentally observed C13=C14 selectivity (i.e. only the C13=C14 isomerization was detected) and low ϕ_{iso} value of TaHeR in the presence of a non-regioselective S_1 isomerization dynamics will be analyzed. This is done using a theoretical framework based on τ : a function of the skeletal torsion α and substituents wag β of the isomerizing double bond described in **Figure 3.12**.

Past and recent studies on Rh [42, 86] have demonstrated the pivotal role of the geometrical parameter *tau* at the moment of the S_1 decay ($d\tau/dt^{decay}$) for the mechanistic interpretation of reaction quantum efficiencies in Rh. More specifically, it has been established that in the CCW isomerization the $d\tau/dt^{decay} < 0$ provides a statistically valid prediction of the corresponding computed ϕ_{iso} value. When considering the stereochemical change in rotation direction, in a CW isomerization such a prediction would be based in the fraction of trajectories decaying with $d\tau/dt^{decay} > 0$. These “rules” are connected and justified by the canonical Landau-Zener (LZ) model for a reactive non-adiabatic transition [138]. In fact, the LZ model relates the probability of non-adiabatic transitions leading to the product (i.e. the probability that after S_1 decay the system would generate a product) to the nuclear velocity along the “local” reaction coordinate.

A comparison of **Figures 3.13A** and **3.13B** demonstrates that such sign rule is valid for the C13=C14 isomerization (i.e. for decay through the CoIn₁₃ channel) of TaHeR but it is not valid for the C11=C12 isomerization (i.e. for decay through the CoIn₁₁ channel), in deep contrast with the reported validity in the case of the CoIn₁₁ channel of Rh. More specifically, in the top panel of **Figure 3.13A** it can be seen that the vast majority of reactive trajectories (red circles) decayed with $d\tau/dt^{decay} > 0$ (upper half of the panel). In deep contrast, **Figure 3.13B** shows that irrespective of the sign of $d\tau/dt^{decay}$ the trajectories were all unreactive. The lower panels of **Figures 3.13A** and **3.13B** report an analysis of the $d\tau/dt^{decay}$ in the vicinity of the TaHeR decay regions. By focusing on a time window of 40 fs centered on the decay event, it is shown that τ velocities behaved differently along the paths leading

to the two distinct channels. In the case of trajectories pointing to the CoIn₁₃ channel, the population displayed a positive average velocity that appears to remain constant until the decay region is reached. This would translate into a constant velocity for the breaking of the π -bond as pictorially illustrated in **Figure 3.13C**. In contrast, consistently with the illustration of **Figure 3.13D**, the corresponding motion towards CoIn₁₁ displayed decreasing τ velocities suggesting formation of the 11-*cis* product. However, about 10 fs before decay, a deceleration occurred, and the trajectories reached the decay region with rapidly decreasing velocities and yielding only small negative values at decay.

Figures 3.13A and **3.13B** also provided information on the TaHeR π -bond making dynamics immediately after the decay. It can be seen that, at CoIn₁₃, the percentage of reactive trajectories generated at decay (63%) was relatively close to that of $d\tau'/dt^{decay} > 0$ (74%). This does not happen in the case of the unreactive CoIn₁₁ channel (0%) where, despite a still large (49%) fraction of $d\tau'/dt^{decay} < 0$, $d\tau'/dt$ displayed, immediately after the decay, large and positive values consistent with the resulting reactant reconstitution. This behavior suggests that, in correspondence of the $d\tau/dt$ decrease seen about 10 fs before the hopping points, the C11=C12 rotation or, better, the bicycle pedal nuclear motion, encounters a restraining force effectively inverting the motion from CCW to CW.

In order to provide a molecular-level understanding of the lack of reactivity of the CoIn₁₁ channel of TaHeR, it is hypothesized that the restraining force mentioned above originates from steric and/or electrostatic interactions. Inspection of the TaHeR trajectories reveals that the C11=C12 rotation led to strongly decreased distances between the plane associated with the bicycle-pedal coordinate and the Phe203, Phe206 and Gln143 residue sidechains (see **Figure 3.14A**). This effect appears to lead to the “collision” of the rotation of the substituents at C10 and C11 with the π -system of Phe206 and the -NH₂ aminic group of Gln143 causing compenetration of VdW radii [139] and, thus, strong repulsion. As anticipated above, these effects may systematically invert the negative sign of $d\tau/dt$ before and after the decay pointing to π -bond reconstitution via CW, rather than CCW, rotation. In contrast, the surround of the C13=C14-C15=N moiety appears more “spaced-saving”. In fact, the rotation of the plane associated to the corresponding bicycle-pedal coordinate is less impaired (compare **Figure 3.14A** with **Figure 3.14B**) as demonstrated by the ≈ 2.5 Å distance kept by the C14 and C15 substituents with the closest residue sidechains during the entire progression toward the CoIn₁₃ channel.

3.2.7 Conclusions

Above it was reported on the first atomistic description of the primary event in the photocycle of a member of the recently discovered HeR family. Since the employed TaHeR *a*-ARM QM/MM model reproduced the observed spectroscopic parameters and measured ϕ_{iso} , it was used to simulate the photoinduced dynamics and reinterpret previously proposed mechanistic hypothesis [129]. The simulations have revealed an equal population of two regiochemically and stereochemically distinct S₁ isomerization channels. The first C13=C14 channel is found in substantially all microbial rhodopsins, while the second C11=C12 channel is related to

3.2. PHOTOISOMERIZATION QUANTUM YIELD MODULATION IN THE HELIORHODOPSIN FAMILY

the photoisomerization of animal rhodopsins but in the reversed all-*trans* to 11-*cis* direction. Since TaHeR is phylogenetically distant from both animal and microbial rhodopsins, the resulting promiscuous photoisomerization provided an example of a high isomerization diversity induced by diverse opsin cavity compositions. A second result was the finding of a new ϕ_{iso} control mechanism in TaHeR, not based on a low ϕ_{iso} but, rather, to the lack of S₁ regioselectivity with respect to an equally populated fully inefficient C11=C12 S₁ isomerization channel. In fact, the suppression of the C11=C12 channel would allow, in principle, an increase of ϕ_{iso} from 0.17 to 0.63. A third result was the confirmation of the universality, i.e. its validity also for a HeRs member, of the space-saving ABP mechanism initially documented for Rh [134], as well as of the validity of the theory based on τ velocity at the point of S₁/S₀ decay as the determinant of ϕ_{iso} [86, 42].

Isomerization mechanisms featuring coexisting isomerization channels might operate in certain microbial rhodopsins but these have not been characterized at the atomistic level. For instance, although the extreme distance between HeRs and microbial rhodopsins make any comparisons difficult, past and recent studies on chloride pump halorhodopsin [140, 141], proteorhodopsin [124, 142, 143], sensory rhodopsin I (SRI) [144], channelrhodopsin chimera C1C2 [145], sodium ion pump rhodopsin KR2 [146] as well as bacteriorhodopsin under acidic conditions [124], have investigated the factors that govern their observed multicomponent excited state decay. Similar to TaHeR, these microbial rhodopsins display an S₁ dynamics well described by fast (subpicosecond) and reactive, and slow and unreactive decay components. Two types of mechanisms were proposed to rationalize the findings: a) the potential heterogeneity of ground state species arising either from the acid-base equilibrium of the rPSB chromophore counterion or the presence of a mixture of chromophore isomers and b) the possibility of reaction branching on S₁. The calculation shown above supports the latter mechanism in the case of TaHeR. In fact, for TaHeR, the presence of different chromophore isomers in the DA state can be excluded on the basis of HPLC measurements that document a dominant presence of all-*trans* rPSB. Furthermore, the TAS measurements are carried out at neutral pH that should ensure the deprotonation of the rPSB counterion (Glu108) [129].

A branched isomerization mechanism has been experimentally established for a microbial rhodopsin from the archaea *haloquadratum walsbyi* called middle rhodopsin (MR) [144] where a mixture of 13-*cis* and 11-*cis* configurations is produced upon illumination starting from all-*trans* rPSB. It is possible that the isomerization mechanism of MR is essentially the same described above for TaHeR but having the all-*trans* to 11-*cis* isomerization channel corresponding to CoIn₁₁, at least partially reactive. The authors of such research proposed that MR might represent “a missing link” in the evolution from microbial to animal rhodopsins. It is apparent that the presented TaHeR investigation indicates that such hypothesis could also apply to HeRs. More specifically, from these calculations it is possible to envision that a mutation at the Gln143 or Phe206 sites, causing a reduction in steric hindrance, could make the two CoIn channels reactive providing a first example of HeRs variant capable of producing the 11-*cis* chromophore isomer typical of animal rhodopsins.

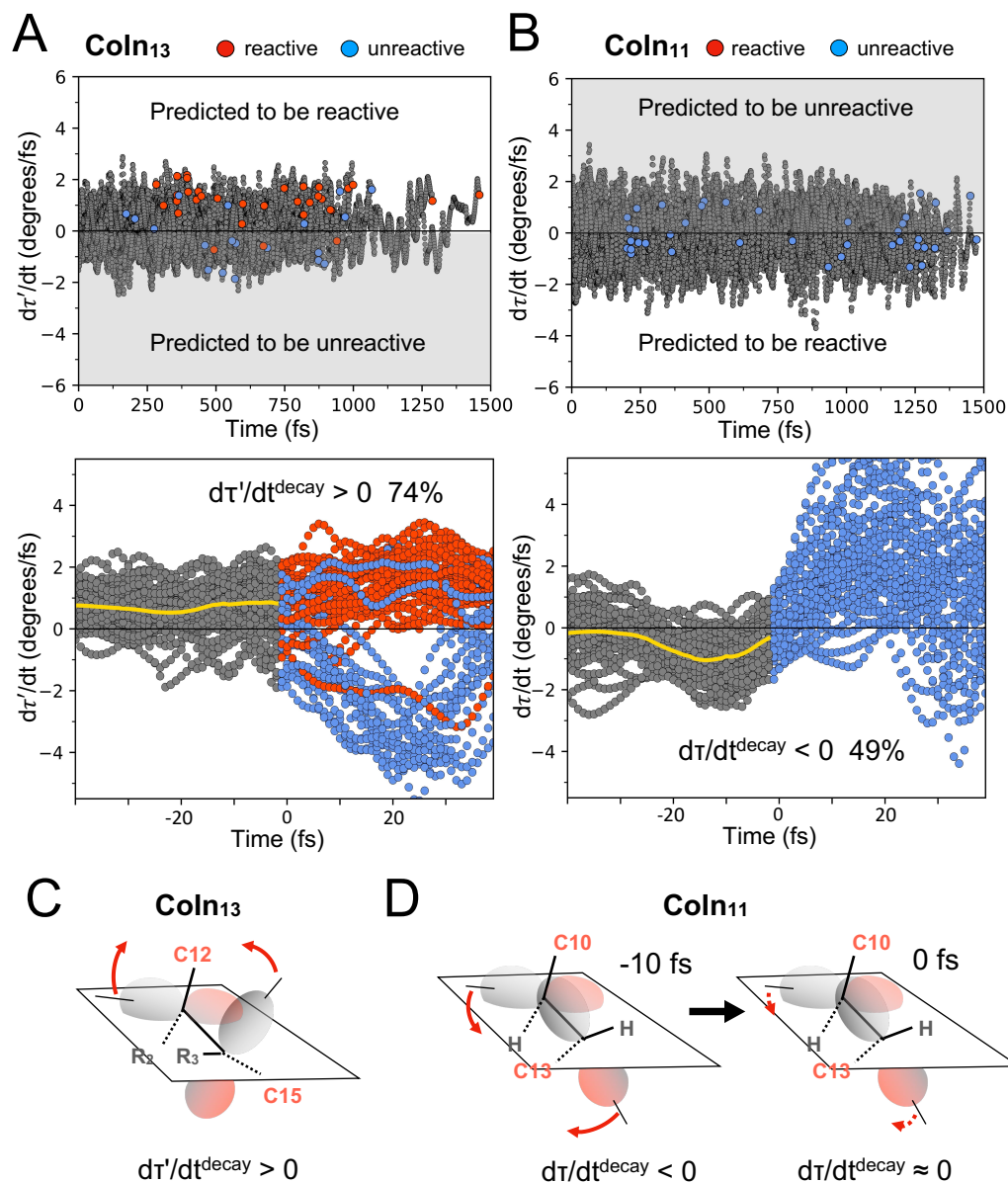


Figure 3.13: Reactivity analysis of TaHeR trajectory sets. A) Time evolution of the overlap index $d\tau'/dt$ towards the CoIn₁₃ channel. The top panel shows the correlation between the distribution of reactive decays (red circles) and the $d\tau'/dt$ sign. The bottom panel displays the same data but with trajectories aligned at the $S_1 \rightarrow S_0$ decay time ($t = 0$ fs) and starting 40 fs before the decay. Gray data refers to the S_1 while red and blue data refers to S_0 depending on whether the trajectory is reactive (red) or unreactive (blue). The yellow curve represents the average value before the $S_1 \rightarrow S_0$ decay. The displayed numerical value refers to the percentage of trajectories decaying with $d\tau'/dt > 0$. B) Same as A for the CCW motion leading to the CoIn₁₁ channel. Notice that now the displayed numerical value refers to $d\tau'/dt < 0$. C) Pictorial illustration of the τ' velocity in terms of decreasing overlap along the CW progression towards the CoIn₁₃ channel. The curly arrow represents the rotational velocity of the orbital axes. D) Same as C for the CoIn₁₁ channel where, in contrast, the velocity of the CCW overlap decrease is decelerating during the progression towards the decay channel.

3.2. PHOTOISOMERIZATION QUANTUM YIELD MODULATION IN THE HELIORHODOPSIN FAMILY

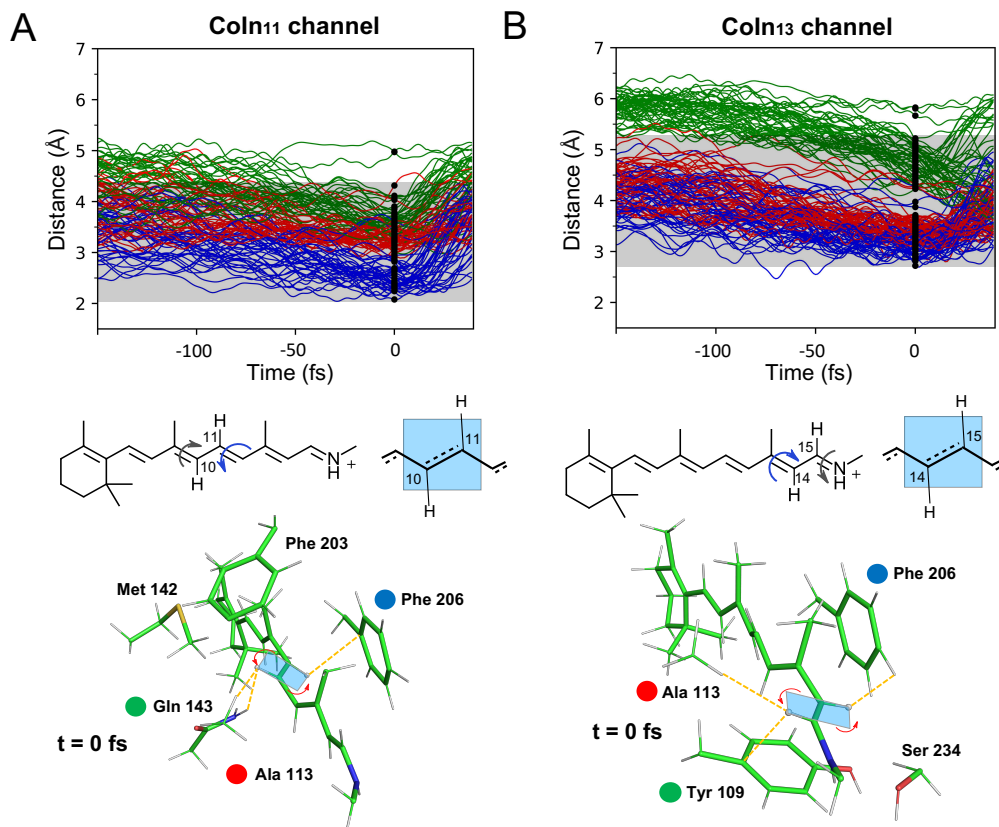


Figure 3.14: Quantum efficiency quenching mechanism in TaHeR. A) Population time evolution (trajectories aligned at the $S_1 \rightarrow S_0$ decay time, $t = 0$ fs) of the closest the distances between the rPSB chromophore atoms characterizing the CW rotation of the bicycle-pedal plane of the CoIn₁₁ channel with neighboring amino acids indicated below. The vertical set of black dots indicate the distribution of distances at the $S_1 \rightarrow S_0$ hopping points. Below, the definition and color coding of the distances monitored in the top. The rotating bicycle pedal plane is also defined pictorially and within a snapshot of the reacting rPSB region at the $S_1 \rightarrow S_0$ decay time ($t = 0$ fs). B) Same as A for the CoIn₁₃ channel.

Chapter 4

Conclusions and Outlook

The computational study outlined in this thesis focuses on the comprehension, at the geometrical and electronic level of detail, of the as-yet-undescribed mechanism by which two recently discovered and phylogenetically distant rhodopsins employ the retinal chromophore to achieve light sensitivity. This undertaking was facilitated by the application of a standardized protocol known as a-ARM, which enabled the generation of basic, yet robust, QM/MM models. These models, proven accurate enough to semi-quantitatively reproduce the experimentally observed spectral and photodynamic behavior, have served as computer models for simulating the photoisomerization of the retinal chromophore; the light induced process that initiates the protein photocycle and the corresponding biological function. Accordingly, the conclusions of this thesis can be divided into the following two parts:

Photoisomerization Quantum Yield Modulation in the Heliorhodopsin Family

Excited state population dynamics was employed to simulate the primary photoisomerization event occurring in the archaea TaHeR, a member of the newly discovered third HeR superfamily, unveiled only in 2018. The calculations successfully replicated the experimentally observed conversion of the all-*trans* to 13-*cis* retinal chromophore with a low ϕ_{iso} . Intriguingly, despite the extreme distance ($< 15\%$ sequence identity) with animal and microbial homologues, this conversion was found accomplished through essentially the same reaction coordinate described by the (aborted) bicycle pedal motion occurring within the -C13=C14-C15=N- chromophore moiety. To this date, this marked the first evidence of the conservation of this archetypal isomerization coordinate in the HeR family.

Unexpectedly, the evolution of the TaHeR photoexcited population exhibited low selectivity for the C13=C14 isomerization. A competing, equally populated bicycle pedal pathway affecting, analogously to animal rhodopsin, the -C9=C10-C11=C12- chromophore moiety was observed. This reduced reaction regioselectivity was deemed responsible for the low C13=C14 ϕ_{iso} , instead predicted to be approximately 0.60 if the C13=C14 isomerization were the sole operating pathway for excited state deactivation. Interestingly, the detected animal-like isomerization motion was found completely unproductive, since leading to internal conversion

exclusively, due to steric interaction with nearby aromatic residues that hindered the rotation of the reactive moiety upon decay to the ground state. Furthermore, the presence of such unproductive decay events was proposed to explain the elusive origin of the "unreactive" signals observed during time resolved absorption spectroscopy (TAS) measurements during the TaHeR excited state deactivation.

In general, the calculations unveiled a previously unreported and effective way by which HeRs control the ϕ_{iso} that, for the best of our knowledge, was never observed in other rhodopsins. Potentially, this mechanism call for TaHeR as a candidate for coherent control experiments. The suppression of one or the other of the two excited state pathways, for instance through targeted mutagenesis, is likely to result in different ϕ_{iso} values and in the simultaneous production C13=C14 and C11=C12 isomerization products.

Red-shift and Fluorescence Enhancement of Neorhodopsin

For NeoR, an in-depth exploration of its electronic structure, coupled with static photoisomerization reaction path computations and resonance Raman (RR) simulations, was employed to elucidate the origin of its unusual far-red spectral properties, particularly desirable within the field of optogenetics. The analysis revealed that the conventional ultrafast C13=C14 photoisomerization, which typically prevents canonical microbial rhodopsins from being fluorescent, can be impeded by the lack in chromophore counterions in proximity of the Schiff base moiety. Instead, a distal chromophore counterions could be eventually located in region located halfway along the chromophore polyene chain.

The primary local effect of this electrostatic constraint was the generation of a DA featuring an highly diffuse positive charge spread all along the π conjugated system. The rest of the described properties, including spectral properties such as the large λ_{max}^a bathochromic shift, small Stokes shift, and sizable S_1 isomerization energy barrier along the canonical C13=C14 isomerizing coordinate, emerged as a consequence of such a change. Specifically, it was recognized that the positive charge re-confinement occurring during the S_1 isomerization event was the primary cause for the increase in the isomerization energy barrier, and consequently, for the enhanced emission fluorescence brightness. This mechanism, termed the "charge delocalization and then confinement" mechanism was therefore indicated as the fluorescence enhancement control mechanism operating in NeoR.

A fascinating avenue for further exploration lies in the possibility of generating QM/MM models of ancestral NeoR sequences (i.e. extinct). Such an endeavor would provide insights into whether fluorescence (or the λ_{max}^a bathochromic shift) was already an ancestral property. If not, it could unravel how natural evolution has shaped the NeoR sequence all along its phylogeny to attain these distinctive spectral features displayed nowadays. Such type of investigation could not only deepen our understanding of NeoR's evolutionary history but also inspire novel strategies in the engineering of optogenetic tools with tailored fluorescence properties.

Chapter 5

Articles

5.1 PAPER 1: Retinal Chromophore Charge Delocalization and Confinement explain the Extreme Photophysics of Neorhodopsin

Publication: Riccardo Palombo, Leonardo Barneschi, Laura Pedraza-Gonzalez, Daniele Padula, Igor Schapiro, Massimo Olivucci, 2022, Nature Communications, 13(1), 6652.

Contribution: Research design, calculations, manuscript writing and revision, data analysis and graphics generation.

Status: Published on November the 4th, 2022.


Copyright: Open access distributed under the terms of the *Creative Commons CC BY*.

Retinal chromophore charge delocalization and confinement explain the extreme photophysics of Neorhodopsin

Received: 21 April 2022

Accepted: 7 October 2022

Published online: 04 November 2022

 Check for updatesRiccardo Palombo^{1,2}, Leonardo Barneschi¹, Laura Pedraza-González¹,
Daniele Padula¹, Igor Schapiro³ & Massimo Olivucci^{1,2}✉

The understanding of how the rhodopsin sequence can be modified to exactly modulate the spectroscopic properties of its retinal chromophore, is a prerequisite for the rational design of more effective optogenetic tools. One key problem is that of establishing the rules to be satisfied for achieving highly fluorescent rhodopsins with a near infrared absorption. In the present paper we use multi-configurational quantum chemistry to construct a computer model of a recently discovered natural rhodopsin, Neorhodopsin, displaying exactly such properties. We show that the model, that successfully replicates the relevant experimental observables, unveils a geometrical and electronic structure of the chromophore featuring a highly diffuse charge distribution along its conjugated chain. The same model reveals that a charge confinement process occurring along the chromophore excited state isomerization coordinate, is the primary cause of the observed fluorescence enhancement.

Modern neuroscience requires membrane-localized signaling tools^{1,2} that could emit intense fluorescence upon irradiation with red light. However, until recently, the available tools, based on engineered microbial rhodopsins, could only generate weak fluorescence signals that impair their performance. At the molecular level, the optical properties of microbial rhodopsins owe to the presence of a covalently bounded all-*trans* retinal protonated Schiff base (rPSB) chromophore and its interaction with the surrounding protein environment. Therefore, a deep molecular comprehension of the factors dictating such properties is highly desirable. In this regard, few studies^{3–5} have formulated rules for tailoring the absorption and emission properties of the retinal chromophore based on the effects of homogeneous electrostatic fields acting on isolated chromophores or via chromophore chemical modifications. However, it is expected that a simple electrostatic picture could not be sufficient to explain the origin of these properties in the complex environment offered by the protein cavity since other factors like non-homogeneous electrostatic fields or chromophore-cavity steric effects could play an important role.

In 2020 the discovery of Neorhodopsin (NeoR) offered an unprecedented case study that could potentially expand our comprehension of red-shifted and highly fluorescent rhodopsins. NeoR is a rhodopsin guanylyl-cyclase (RGC) expressed in the *Rhizoclostridium globosum* from Chytridiomycota, the only phylum of fungi producing motile and flagellated spores (zoospores)^{6,7}. It heterodimerizes with other two RGCs, called RGC1 and RGC2, that have sensitivity in the blue-green spectrum with 550 and 480 nm absorption maxima ($\lambda_{\text{max}}^{\text{a}}$), respectively. In contrast, NeoR displays the strongest bathochromic shift among all known microbial rhodopsins, yielding an extremely red-shifted ($\lambda_{\text{max}}^{\text{a}} = 690$ nm) absorption band. Such a band is mirrored by an intense emission band with a maximum ($\lambda_{\text{max}}^{\text{f}}$) at 707 nm yielding Stokes shift of only 17 nm (350 cm⁻¹). The emission brightness is quantified by a fluorescence quantum yield (FQY) of 20% and by an extinction coefficient (ϵ) of $129,000$ M⁻¹ cm⁻¹. In addition, the excited state lifetime (ESL) of 1.1 ns points to a slow excited state deactivation. The FQY of NeoR, only ca. four times weaker than that of the green fluorescent protein⁸ (GFP), represents an anomaly in the rhodopsin superfamily and suggests an evolution-driven origin. More specifically,

¹Dipartimento di Biotecnologie, Chimica e Farmacia, Università di Siena, via A. Moro 2, I-53100 Siena, Italy. ²Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403, USA. ³Fritz Haber Center for Molecular Dynamics, Institute of Chemistry, The Hebrew University of Jerusalem, 9190401 Jerusalem, Israel. ✉e-mail: olivucci@unisi.it

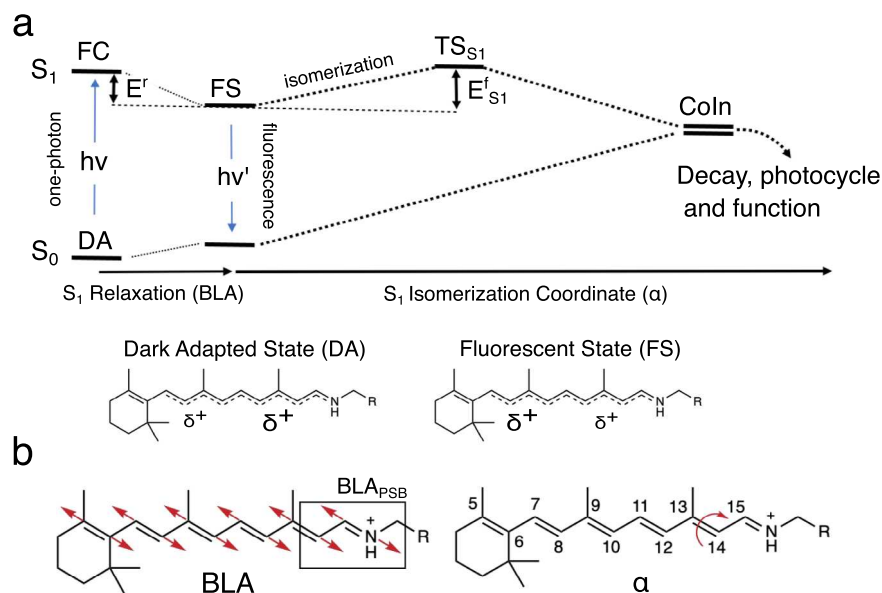


Fig. 1 | Geometrical and electronic structure changes in NeoR. a Schematic representation of the hypothetical S_0 and S_1 energy changes occurring along the S_1 relaxation that involves the bond length alternation (BLA, quantified by the difference between the average of the double-bond lengths and the average of the single-bond lengths of the conjugated chain) and isomerization (α) coordinates. The rPSB resonance hybrids show a delocalized positive charge at the S_0 and S_1

energy minima corresponding to the Dark Adapted State (DA) and Fluorescent State (FS), respectively. The symbol “ δ^+ ” gives a qualitative measure of the amount of positive charge located along the rPSB-conjugated chain. **b** Representation of the bond length alternation (BLA) mode and the torsion mode (α) along the C13=C14 double bond. BLA_{PSB} is the -C14-C15 and C15-N bond lengths difference.

since the emission competes with the photoisomerization of its rPSB chromophore, a presently unknown adaptation process must have decreased the efficiency of the protein function. This hypothesis is in line with the fact that wild-type (WT) rhodopsins commonly exhibit FQYs spanning the 0.0001%–0.01%^{9–12} range while engineering efforts yielded variants with only modest increases up to a 1.2% value^{13–16}.

Deciphering how natural evolution in NeoR has tuned these extreme spectroscopic properties of the rPSB chromophore could expand our ability to design optogenetic tools with augmented functionality. Therefore, the modeling of NeoR represents a new promising learning opportunity that can be also used to assess the transferability of the rules mentioned above. In particular, NeoR offers the opportunity to disclose the molecular-level mechanism controlling the branching between fluorescence emission and photoisomerization. Such branching, which is schematically illustrated in Fig. 1a for all-*trans* rPSB, has been shown to dominate the fluorescence modulation in a set of GFP-like protein variants^{8,17}. More specifically, in these systems, the FQY appears to be directly proportional to the energy barrier (E_{S1}^f) controlling both access to a conical intersection (CoIn) located along the first singlet excited state (S_1) isomerization coordinate and the decay to the ground state (S_0). Here we assume that the same mechanism operates in NeoR is then used as a “laboratory” model for proposing a mechanism capable to connect sequence variation and rPSB emission. To do so, we also assume, in line with the evidence coming from a set of Arch3 variants displaying enhanced fluorescence^{18,19}, that the NeoR emission is a one-photon process and that, therefore, originates directly from its dark adapted state (DA).

In order to pursue the objectives above, we construct a quantum-mechanics/molecular mechanics (QM/MM) model of NeoR based on multiconfigurational quantum chemistry. Since a crystallographic structure is not available, we employ, for the model construction, a previously reported comparative model⁷. While this may limit, in principle, the fidelity of the environment description with respect to that found in nature, our target here is to achieve an atomistic model capable to replicate all relevant spectroscopic and photochemical observables and use it to explain the high FQY of NeoR in terms of geometrical, electrostatic and steric effects.

Accordingly, here we firstly use the QM/MM model to investigate the electronic structures of the NeoR DA and fluorescent state (FS) and, secondly, we use it to investigate the NeoR photoisomerization with the target of documenting the magnitude and origin of E_{S1}^f . We show that the confinement of the delocalized positive charge on the Schiff base moiety of the rPSB backbone can explain the existence of large E_{S1}^f values and, in turn, the high FQY of NeoR.

Results and discussion

Model construction and validation

An initial QM/MM model of NeoR was constructed using the Automatic Rhodopsin Modeling (*a*-ARM) technology^{20–22} starting from the comparative model mentioned above. *a*-ARM models have been shown to yield congruous (i.e., built by employing exactly the same protocol) animal and microbial rhodopsin models that correctly reproduce trends in λ_{\max}^a values^{9,20,21,23–27}. The model showed that the NeoR all-*trans* rPSB is embedded in a cavity featuring a peculiar amino acid composition with two glutamic (E136 and E262) and one aspartic (D140) acid residues located in the vicinity of its Schiff base moiety. However, due to the lack of experimental information on the residue protonation state, the chromophore counterion assignment remains ambiguous²⁸. For this reason, a set of customized *a*-ARM models featuring different protonation states for the E136, E262, D140 plus E141, a residue located halfway along the rPSB conjugated chain (see Fig. 2a), were built and ranked by computing the absorption (λ_{\max}^a) and emission (λ_{\max}^f) maxima as well as the relaxation energy (E^f) defined by the basic mechanism of Fig. 1a. The λ_{\max}^a and λ_{\max}^f values were computed in terms of vertical excitation energies (ΔE_{S0-S1}) between S_0 and S_1 at the DA and FS equilibrium geometries, respectively. E^f was instead computed as the energy difference between the Franck–Condon (FC) point and FS state and, therefore, quantifies the energy decrease associated with S_1 relaxation. The results collected in Fig. 2b that display the λ_{\max}^a , λ_{\max}^f , and E^f values for models where the “counterion tetrad” defined above have total charges of 0, –1, –2, –3. For completeness, we have also reported the scenario with a total charge –3 even if the two transitions displayed by these models are not allowed, being the oscillator strengths (f_{S0-S1}) close to zero.

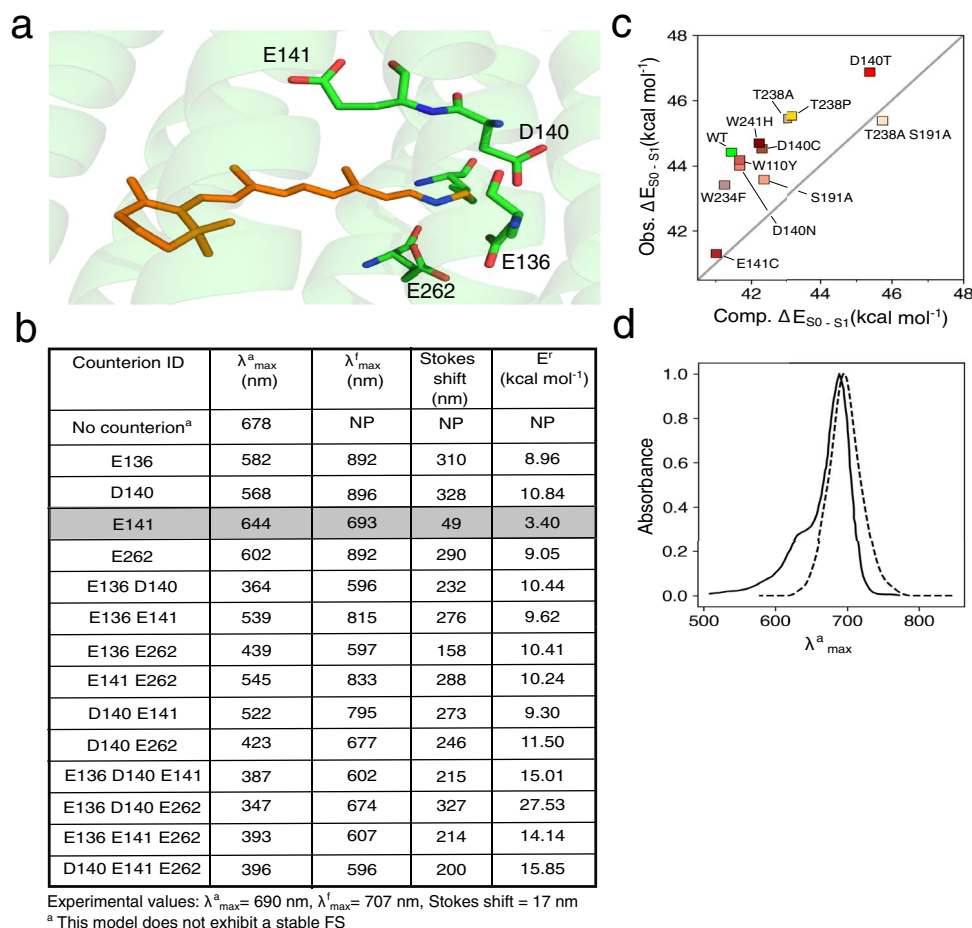


Fig. 2 | Choice of the NeOR chromophore counterion and model assessment.

a Overview of the structure of the all-trans rPSB chromophore (orange) and its four potential residue counterions (in green). The lysine residue (in green) bounded to the rPSB chromophore is also displayed. **b** Computed (CASPT2 level) maximum absorption wavelength (λ_{\max}^a), maximum emission wavelength (λ_{\max}^f) and relaxation energy (E^r) of NeoR with varying counterion choices. **c** Correlation between

experimental (Obs. $\Delta E_{S_0-S_1}$) and computed (Comp. $\Delta E_{S_0-S_1}$) values of vertical excitation energies defining λ_{\max}^a in the wild type (indicated as WT) and a set of NeoR mutants. **d** Superimposition of experimental and computed (dotted line) absorption band of wild type NeoR. The experimental band has been digitalized from the corresponding ref. 7.

Figure 2b shows that the model with a deprotonated E141 and neutral E136, D140, and E262 (from now on α -ARM_{E141}) is the most accurate. In fact, α -ARM_{E141} yields λ_{\max}^a and λ_{\max}^f values only 46 and 14 nm blue-shifted with respect to the experimental value as well as the smallest difference (49 nm) between those values consistent with the tiny Stokes shift experimentally observed for NeoR. All other assessed protonation states yielded a poor comparison with the observed quantities. For instance, although the model with all four residues protonated (i.e., with no counterion) produces a λ_{\max}^a close to the experimental one, it lacks a stable FS structure since no energy barrier could be located preventing access to the CoIn along α . Also, consistently with the high intensity of the observed absorption and emission bands⁷, the α -ARM_{E141} computed oscillator strengths are found to be very high (see Supplementary Tables 1 and 2): 1.71 and 1.90, respectively. Such values were confirmed via multistate XMCQDPT2 calculations that yielded values close to 1.66 and 1.80.

In order to further assess the quality of α -ARM_{E141}, we constructed the models of a set of NeoR variants whose λ_{\max}^a values have been experimentally measured (see Supplementary Table 3). As shown in Fig. 2c the models reproduce the observed trend indicating that α -ARM_{GLU141} describes, qualitatively, the effect of cavity residue replacements. Notice that the trend is reproduced with a systematic blue shift, which is typical of α -ARM models^{20,21,23,25–27,29}. We also used

α -ARM_{E141} to simulate the WT NeoR absorption band at room temperature by computing the $\Delta E_{S_0-S_1}$ and $S_0 \rightarrow S_1$ transition probability values for 200 snapshots representing the Boltzmann distribution (see Supplementary Section 3). Comparison between the simulated and observed data in Fig. 2d shows that the center of the computed band (703 nm) is only 13 nm red-shifted with respect to the experimental λ_{\max}^a value and the computed band half-width is close to that seen experimentally. α -ARM_{GLU141} only appears to miss a shoulder at 640 nm that is, likely, of vibronic origin⁷ and therefore not captured by a simulation based on the Condon approximation.

Electronic character of the DA and FS vertical transitions

The agreement between experimental and computed data allows to use α -ARM_{E141} to investigate the large bathochromic shift, negligible Stokes shift, and intense fluorescence of NeoR. The aim of Fig. 3b, c is to document the variation in electronic character upon light absorption and emission by looking at the Mayer bond order analysis (see Supplementary Table 6) and vertical electron density changes at DA and FS ($\delta\rho_{abs}$ and $\delta\rho_{emi}$, respectively). The results support the hypothesis that both the S_1 and S_0 electronic characters are combinations of putative covalent (COV) and charge transfer (CT) diabatic states loosely associated with the limiting resonance structures of Fig. 3a. In particular, the results reveal that, the DA structure displays an unusually large CT weight in S_0 yielding a positive charge spread

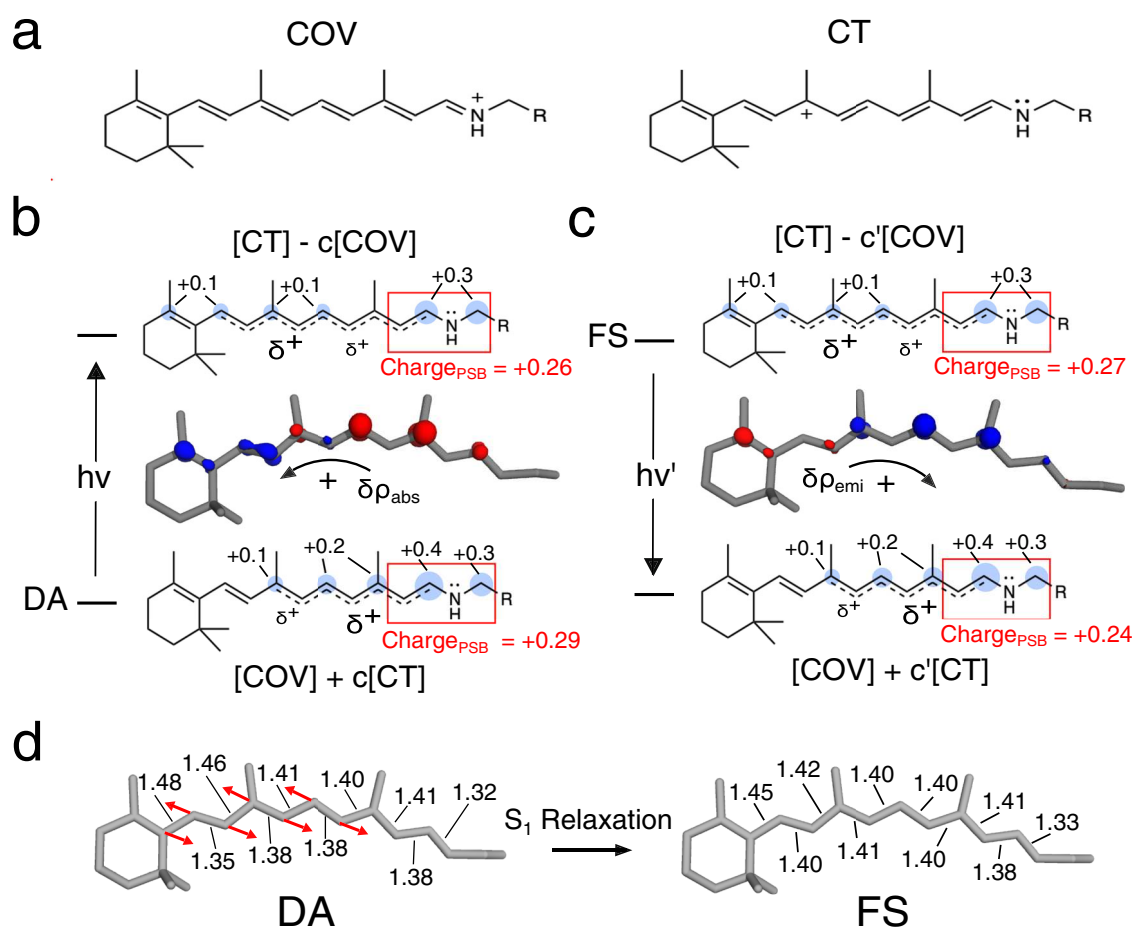


Fig. 3 | Electronic and geometrical character of S_1 relaxation in Neor chromophore. **a** Representation of the two limiting resonance formulas adopted to describe the electronic character of the rPSB chromophore. **b** Electron density variation ($\delta\rho_{abs}$) characterizing the vertical $S_0 \rightarrow S_1$ transition from the Dark Adapted State (DA). Blue and red clouds correspond to electron density decrease and increase respectively. Isovalue set to 0.002 a.u. The associated resonance formulas correspond to resonance hybrids also anticipated in Fig. 1a. Blue bubbles represent the QM positive charge (in e unites). Only absolute values > 0.05 e are reported. As

indicated by the red box, the total charge residing in the -C14-C15-N-C ϵ - rPSB fragment is also given. **c** Same data for the $S_1 \rightarrow S_0$ emission from the Fluorescent State (FS). **d** Geometrical comparison between DA and FS rPSB structures. The arrows indicate the dominant geometrical change corresponding, clearly, to a variation in the bond length alternation (BLA, see definition in the caption of Fig. 1) in a region of the conjugated chain distant from the Schiff base moiety. The relevant bond lengths are given in Å.

along the carbon atoms of the chromophore (see the bubble representation from Mulliken charges) with a limited +0.29 e charge residing in the C14-C15-N-C ϵ moiety (from now on, the charge residing on such moiety will be called Charge_{PSB}). The vertical transition to S_1 only slightly modifies such charge distribution. For instance, when taking the mid C13=C14 as a reference one can see only a small 0.03 e translocation towards the β -ionone ring. The same behavior is seen at the FS when looking at the vertical emissive transition for which one finds Charge_{PSB} values of +0.27 e and +0.24 e for S_1 and S_0 , respectively. In conclusion, as illustrated in Fig. 3b, c, the DA and FS transitions can be both qualitatively interpreted as transitions between adiabatic states (i.e., S_0 and S_1) corresponding to in-phase and out-of-phase mixing of two diabatals (or resonance formulas) close in energy. Such an interpretation appears to be related to the one proposed for explaining the observed absorption and emission trends of GFP-like proteins¹⁷.

The description above is not in line with the consensus electronic structure of the rPSB chromophore^{30,31} of rhodopsins. In fact, the DA $S_0 \rightarrow S_1$ transition, is usually described as a transition starting from a COV-dominated state featuring a positive charge localized on the -C15=N- moiety and not a delocalized charge spread on the -C9=C10-C11=C12-C13=C14-C15=N- chain as seen in Fig. 3b. Starting from such a

state, an at least three times larger charge translocation has been computed upon $S_0 \rightarrow S_1$ excitation^{32,33}.

Notice that the chromophore charge delocalization seen in the selected α -ARM_{E141} model is modulated by the position of the counterion. In fact, the charge distribution of the model featuring E262 as the only charged residue of the tetrad (see Fig. 2b) features, in the DA state, a blue-shifted λ^{max} and a reduced charge delocalization. These values are accompanied by a much larger +0.68 e to +0.27 e change in Charge_{PSB} value upon vertical excitation (see Supplementary Fig. 2) and are, therefore, more in line with the consensus rPSB charge distribution mentioned above.

FC \rightarrow FS geometrical and electronic relaxation

Consistently with the computed negligible (0.01 e) difference in Charge_{PSB} value between the FC point and the FS state, the S_1 electronic relaxation of α -ARM_{E141} can be interpreted as a relatively minor change in the weights of the COV and CT diabatals. The geometrical variation accompanying such a process is documented in Fig. 3d and corresponds to a minor progression along the BLA coordinate (this is defined as the difference between the average single-bond length and the average double-bond length of a conjugated chain, see Fig. 1b) of the chromophore, leading to an $E^{\ddagger} \approx 3.5$ kcal mol⁻¹

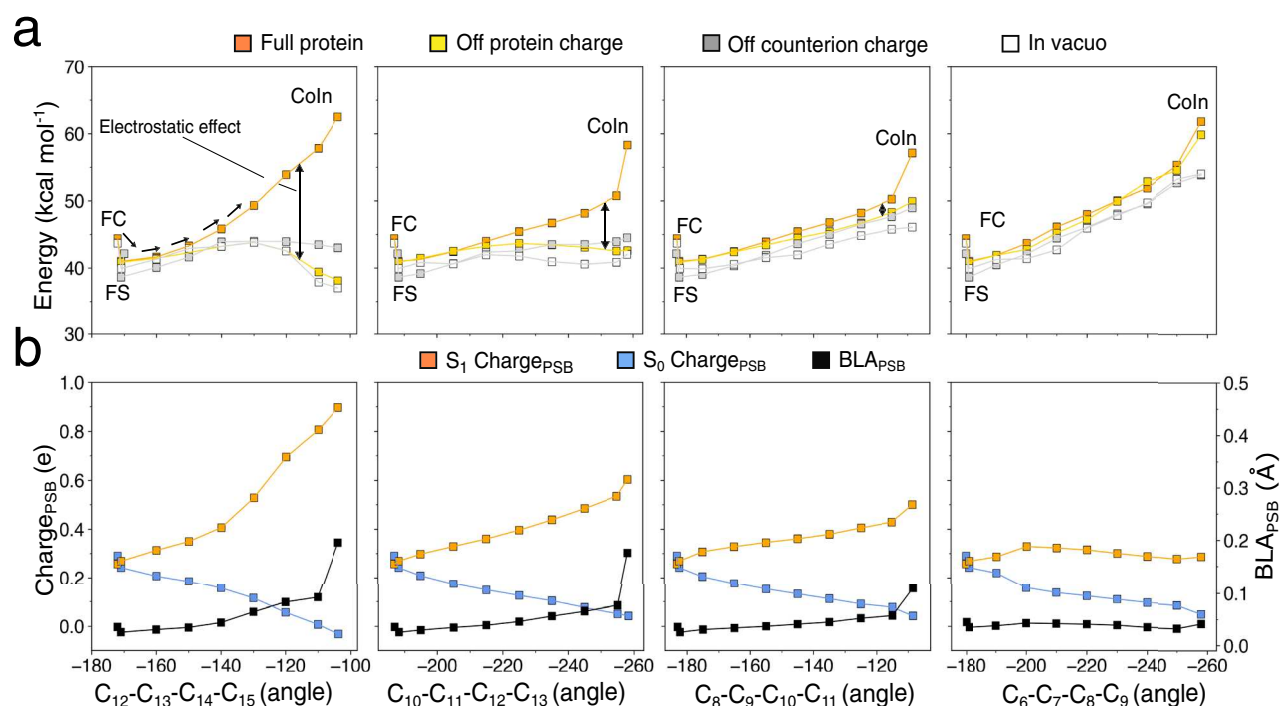


Fig. 4 | S_1 energy, charge and BLA profiles along the photoisomerization of, from left to right, C13=C14, C11=C12, C9=C10 and C7=C8 double bonds. **a** CASPT2 S_1 energy profiles computed in presence of the full protein environment (orange squares), after setting to zero the MM charges of the entire protein (gold squares), after setting to zero only the MM charges of the rPSB counterion (i.e.,

E141, gray squares) and after removing the whole protein in a full QM calculation (i.e., in vacuo, empty squares). **b** Evolution of S_0 and S_1 $\text{Charge}_{\text{PSB}}$ (the charge residing in the C14-C15-N-C ϵ rPSB moiety) and evolution of BLA_{PSB} (C14-C15 and C15-N bond lengths difference, see Fig. 1b).

value (see Fig. 2b). Such assignment is based on the hypothesis of a negligible contribution from the surrounding protein environment, in line with the results reported for GFPs¹⁷. Interestingly, the BLA-driven nature of the S_1 relaxation is a known feature among GFPs where the only geometrical difference between the DA and FS states is mainly attributed to a different BLA displacement¹⁷. In NeoR the BLA change is mostly described by a contraction and elongation of C6-C7/C8-C9/C10-C11 and the C7=C8/C9=C10/C11=C12 bonds respectively, while the BLA of the C14-C15-N Schiff base moiety (from now on simply BLA_{PSB} , see Fig. 1b) is minimally involved and only changes of 0.01 Å.

NeoR photoisomerization

As anticipated above, here we assume that the branching (see Fig. 1a) between the canonical rPSB double bond photoisomerization of microbial rhodopsin and emission determines the fluorescence brightness. In other words, since the S_1 isomerization is intrinsically a non-radiative decay process, the $E_{S_1}^{\ddagger}$ barrier controlling its rate must modulate the ESL and, in turn, the FQY value. Accordingly, we have used $\alpha\text{-ARM}_{\text{E141}}$ to compute the approximate S_1 minimum energy path (MEP) describing the clockwise (CW) torsional deformation along α , namely the C12-C13-C14-C15 dihedral angle describing the C13=C14 double bond torsion connecting the FS to the ca. 90° twisted conical intersection (CoIn) giving access to S_0 . The energy maximum located along the MEP energy profile must therefore reflect the barrier height (i.e., $E_{S_1}^{\ddagger}$).

Given the current uncertainty about the regiochemistry of the photoisomerization reaction in NeoR^{34,35}, we also computed three additional MEPs (see Supplementary Section 5) corresponding to the counterclockwise (CCW) isomerization of the C11=C12 and C7=C8 double bond and to the CW isomerization of the C9=C10 double bond. The choice of the CW/CCW pattern for adjacent double bonds conforms to the well-known aborted bicycle-pedal motion^{30,36}, the archetypal space-saving reaction coordinate for the rPSB chromophore

isomerization. The selected direction of the twisting is consistent with the stereochemistry found in the four computed CoIn's where the constrained rotation of the reactive bond is assisted by the opposite rotation of the adjacent double bonds (see Supplementary Fig. 3).

Figure 4 displays the CASPT2 S_1 energy profiles (top panels) together with the evolution of the S_0 and S_1 $\text{Charge}_{\text{PSB}}$ and the BLA_{PSB} coordinate (bottom panels). As reported in the previous works^{30,32,37}, and also above, such charges are used as indicators of the weights of the COV and CT diabatic states in the adiabatic S_0 and S_1 energy profiles. As detailed in section S5, the energy trends that emerged from the CASPT2 level were also confirmed at the XMCQDPT2 level. All four S_1 isomerization paths point to the presence of a barrier (see orange energy profiles). However, the S_1 energy profile is qualitatively different from the one hypothesized in Fig. 1a. In fact, in all cases the energy increases monotonically from the FS state with the CoIn corresponding to the highest point along the MEP. The plot is consistent with a sloped, rather than peaked, topography of the ca. 90° twisted CoIn structures³⁸ (see also the branching plane map in Supplementary Fig. 13 for the CoIn located along the C13=C14 MEP). As shown in Supplementary Fig. 5, we estimated the $E_{S_1}^{\ddagger}$ magnitude as the difference in energy between the CoIn and the FS states. $E_{S_1}^{\ddagger}$ is found of 21 (25), 17 (22), 16 (16), 20 (16) kcal mol⁻¹ at the CASPT2 (XMCQDPT2) level for respectively the C13=C14, C11=C12, C9=C10 and C7=C8 isomerizations. Since the lowest energy barrier is still relatively large, our model supports a barrier-controlled mechanism for the FS emission. Thus, NeoR would be an analog of the GFP-like fluorescent reporters^{8,30} as, in contrast with microbial rhodopsins such as bacteriorhodopsin³⁰, its relatively high barrier would induce slow internal conversion kinetics. This conclusion is consistent with the nanosecond ESL of NeoR while, most known fluorescent rhodopsins feature an ESL that does not exceed^{13,30,39-41} the picosecond range.

Notice that the model predicts lower energy barriers with respect to the canonical microbial C13=C14 isomerization which appears

highly disfavored. Hence, alternative photoisomerization pathways might originate as recently observed by Sugiura et al.³⁵ in the NeoR from *Obelidium mucronatum* (OmNeoR) that shares 78% of sequence identity with the NeoR studied in this work. This result appears consistent to what was found in the past by Cembran et al.³ aimed at investigating the relationship between the position of an acetate counterion and the photoisomerization of a nearly isolated protonated polyene chain. They found that placing the counterion above the polyene favors the isomerization of the double bonds closest to the counterion; this result can be loosely associated with the favored C11=C12, C9=C10 and C7=C8 isomerization in our a-ARM_{E141} model. However, we need to stress that the obtained $E_{S_1}^f$ values possibly represent upper limits as a sloped Coln features, in its close vicinity, a slightly lower energy region with the same 90° twisted conformation (see Supplementary Fig. 13).

In order to check the existence of alternative and lower energy photoisomerization channels, we computed the C8-C9, C10-C11, and C12-C13 single bond MEPs and found that in all cases $E_{S_1}^f$ is >20 kcal/mol⁻¹ (see Supplementary Figs. 9 and 10). This supports the high stability of the FS displayed by our a-ARM_{E141} model.

The evolution of the Charge_{PSB} along the C13=C14 isomerization coordinate (see the orange curve in the corresponding panel of Fig. 4) reveals that the S_1 weight of the COV diabatic state increases monotonically along the S_1 MEP until it dominates the region approaching the Coln. This corresponds to confinement (or localization) of the charge in the small Schiff base moiety that, along the terminal part of the MEP (i.e., near the 90° twisted conformation), hosts a π -system orthogonal to the one residing along the rest of the rPSB conjugated chain. This is not a general behavior that depends on the isomerizing double bond. In fact, the C7=C8 MEP in Fig. 4 appears to feature, along the entire S_1 profile, a steady mixed COV/CT character. These results point to a change in the origin of the critical $E_{S_1}^f$ barrier along different isomerization coordinates. More specifically, it is expected that the electrostatic effect imposed by the NeoR cavity may have different effects along different MEPs with a maximal effect on the canonical C13=C14 energy profile and a minimal effect on the C7=C8 energy profile. It is thus necessary to also evaluate steric effects.

In our ARM_{E141} model the electrostatic effect is due to the protein point charges including those describing the negatively charged E141 counterion. In order to assess the impact of such an effect on the isomerization energy profiles, these have been re-evaluated after setting to zero all protein point charges while keeping the geometrical progression unchanged (see gold energy profiles). Consistently, with a dominant role played by the protein electrostatics, the slope in the S_1 profile associated with the C13=C14 coordinate is strongly decreased and even inverts from positive to negative in its last part. This effect is gradually reduced along the C11=C12 and C9=C10 coordinates (compare the vertical double arrows) and disappears along the C7=C8 coordinate of Fig. 4d. The energy profiles were also re-computed after the removal of the E141 counterion charge exclusively (see gray energy profiles). When compared to the energy progression seen in absence of the protein electrostatics, the effect is reduced but maintained, indicating that the leading electrostatic contribution is due to the negative charge in the E141 position. This behavior is consistent with the lack of a stable FS (i.e., due to the absence of an S_1 energy barrier controlling access to the Coln) displayed by the model with no counterions (see Fig. 2b). In fact, switching off the charge of the E141 counterion roughly replicates the electrostatic embedding imposed by that model.

The models featuring a counterion configuration different from that of a-ARM_{E141}, display flat, and substantially barrierless, S_1 isomerization energy profiles (see Supplementary Fig. 11). As stated above, this is not consistent with the ESL of NeoR estimated to be 1.1 ns. Therefore, our data indicate that an E141 counterion appears not only critical for tuning the extreme spectroscopy of NeoR (see Fig. 2b) but also for the generation of a barrier.

To disentangle the electrostatic and steric contributions to the computed $E_{S_1}^f$ value, the same energy profiles have been re-evaluated in the absence of a whole protein environment (see the energy profiles marked with empty squares). The results demonstrate that while in C11=C12 and C9=C10 MEPs the S_1 profile becomes completely flat, in C7=C8 MEP the S_1 energy barrier is only reduced but persists, indicating a destabilization that originates from the rPSB geometrical progression. Notice that such progression is due, in all cases, to indirect electrostatic and steric effects determining the DA, FS, and Coln geometries (i.e., determining the isomerization coordinate) and include the effect of the polarization of the rPSB π -electron density due to the counterion.

Our conclusion is that a small barrier increasing along the C13=C14 to C7=C8 series, is an intrinsic feature of the isomerization coordinate computed using a-ARM_{E141}. While such an increase is clearly enhanced when switching on the *direct* steric interactions (i.e., due to the Lennard-Jones potentials between QM and MM atoms) are considered, the $E_{S_1}^f$ value along the C13=C14 and C11=C12 paths remain flat, and inconsistent with bright emission. To enhance these barriers and restore consistency, a direct electrostatic contribution (i.e., due to the interaction between QM electron density and MM point charges) is critical. In the next section, we look at the mechanism driving such a critical electrostatic effect.

Fluorescence enhancement mechanism

In microbial rhodopsins the canonical S_1 isomerization produces the 13-*cis* rPSB chromophore. In general, this is an ultrafast (sub-picosecond) reaction only allowing a negligible fluorescence emission from the DA state. We now use the results above to formulate a mechanistic theory for the fluorescence enhancement explaining how the NeoR electrostatics generates the high C13=C14 isomerization barrier of Fig. 4. Such theory takes the progressive confinement of the initially delocalized rPSB charge described above as the key event blocking the C13=C14 isomerization.

We start by employing the Charge_{PSB} and BLA_{PSB} quantities defined above to follow the chromophore geometrical and electronic changes along the S_1 isomerization coordinate. The first index displays a monotonic charge increase from 0.27 to 0.90 e (at FS and Coln respectively), consistently with a monotonic increase of the positive charge on the Schiff based chromophore moiety. The second index points to a 0.03 to 0.20 Å change consistently with the reconstitution of a C=N double bond along the path and full localization of the charge on such a bond. Such progressive charge confinement is directly proportional to the increase in the electrostatic effect along the MEP of C13=C14 of Fig. 4 (i.e., the one indicated by the double-headed vertical arrow) and, therefore, to the energy increase leading to the large computed $E_{S_1}^f$ value. We now propose that the molecular mechanism driving the energy increase is the progressive increase in distance between the negative E141 counterion charge and the centroid of the confining charge. As illustrated in Fig. 5a the progressive positive charge confinement shifts the centroid of the positive charge away from the E141 residue, unavoidably leading to destabilization. This mechanism is supported by the computed decrease in electrostatic effect (i.e., again, the destabilization indicated by the vertical arrow) along the C13=C14, C11=C12 and C7=C8 MEPs of Fig. 4. As an example, in Fig. 5a we also show that the C9=C10 isomerization could not lead to the same electrostatic effect as, in this case, the charge does not get confined far from E141 but remains delocalized along the extended C10-C11-C12-C13-C14-C15-N moiety. This causes only a limited change in the counterion-chromophore interaction consistently with the computed decrease in electrostatic stabilization. The charge confinement on the Schiff base moiety is thus critical.

The mechanism described above can be reinterpreted in terms of changes in the energy of the COV diabatic state (see H_{COV} in Fig. 5b) featuring a positive charge permanently located on the Schiff

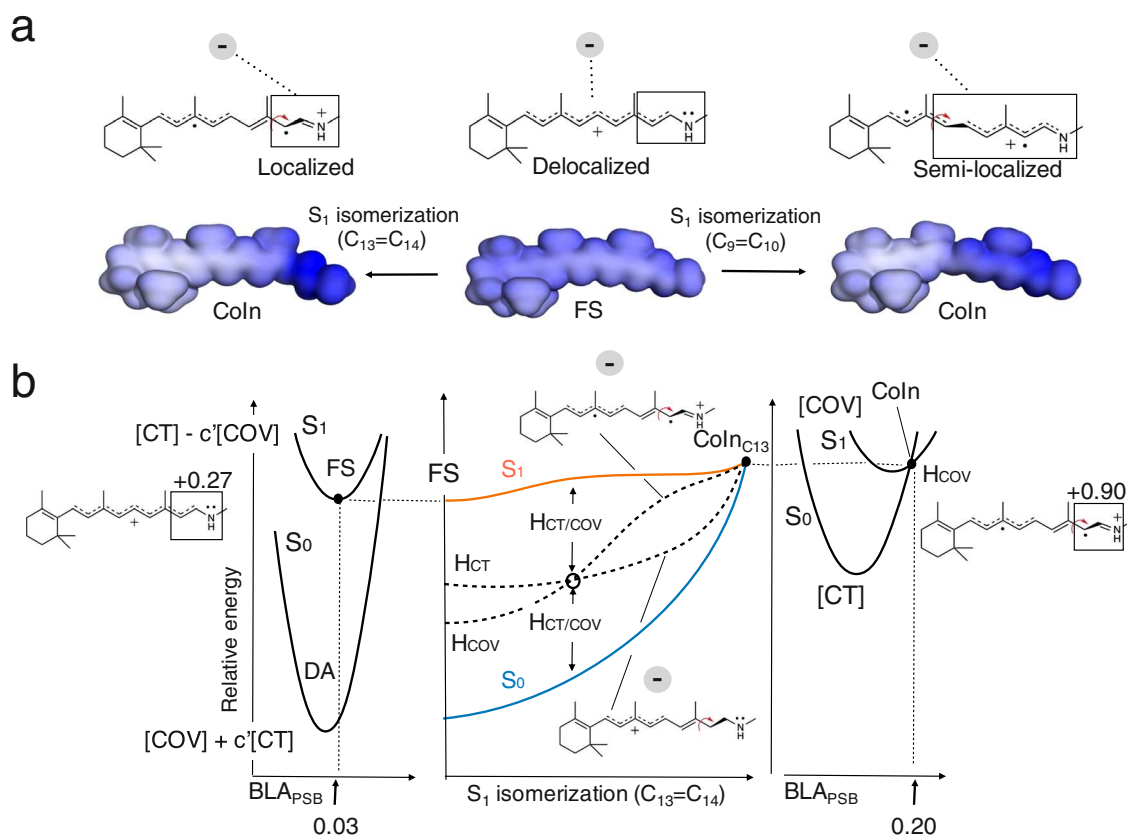


Fig. 5 | Origin of the E_{S_1} barrier in NeOR. **a** Representation of the increase in the distance between the negative and positive charge centroids due to the positive charge confinement along the C13=C14 and C9=C10 photoisomerization paths. Comparison between electrostatic potential (ESP) maps indicates that along the C9=C10 coordinate the extent of the confinement is less pronounced being the charge at the Coln spread on a longer rPSB chromophore moiety (i.e., C10-C11-C12-C13-C14-C15-N). **b** Proposed origin of the isomerization barrier in terms of COV energy (H_{COV} , dashed line) destabilization due to the charge confinement resulting from the mixed $[CT] - c[COV]$ to the pure $[COV]$ electronic structure change along

the S_1 adiabatic energy profile (in orange). We hypothesize that the diabatic energy curves cross halfway along the isomerization coordinate α , which therefore corresponds to the point with the highest diabatic coupling. The such diabatic coupling will then vanish at the Coln. The left and right panels display the shape of the S_1 and S_0 adiabatic potential energy curves along the BLA_{PSB} coordinate (see definition in the caption of Fig. 4) at FS (left) and Coln (right) and are in line with the presented FC→FS and Coln computations. The BLA_{PSB} and α coordinates are substantially orthogonal.

base moiety (see the resonance formula in Fig. 3a). The highly delocalized electronic structure of the FS points to close H_{COV} and H_{CT} values consistent with the small $Charge_{PSB}$ value of S_1 . On the other hand, the large $S_1 Charge_{PSB}$ computed in the Coln region points to an adiabatic state dominated by the COV diabatic. Thus, the charge confinement effect described above is translated into an increase in the weight of the COV diabatic state along the isomerization coordinate. This justifies the steep S_1 energy increase in the Coln region that would originate from the simultaneous increase in the COV weight and COV destabilization due to the offset in the electrostatic interaction with E141.

Such a delocalization-confinement mechanism (i.e., without an initially delocalized charge there is no progressive confinement) also suggests a lesser sensitivity of the S_0 state to the progression along the isomerization coordinate. In fact, while S_0 becomes progressively dominated by the CT diabatic state, in CT the rPSB charge remains relatively unconfined along a long segment (-C7-C8-C9-C10-C11-C12-C13-) of the rPSB backbone even in the Coln region.

Above we have shown that, the conventional C13=C14 isomerization preventing microbial rhodopsin to be highly fluorescent, can be blocked or slowed down by a suitable change in the electrostatic environment of the all-*trans* rPSB chromophore. More specifically, the presented α -ARM_{E141} model, indicates that NeOR is the product of an evolutionary process driven by the translocation of a negatively charged residue from the chromophore Schiff base region

to a region located halfway along the chromophore conjugated chain (E141). The main local effect of this process is the generation of a DA state featuring a delocalized rPSB-positive charge. The rest of the described properties, including spectral properties such as the large bathochromic shift, small Stokes shift, and sizable S_1 isomerization energy barrier along the canonical C13=C14 torsional coordinate, are a consequence of such a change. This behavior was recently documented by El-Tahawy et al. in isolated rPSB chromophores subject to homogeneous, strongly negative red-shifting electric fields and, therefore, the “two electron-two orbital model” theory that is shown in there to account for the electrostatic origin of the C13=C14 photoisomerization energy barrier appears also operative in the here presented NeOR². Notice that the described E141 counterion location is presently a theoretical result that remains to be experimentally demonstrated.

The effect associated with the repositioning of the rPSB counterion was previously proposed to explain the λ_{max}^a changes observed in a set of rhodopsin mimics based on the human cellular retinol-binding protein II (hCRBP_{II})⁴². Similar to our NeOR model, the members of the set displaying a large red shift were found to be associated with a counterion located far from the Schiff base moiety. This conclusion was reached through X-ray crystallographic analysis supporting the hypothesis that a red-shifted λ_{max}^a must be associated with an even distribution of the iminium charge along the chromophore π -conjugated chain. Such delocalization can be associated with the rPSB

delocalized charge seen in the α -ARM_{E141} calculation and interpreted as a COV \leftrightarrow CT resonance hybrid (see Fig. 3a) or, in a different language, to a near cyanine limit situation⁴³.

α -ARM_{E141} leads to a possible general principle for the engineering of other highly fluorescent rhodopsins that we call “delocalization-confinement”. Such principle establishes that the electrostatic field generated by the cavity, for instance via a specific counterion localization, must yield a vastly delocalized geometrical and electronic structure of the rPSB conjugated chain in both the DA and the FS state of the protein. In this condition, an electrostatically induced high reaction barrier can be generated via a rPSB charge confinement process occurring, unavoidably, along the canonical C13=C14 isomerization path in the region entering the corresponding Coln channel.

Finally, the reported results provide evidence that the spectroscopy of retinal proteins is regulated by the same principles regulating GFP-like fluorescence. More specifically, it was proposed that the GFP variants achieving maximal π -electron delocalization (called the cyanine limit) are the ones where a COV and CT configurations of the protein chromophore have exactly the same weight, thus pushing the λ_{max} value to the extreme red and culminating in a null Stokes shift.

Methods

The employed hybrid QM/MM modeling of NeoR was performed using the α -ARM protocol^{20–22} and based on the comparative model structure built and validated by S. Adam et al.⁷ Further details about the α -ARM protocol are given as Supplementary Informations (Supplementary Section 1). After initially producing, automatically, default α -ARM models for WT and mutant NeoR, the equilibrium geometries of the DA were obtained via re-assignment of the counterion before carrying out ground state geometrical relaxation with energy gradients calculated at the 2 root state average CASSCF(12,12)/6-31 G*/AMBER94 level of theory^{44–46} using the Molcas/Tinker^{47,48} interface (Supplementary section 2). The relevant energies were instead computed, again employing the Molcas/Tinker interface, at the single-state and, in specified cases, multistate multiconfigurational levels. These correspond to the 3-root state average CASPT2(12,12)/6-31 G*/AMBER94 and 3-root state average XMCQDPT2/CASSCF(12,12)/6-31 G*/AMBER94 levels, respectively. The XMCQDPT2 calculation was based on Firefly v8.2⁴⁹. The collection of geometries connecting the FS to the different Coln's and defining the MEPs discussed above, were obtained via constrained geometry optimization at the 2 root state average CASSCF(12,12)/6-31 G*/AMBER94 level of theory. As shown in Supplementary Section 5, for each MEP the S_1 isomerization barrier $E_{S_1}^{\ddagger}$ is estimated after reevaluating the S_1 energy profiles at the above CASPT2 and XMCQDPT2 levels via single point energy calculations and measuring the energy difference between the highest S_1 energy value (i.e., corresponds to the Coln) and the S_1 energy value of the FS.

Data availability

The cartesian coordinates of the DA and the FS of the QM/MM model of NeoR generated in this study are provided, respectively, in Supplementary Data 1 and Supplementary Data 2. Source data are provided with this paper.

Code availability

As stated in the Method section, all the calculations carried out in this work were performed using a combination of the quantum chemical program MOLCAS and molecular mechanics program TINKER except for XMCQDPT2 calculations which were performed using the computational chemistry program Firefly.

References

- Xu, Y., Zou, P. & Cohen, A. E. Voltage imaging with genetically encoded indicators. *Curr Opin Chem Biol.* **39**, 1–10 (2017).
- Lehtinen, K., Nokia, M. S. & Takala, H. Red light optogenetics in neuroscience. *Front. Cell. Neurosci.* **15**, (2022).
- Cembran, A., Bernardi, F., Olivucci, M. & Garavelli, M. Counterion controlled photoisomerization of retinal chromophore models: a computational investigation. <https://doi.org/10.1021/ja048782> (2004).
- El-Tahawy, M. M. T., Nenov, A. & Garavelli, M. Photoelectrochromism in the retinal protonated schiff base chromophore: photoisomerization speed and selectivity under a homogeneous electric field at different operational regimes. *J. Chem. Theory Comput.* **12**, 4460–4475 (2016).
- El-Tahawy, M. M. T., Conti, I., Bonfanti, M., Nenov, A. & Garavelli, M. Tailoring spectral and photochemical properties of bioinspired retinal mimics by in silico engineering. *Angew. Chem. Int. Ed.* **59**, 20619–20627 (2020).
- James, T. Y. et al. A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blas-tocladiomycota). *Mycologia* **98**, 860–871 (2006).
- Broser, M. et al. NeoR, a near-infrared absorbing rhodopsin. *Nat. Commun.* **11**, 5682 (2020).
- Lin, C.-Y., Romei, M. G., Mathews, I. I. & Boxer, S. G. Energetic basis and design of enzyme function demonstrated using gfp, an excited-state enzyme. *J. Am. Chem. Soc.* **144**, 3968–3978 (2022).
- Andruniów, T.; O. M. QM/MM studies of light-responsive biological systems. (Springer, 2021).
- Doukas, A. G. et al. Fluorescence quantum yield of visual pigments: Evidence for subpicosecond isomerization rates (primary event/cis-trans isomerization/excited-state potential surfaces/barrier-less rotation). *Biophysics* **81**, 15 (1984).
- Kralj, J. M., Douglass, A. D., Hochbaum, D. R., MacLaurin, D. & Cohen, A. E. Optical recording of action potentials in mammalian neurons using a microbial rhodopsin. *Nat. Methods* **9**, 90–95 (2012).
- Kouyama, T., Kinosita, K. & Ikegami, A. Excited-state dynamics of bacteriorhodopsin. *Biophys. J.* **47**, 43–54 (1985).
- Engqvist, M. K. M. et al. Directed evolution of gloeobacter violaceus rhodopsin spectral properties. *J. Mol. Biol.* **427**, 205–220 (2015).
- Mclsaac, R. S. et al. Directed evolution of a far-red fluorescent rhodopsin. *Proc. Natl Acad. Sci. USA* **111**, 13034–13039 (2014).
- Silapetere, A. et al. QuasAr Odyssey: the origin of fluorescence and its voltage sensitivity in microbial rhodopsins. *Nat. Commun.* **13**, 5501 (2022).
- Ganapathy, S. et al. Retinal-based proton pumping in the near infrared. *J. Am. Chem. Soc.* **139**, 2338–2344 (2017).
- Lin, C. Y., Romei, M. G., Oltrogge, L. M., Mathews, I. I. & Boxer, S. G. Unified model for photophysical and electro-optical properties of green fluorescent proteins. *J. Am. Chem. Soc.* **141**, 15250–15265 (2019).
- Hochbaum, D. R. et al. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat. Methods* **11**, 825–833 (2014).
- Piatkevich, K. D. et al. A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat. Chem. Biol.* **14**, 352–360 (2018).
- Melaccio, F. et al. Toward automatic rhodopsin modeling as a tool for high-throughput computational photobiology. *J. Chem. Theory Comput.* **12**, 6020–6034 (2016).
- Pedraza-González, L. et al. α -ARM: automatic rhodopsin modeling with chromophore cavity generation, ionization state selection, and external counterion placement. *J. Chem. Theory Comput.* **15**, 3134–3152 (2019).
- Pedraza-González, L., Barneschi, L., Padula, D., de Vico, L. & Olivucci, M. Evolution of the automatic rhodopsin modeling (ARM) protocol. *Top. Curr. Chem.* **380**, 21 (2022).

23. Gholami, S. et al. Multistate multiconfiguration quantum chemical computation of the two-photon absorption spectra of bovine rhodopsin. *J. Phys. Chem. Lett.* **10**, 6293–6300 (2019).
24. (17)Pedraza-González, L.; del C. M. M.; D. V. L.; Y. X.; O. M. *QM/MM Studies of Light-responsive Biological Systems*. (Springer, 2020).
25. Aquilante, F. et al. Modern quantum chemistry with [open]molcas. *J. Chem. Phys.* **152**, 214117 (2020).
26. Mrogiński, M. et al. Frontiers in multiscale modeling of photo-receptor proteins. *Photochem. Photobiol.* **97**, 243–269 (2021).
27. Nakajima, Y. et al. Pro219 is an electrostatic color determinant in the light-driven sodium pump KR2. *Commun. Biol.* **4**, 1185 (2021).
28. Gozem, S., Melaccio, F., Luk, H. L., Rinaldi, S. & Olivucci, M. Learning from photobiology how to design molecular devices using a computer. *Chem. Soc. Rev.* **43**, 4019–4036 (2014).
29. Inoue, K. et al. Red-shifting mutation of light-driven sodium-pump rhodopsin. *Nat. Commun.* **10**, 1993 (2019).
30. Gozem, S., Luk, H. L., Schapiro, I. & Olivucci, M. Theory and simulation of the ultrafast double-bond isomerization of biological chromophores. *Chem. Rev.* **117**, 13502–13565 (2017).
31. Ernst, O. P. et al. Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. *Chem. Rev.* **114**, 126–163 (2014).
32. Gozem, S., Schapiro, I., Ferré, N. & Olivucci, M. The molecular mechanism of thermal noise in rod photoreceptors. *Science (1979)* **337**, 1225–1228 (2012).
33. Altoè, P., Cembran, A., Olivucci, M. & Garavelli, M. Aborted double bicycle-pedal isomerization with hydrogen bond breaking is the primary event of bacteriorhodopsin proton pumping. *Proc. Natl Acad. Sci.* **107**, 20172–20177 (2010).
34. Mei, G. et al. Optical switching between long-lived states of opsin transmembrane voltage sensors. *Photochem. Photobiol.* **97**, 1001–1015 (2021).
35. Sugiura, M. et al. Unusual Photoisomerization Pathway in a Near-Infrared Light Absorbing Enzymic Rhodopsin. *J. Phys. Chem. Lett.* **13**, 9539–9543 (2022).
36. Warshel, A. Bicycle-pedal model for the first step in the vision process. *Nature* **260**, 679–683 (1976).
37. Luk, H. L. et al. Modulation of thermal noise and spectral sensitivity in Lake Baikal cottoid fish rhodopsins. *Sci. Rep.* **6**, 38425 (2016).
38. Atchity, G. J., Xantheas, S. S. & Ruedenberg, K. Potential energy surfaces near intersections. *J. Chem. Phys.* **95**, 1862–1876 (1991).
39. Marín, M. D. C. et al. Fluorescence enhancement of a microbial rhodopsin via electronic reprogramming. *J. Am. Chem. Soc.* **141**, 262–271 (2019).
40. Laricheva, E. N. et al. Origin of fluorescence in 11- cis locked bovine rhodopsin. *J. Chem. Theory Comput.* **8**, 2559–2563 (2012).
41. Liang, R., Yu, J. K., Meisner, J., Liu, F. & Martinez, T. J. Electrostatic control of photoisomerization in channelrhodopsin 2. *J. Am. Chem. Soc.* **143**, 5425–5437 (2021).
42. Wang, W. et al. Tuning the electronic absorption of protein-embedded all- trans -retinal. *Science (1979)* **338**, 1340–1343 (2012).
43. Olsen, S. & McKenzie, R. H. Conical Intersections, charge localization, and photoisomerization pathway selection in a minimal model of a degenerate monomethine dye. *J. Chem. Phys.* **131**, 234306 (2009).
44. Roos, B. O., Taylor, P. R. & Sigbahn, P. E. M. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chem. Phys.* **48**, 157–173 (1980).
45. Cornell, W. D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
46. Hariharan, P. C. & Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **28**, 213–222 (1973).
47. Aquilante, F. et al. MOLCAS 7: the next generation. *J. Comput. Chem.* **31**, 224–247 (2010).
48. Rackers, J. A. et al. Tinker 8: software tools for molecular design. *J. Chem. Theory Comput.* **14**, 5273–5289 (2018).
49. Alex A. Granovsky, Firefly version 8, [www http://classic.chem.msu.su/gran/firefly/index.html](http://classic.chem.msu.su/gran/firefly/index.html).

Acknowledgements

The research has been in part supported by Grants NSF CHE-CLP-1710191 and NIH GM126627-01. The authors are also grateful for a Department of Excellence Grant 2018-2022 funded by the Italian MIUR and by Fondazione Banca d'Italia for providing equipment funds. D.P. acknowledges the Italian Ministry of University and Research (MUR) for a Rita Levi Montalcini grant. We thank Suliman Adam for having provided the NeoR homology model used in this work. I.S. acknowledges support by the DFG through SFB 1078, project C6. I.S. thanks the Israel Science Foundation (Research Center grant no. 3131/20). L.B. and M.O. acknowledge partial support from European Union, Next Generation EU, MUR Italia Domani Progetto mRNA Spoke 6 del "National Center for Gene Therapy and Drugs based on RNA Technology", CUP di progetto B63C22000610006.

Author contributions

M.O. and R.P. conceived and designed the work. R.P. generated the QM/MM models and carried out all the calculations. All the authors R.P., L.B., L.P.G., D.P., I.S., and M.O. contributed to analyzing, discussing, and interpreting the results of the calculations. M.O. and R.P. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-33953-y>.

Correspondence and requests for materials should be addressed to Massimo Olivucci.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

5.2 PAPER 2: Experimental Assessment of the Electronic and Geometrical Structure of a Near-Infrared Absorbing and Highly Fluorescent Microbial Rhodopsin

Publication: Matthias Broser, Tadeusz Andruniów, Anastasia Kraskov, Riccardo Palombo, Sagie Katz, Miroslav Kloz, Jakub Dostál, César Bernardo, John T M Kennis, Peter Hege-
mann, Massimo Olivucci, Peter Hildebrandt, 2023, The Journal of Physical Chemistry Letter,
14(41), 9291-9295.

Contribution: Theoretical calculations, manuscript writing, theoretical data analysis and
graphics generation.

Status: Published on October the 10th, 2023.

Copyright: Reprinted with permission from Broser, Matthias, et al. "Experimental Assess-
ment of the Electronic and Geometrical Structure of a Near-Infrared Absorbing and Highly
Fluorescent Microbial Rhodopsin." The journal of physical chemistry letters 14.41 (2023):
9291-9295. Copyright 2023 American Chemical Society.

Experimental Assessment of the Electronic and Geometrical Structure of a Near-Infrared Absorbing and Highly Fluorescent Microbial Rhodopsin

Matthias Broser,^{*,∇} Tadeusz Andruniow,[∇] Anastasia Kraskov,[∇] Riccardo Palombo, Sagie Katz, Mirosław Kloz, Jakub Dostál, César Bernardo, John T. M. Kennis, Peter Hegemann, Massimo Olivucci,^{*} and Peter Hildebrandt^{*}



Cite This: <https://doi.org/10.1021/acs.jpcllett.3c02167>



Read Online

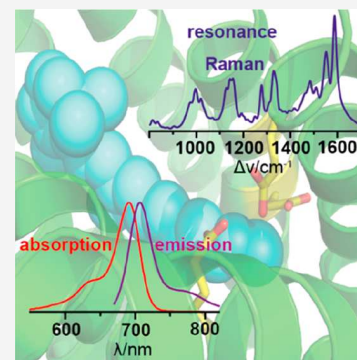
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The recently discovered Neorhodopsin (NeoR) exhibits absorption and emission maxima in the near-infrared spectral region, which together with the high fluorescence quantum yield makes it an attractive retinal protein for optogenetic applications. The unique optical properties can be rationalized by a theoretical model that predicts a high charge transfer character in the electronic ground state (S_0) which is otherwise typical of the excited state S_1 in canonical retinal proteins. The present study sets out to assess the electronic structure of the NeoR chromophore by resonance Raman (RR) spectroscopy since frequencies and relative intensities of RR bands are controlled by the ground and excited states' properties. The RR spectra of NeoR differ dramatically from those of canonical rhodopsins but can be reliably reproduced by the calculations carried out within two different structural models. The remarkable agreement between the experimental and calculated spectra confirms the consistency and robustness of the theoretical approach.



In the past years, we have witnessed a renaissance of research on microbial rhodopsin due to the discovery of new representatives from various species and prospective biomedical applications.^{1,2} In this context, the discovery of red-light-absorbing and highly fluorescent rhodopsins has raised widespread interest as “models” for engineering of more effective optogenetics tools. Moreover, on a fundamental level, such a discovery provides an opportunity to shed new light on how proteins control the properties of embedded chromophores. Despite their diverse functions, microbial rhodopsins share a similar structure and harbor a protonated retinal Schiff base (RSBH⁺) chromophore that is covalently bound to a cavity of the protein and absorbs light in the visible spectral range between 400 and 600 nm.³ The specific absorption maxima of the chromophore are determined by the interactions with the surrounding protein residues.^{4–8} Mainly carboxylates function as counterions and affect the energy levels of the delocalized RSBH⁺ π -electron system and may alter the low fluorescence quantum yield (FQY) between 10^{-5} and 10^{-3} ,⁹ and the photoisomerization quantum yield of 0.2–0.5 of microbial rhodopsins.^{3,10–12}

The canonical properties described above have been challenged by the discovery of neorhodopsin from *Rhizoclostridium globosum* (NeoR), which features an unprecedented red-shifted absorption ($\lambda_{\text{exp}}^{\text{a}} = 690 \text{ nm}$), a dramatically increased FQY of 20%,¹³ and a drastically reduced photoisomerization efficiency of $\sim 3 \times 10^{-4}$ (Figure S1). These

properties do not fit the explanatory models of canonical retinal proteins.

Recent hybrid quantum mechanics/molecular mechanics (QM/MM) studies have suggested that in NeoR, the electrostatic field of the protein induces substantial changes in both the RSBH⁺ ground (S_0) and excited (S_1) electronic states,¹⁴ corresponding to a partial inversion of the covalent and charge-transfer character (cf. Supporting Information, Figure S3), which prevails in the S_0 and S_1 states of canonical rhodopsins, respectively. In canonical rhodopsins the S_0 electronic structure features a positive charge almost completely localized in the chromophore $-\text{C}=\text{N}$ fragment while the S_1 electronic structure displays a positive charge delocalized along the chromophore backbone toward its β -ionone ring.^{15–17} Such an inversion in NeoR would explain the far-red-shifted spectrum, the FQY enhancement, and a reduced photoisomerization efficiency that yielded isomerization around the C7=C8 bond,¹⁸ a process that can only occur as a marginal byproduct in canonical rhodopsins. However, these

Received: August 3, 2023

Accepted: September 28, 2023

changes in the S_0 and S_1 electronic structures lack any experimental assessment.

The present study focuses on resonance Raman (RR) spectroscopy as the critical benchmark for characterizing the unusual NeoR electronic structure. RR spectroscopy is a powerful tool for investigating microbial rhodopsins and has been widely used to study the RSBH⁺ structure and dynamics.^{19–22} The main observables of this technique are frequencies and intensities depending on the details of the S_0 and S_1 states. Thus, comparing the experimental RR with the spectra calculated within the framework of this theoretical approach provides an excellent test for its accuracy and consistency and may specifically verify or reject the postulated inversion in the electronic character of the S_1 and S_0 states.

The RR spectra of NeoR were measured at 80 K with 1064 nm excitation (Figure 1) to achieve sufficiently strong

variant H134 K (ChR2-H134 K) and measured under rigorous resonance conditions.²⁴ Most strikingly, the RR spectra of the two rhodopsins differ drastically. The RR spectrum of ChR2-H134 K, a type of canonical rhodopsin, is dominated by a single band at 1554 cm^{-1} . A small number of distinctly weaker bands between 1100 and 1300 cm^{-1} accompanies it, and there are low-intensity bands at 1007 cm^{-1} , which is essentially invariant in all microbial rhodopsins and at 1661 cm^{-1} . In contrast, the RR spectrum of NeoR displays a complex band pattern between 900 and 1600 cm^{-1} with a series of ~ 20 partially overlapping bands at strong and medium intensities. Also, the spectral changes upon H/D exchange at the Schiff base are strikingly different. While, for ChR2-H134 K, only very few bands are affected, with the furthest shift from 1661 to 1632 cm^{-1} is observed for the C=N stretching of the protonated Schiff base. Significantly more spectral changes are noted in NeoR throughout the region from 900 to 1600 cm^{-1} , but no band attributable to the Schiff base (C15=N) stretching has been detected. In addition, all canonical rhodopsins studied so far obey a linear reciprocal relationship between the electronic absorption maximum and the frequency of the main band C=C stretching mode.^{25,26} This relation holds for absorption wavelengths from the near-UV to the red and even for opsin-bound retinal analogues that absorb in the near-IR.⁷ Clearly, NeoR does not follow this relation.

We are primarily interested in the modes dominated by the C=C and C–C stretching coordinates of the polyene chain. Upon excitation in resonance with a strong electronic transition, the Franck–Condon mechanism (Albrecht's A-term) holds,^{19,27} and thus these modes should be associated with the highest RR intensities since the stretching coordinates are expected to display the highest excited-state displacement measurements Δd , i.e., the difference between the bond lengths in S_1 and S_0 ($\Delta d = d_{S_0} - d_{S_1}$). Hence, we consider Δd in more detail.

In canonical rhodopsins, the S_0 electronic structure of RSBH⁺ shows a covalent character and a positive charge residing in the Schiff base region. Such an electronic structure leads to the typical alternating double and single bond lengths of polyenes with negative and positive Δd for double and single bonds, respectively.²⁸ Therefore, the RR spectra show the highest intensity for the mode that involves the in-phase stretchings of essentially all C=C and C–C coordinates of the polyene chain (1554 cm^{-1} in ChR2-H134 K).^{22,29} Additional modes involving C=C stretching coordinates are restricted to the region between 1500 and 1650 cm^{-1} , whereas modes with C–C but not C=C stretching character are clustered between 1100 and 1300 cm^{-1} .

The present calculations of the RR spectra rely on a QM/MM model (reported in ref 14, in the following referred to as MDR) built from a homology structure predicted by Modeler³⁰ (see Section 1.3.2 in the Supporting Information), that has served as the basis for the computational modeling of NeoR in multiple studies.^{13,14} The MDR model for NeoR reveals an unusual RSBH⁺ structure showing an alternating pattern of single and double bond lengths only in the region close to the β -ionone ring (i.e., C1=C6 to C9=C10). Instead, upon approaching the Schiff base terminus (C15=N), the length differences between single and double bonds decrease as the C12–C13 and C14–C15 bond lengths become shorter than those of the C13=C14 bond (Figure 2).

Concomitantly, in NeoR, the bond length alternation in the S_1 state is damped in the same direction from the β -ionone

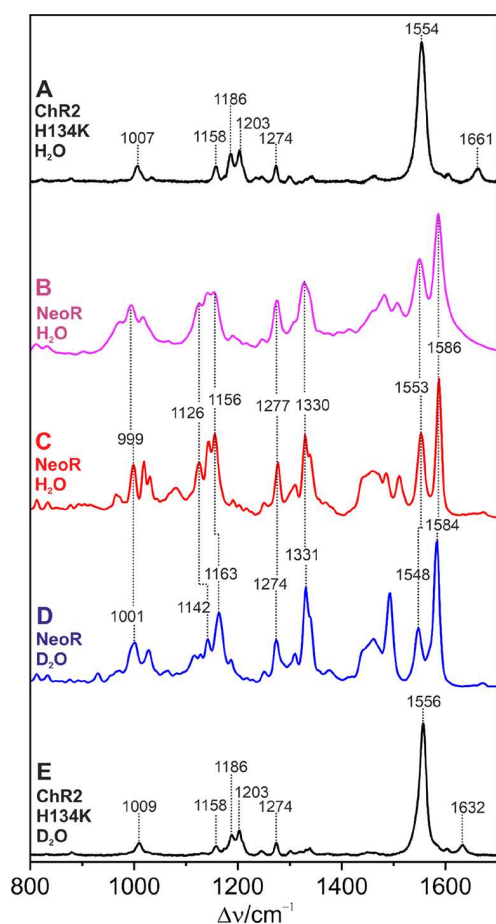


Figure 1. RR spectra of the parent (dark) states of NeoR in H₂O (C, red trace) and D₂O (D, blue trace) compared with those of canonical rhodopsin (ChR2-H134 K: A, E; black traces) from previous work.²⁴ Trace B shows the stimulated Raman spectrum of NeoR measured with 800 nm excitation.

resonance enhancement, such that the spectra exclusively display the RR bands of RSBH⁺ without any interference from the protein. This is also illustrated by the agreement with the stimulated Raman spectrum, measured with a probe beam at 800 nm,²³ and thus even in closer resonance with the $S_0 \rightarrow S_1$ transition.

In comparison, Figure 1 also includes the RR spectrum of the all-*trans* isomer of dark-adapted channelrhodopsin-2

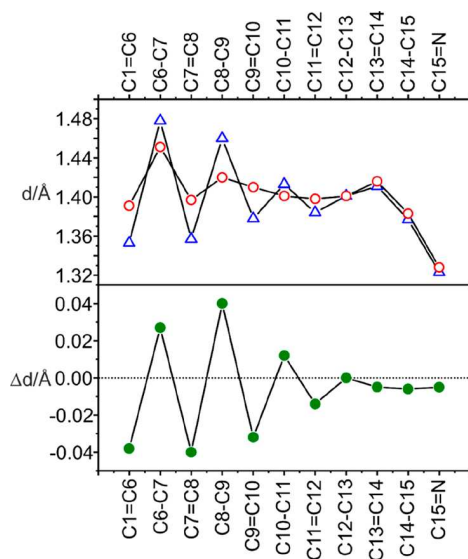


Figure 2. Top, calculated bond lengths of the polyene chain of the retinal chromophore of the MDR model. The open triangles and circles refer to the S_0 ground and S_1 excited states, respectively. Bottom, the bond length differences $\Delta d = d_{S_0} - d_{S_1}$ (green) representing the excited state displacements of the stretching coordinates. The red and blue colors refer to double and single bonds, respectively.

ring to the Schiff base, and thus Δd decreases and eventually approaches zero. Thus, the MDR model points to opsin electrostatics (i.e., residue charge distribution), creating a predominantly delocalized charge along the RSBH⁺ chain that can also be described in terms of a high percentage of charge transfer (CT) already in the S_0 state.¹⁴ Such a state refers to the typical S_1 charge distribution found in canonical rhodopsins where the positive charge, usually located on the chromophore $-C=NH-$ moiety, is delocalized toward the β -ionone ring (see ref 16).

The unique electronic and structural properties of the S_0 and S_1 states of NeoR have distinct consequences for the RR spectra (Figure 3).

Both C=C and C-C stretchings contribute notably to a much greater number of normal modes in a large spectral region, from 1100 to 1650 cm^{-1} . Furthermore, the prominent RR bands are due to modes with significant contributions of two stretching coordinates on average. Among them is the C7=C8 stretching, which is involved in most of the modes with high RR intensity. Therefore, the intense band at 1586 cm^{-1} observed in NeoR appears to be a consequence of the CT character of S_0 RSBH⁺. Notably, a mode with contributions of the C15=N stretching is calculated at 1646 cm^{-1} (see Table S3) and thus at a frequency similar to that of canonical rhodopsins. However, consistent with the experimental observation, essentially no RR intensity is predicted for this mode, probably because the two main stretching coordinates involved (i.e., C15=N and C14-C15) have Δd values of nearly zero. In conclusion, the calculations reproduce the experimental RR spectra in both H₂O and D₂O solutions (Figure 3, Figure S7).

We carried out our calculations of NeoR using the previously reported MDR model,¹⁴ which was based on homology modeling (see Section 1.3.1 in the Supporting Information). To reduce any model-linked bias, we constructed a second QM/MM model (AP2) from a structure

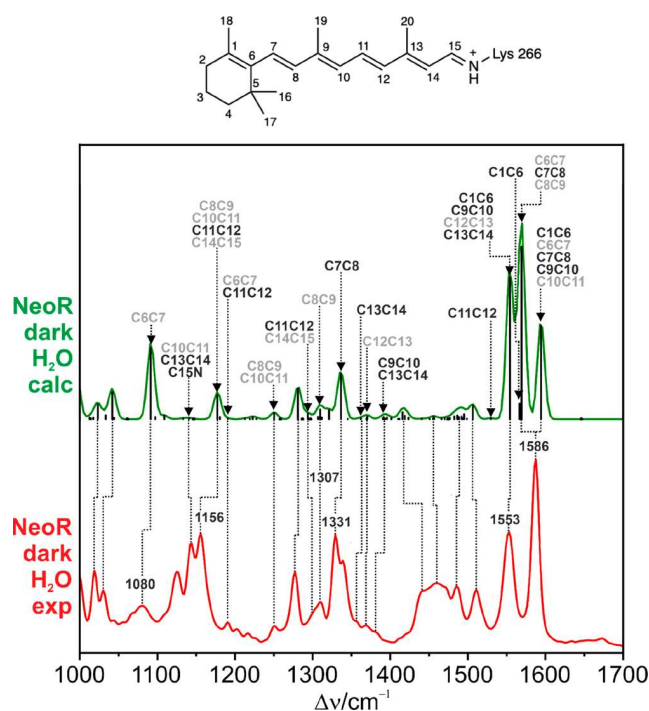


Figure 3. Calculated and experimental RR spectrum of NeoR in H₂O. The calculated spectrum was generated by Gaussian band shapes of the vibrational transition derived from the QM partition of the MDR model (green trace). The spectrum also includes the line presentation of the vibrational transitions. Frequencies were scaled by a factor of 0.9. Modes with major C=C and C-C stretching characters (factoring to >50% of the contribution of a coordinate) are indicated by black and gray letters, respectively. The notation follows the numbering of the structural formula on top of the figure. Details of the calculated RSBH⁺ modes are listed in Table S3 (Supporting Information, Section 3.2).

predicted by AlphaFold2.³¹ The main differences between the two models refer to residues E141 and W234. In the MDR model, E141 is deprotonated (anionic), hydrogen bonded to W234, and points away from RSBH⁺. In the AP2 model, E141 is neutral and approaches closer to the RSBH⁺ (Figure 4; Figure S2). This different charge distribution leads to increased localization of the positive charge on the Schiff base unit in MDR concerning AP2, likely causing the slightly different bond length pattern in the Schiff base region (Figure 2, Figure S3; a comprehensive comparison of both models is given in the Supporting Information). The different bond length pattern specifically refers to the C14-C15 and C15=N bonds for which the MDR model yields $\Delta d \approx 0$ in contrast to the appreciable Δd value derived from the AP2 model (Figure S4). As a result, only the MDR model correctly predicts the lack of RR intensity for the Schiff base stretching mode at 1648 cm^{-1} (*vide supra*). However, besides this mode, the AP2 model also provides a satisfactory simulation of the experimental RR spectra (Figures S8 and S9).

In both models, the S_1 potential energy surface (PES) mapping shows high energy barriers ($E_{S_1}^{\ddagger}$) along both the canonical C13=C14 and C11=C12 photoisomerization paths (reported in Figure S6). This promptly explains the dramatic increase in FQY assigned to a hindered S_1 isomerization motion. A high $E_{S_1}^{\ddagger}$ also points to the opening of alternative C7=C8 and C9=C10 isomerization routes present in the MDR and AP2 models associated with the

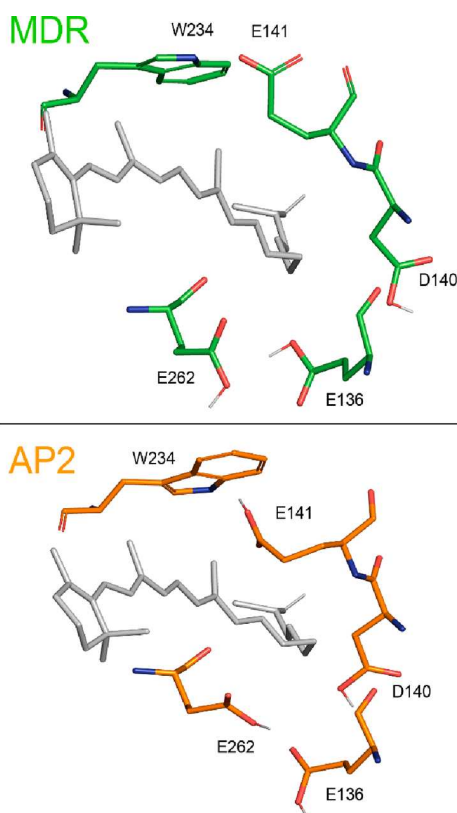


Figure 4. NeoR retinal binding cavities of the MDR and the AP2 model, with E136, D140, and E262 protonated in both models. In the MDR, deprotonated E141 forms a hydrogen bond with W234. This interaction is missing in AP2 with protonated (neutral) E141 (for more details, see Figure S2).

lowest barriers, even though they still amount to ~ 15 kcal/mol $E_{S_1}^f$ (Table S1). This prediction is in line with the photoisomerization from all-*trans* to 7-*cis* upon irradiation of NeoR with red light (Figure S10), whereas no isomerization around the C13=C14 double bond could be detected.¹⁸

In conclusion, this study provides experimental support for the unusual RSBH⁺ electronic structure of NeoR predicted by two different QM/MM models. The fact that the models reproduce the observed RR spectrum implies that the chromophore S_0 and S_1 electronic structures are well described. Thus, unlike canonical microbial rhodopsins, RgNeoR (*R. globosum*) and probably all related NeoRs exhibit an S_0 state with strong CT character, limiting a clear bond length alternation pattern to the segment close to the β -ionone ring. Conversely, the same theoretical methodology allows to reproduce the experimental RR spectra and excited state properties of canonical rhodopsins¹⁵ with their “classical” charge distribution in the ground and excited states.^{16,17} The charge distribution in NeoR significantly impacts the S_1 geometrical displacement, progression, and, ultimately, its reactivity. The present results provide strong support for the “chromophore charge delocalization and confinement” mechanism underlying the origin of the NeoR high energy barrier along the C13=C14 photoisomerization path and explaining the observed unprecedented FQY value of this microbial rhodopsin.^{13,14}

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcclett.3c02167>.

Additional experimental and computational details, materials, and methods, including a comparison of the MDR and AP2 models referred to in the text (PDF) Transparent Peer Review report available (PDF)

AUTHOR INFORMATION

Corresponding Authors

Matthias Broser – Institut für Biologie, Experimentelle Biophysik, Humboldt-Universität zu Berlin, D-10115 Berlin, Germany; orcid.org/0000-0002-7133-4548; Email: matthias.broser@hu-berlin

Massimo Olivucci – Dipartimento di Biotechnologie, Università di Siena, 53100 Siena, Italy; Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43403, United States; orcid.org/0000-0002-8247-209X; Email: molivuc@bgsu.edu

Peter Hildebrandt – Institute für Chemie, Technische Universität Berlin, D-10623 Berlin, Germany; orcid.org/0000-0003-1030-5900; Email: peter.hildebrandt@tu-berlin.de

Authors

Tadeusz Andruniow – Department of Chemistry, Wrocław University of Science and Technology, 50-370 Wrocław, Poland; orcid.org/0000-0002-2215-5287

Anastasia Kraskov – Institute für Chemie, Technische Universität Berlin, D-10623 Berlin, Germany

Riccardo Palombo – Dipartimento di Biotechnologie, Università di Siena, 53100 Siena, Italy

Sagie Katz – Institute für Chemie, Technische Universität Berlin, D-10623 Berlin, Germany

Miroslav Klotz – ELI Beamlines Facility, The Extreme Light Infrastructure ERIC, 25241 Dolní Břežany, Czech Republic

Jakub Dostál – ELI Beamlines Facility, The Extreme Light Infrastructure ERIC, 25241 Dolní Břežany, Czech Republic

César Bernardo – ELI Beamlines Facility, The Extreme Light Infrastructure ERIC, 25241 Dolní Břežany, Czech Republic

John T. M. Kennis – Department of Physics and Astronomy, Faculty of Science, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands; orcid.org/0000-0002-3563-2353

Peter Hegemann – Institut für Biologie, Experimentelle Biophysik, Humboldt-Universität zu Berlin, D-10115 Berlin, Germany; orcid.org/0000-0003-3589-6452

Complete contact information is available at <https://pubs.acs.org/10.1021/acs.jpcclett.3c02167>

Author Contributions

[▽](M.B., T.A., and A.K.) These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Melanie Meiworm and Christina Schnick for their excellent technical assistance. This work was supported by the German Research Foundation, DFG Grant No. 509731234 to M.B., CRC1078 “Protonation Dynamics in Protein Function”, project number 221545957, subproject B06 to P.Hi., and the

299 DFG under Germany's Excellence Strategy—EXC 311 2008/1
300 (UniSysCat)—390540038 (Research Unit E to P.He. and
301 P.Hi.). P. He. is a Hertie Professor and is supported by the
302 Hertie Foundation. M.O. acknowledges support from NSF
303 CSDM-A Grant No. 2102619 and European-Union, Next
304 Generation EU, MIUR Italia Domani Progetto mRNA Spoke 6
305 del "National Center for Gene Therapy and Drugs based on
306 RNA Technology"—CUP B63C22000610006. M.K. acknowl-
307 edges support from the Czech Science Foundation (Project
308 No. 21-09692M). M.B., J.T.M.K., and P. He. acknowledge
309 beam time access at ELI.

310 ■ REFERENCES

- 311 (1) Kandori, H. Retinal Proteins: Photochemistry and Optogenetics.
312 *Bull. Chem. Soc. Jpn.* **2020**, *93*, 76–85.
- 313 (2) Govorunova, E. G.; Koppel, L. A. The Road to Optogenetics:
314 Microbial Rhodopsins. *Biochemistry (Mosc)* **2016**, *81* (9), 928–940.
- 315 (3) Ernst, O. P.; Lodowski, D. T.; Elstner, M.; Hegemann, P.;
316 Brown, L. S.; Kandori, H. Microbial and Animal Rhodopsins:
317 Structures, Functions, and Molecular Mechanisms. *Chem. Rev.* **2014**,
318 *114* (1), 126–163.
- 319 (4) Kochendoerfer, G. G.; Verdegem, P. J.; van der Hoef, I.;
320 Lugtenburg, J.; Mathies, R. A. Retinal Analog Study of the Role of
321 Steric Interactions in the Excited State Isomerization Dynamics of
322 Rhodopsin. *Biochemistry* **1996**, *35* (50), 16230–16240.
- 323 (5) Engqvist, M. K.; McIsaac, R. S.; Dollinger, P.; Flytzanis, N. C.;
324 Abrams, M.; Schor, S.; Arnold, F. H. Directed Evolution of
325 *Gloeobacter Violaceus* Rhodopsin Spectral Properties. *J. Mol. Biol.*
326 **2015**, *427* (1), 205–220.
- 327 (6) McIsaac, R. S.; Engqvist, M. K.; Wannier, T.; Rosenthal, A. Z.;
328 Herwig, L.; Flytzanis, N. C.; Imasheva, E. S.; Lanyi, J. K.; Balashov, S.
329 P.; Gradinaru, V.; et al. Directed Evolution of a Far-Red Fluorescent
330 Rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (36), 13034–
331 13039.
- 332 (7) Hontani, Y.; Ganapathy, S.; Frehan, S.; Kloz, M.; de Grip, W. J.;
333 Kennis, J. T. M. Strong Ph-Dependent near-Infrared Fluorescence in a
334 Microbial Rhodopsin Reconstituted with a Red-Shifting Retinal
335 Analogue. *J. Phys. Chem. Lett.* **2018**, *9* (22), 6469–6474.
- 336 (8) Laricheva, E. N.; Gozem, S.; Rinaldi, S.; Melaccio, F.; Valentini,
337 A.; Olivucci, M. Origin of Fluorescence in 11-Cis Locked Bovine
338 Rhodopsin. *J. Chem. Theory Comput.* **2012**, *8* (8), 2559–2563.
- 339 (9) Penzkofer, A.; Silapetere, A.; Hegemann, P. Absorption and
340 Emission Spectroscopic Investigation of the Thermal Dynamics of the
341 Archaerhodopsin 3 Based Fluorescent Voltage Sensor Quasar1. *Int. J.*
342 *Mol. Sci.* **2019**, *20* (17), 4086.
- 343 (10) van Stokkum, I. H. M.; Hontani, Y.; Vierock, J.; Krause, B. S.;
344 Hegemann, P.; Kennis, J. T. M. Reaction Dynamics in the Chrimson
345 Channelrhodopsin: Observation of Product-State Evolution and Slow
346 Diffusive Protein Motions. *J. Phys. Chem. Lett.* **2023**, *14* (6), 1485–
347 1493.
- 348 (11) Hontani, Y.; Marazzi, M.; Stehfest, K.; Mathes, T.; van
349 Stokkum, I. H. M.; Elstner, M.; Hegemann, P.; Kennis, J. T. M.
350 Reaction Dynamics of the Chimeric Channelrhodopsin C1c2. *Sci. Rep.*
351 **2017**, *7* (1), 7217.
- 352 (12) Malakar, P.; Das, I.; Bhattacharya, S.; Harris, A.; Sheves, M.;
353 Brown, L. S.; Ruhman, S. Bidirectional Photochemistry of Antarctic
354 Microbial Rhodopsin: Emerging Trend of Ballistic Photoisomeriza-
355 tion from the 13-Cis Resting State. *J. Phys. Chem. Lett.* **2022**, *13* (34),
356 8134–8140.
- 357 (13) Broser, M.; Spreen, A.; Konold, P. E.; Peter, E.; Adam, S.;
358 Borin, V.; Schapiro, I.; Seifert, R.; Kennis, J. T. M.; Bernal Sierra, Y.
359 A.; et al. Neor, a near-Infrared Absorbing Rhodopsin. *Nat. Commun.*
360 **2020**, *11* (1), 5682.
- 361 (14) Palombo, R.; Barneschi, L.; Pedraza-González, L.; Padula, D.;
362 Schapiro, I.; Olivucci, M. Retinal Chromophore Charge Delocaliza-
363 tion and Confinement Explain the Extreme Photophysics of
364 Neorhodopsin. *Nat. Commun.* **2022**, *13* (1), 6652.
- (15) Yang, X.; Manathunga, M.; Gozem, S.; Léonard, J.; Andruniów, 365
T.; Olivucci, M. Quantum-Classical Simulations of Rhodopsin Reveal 366
Excited-State Population Splitting and Its Effects on Quantum 367
Efficiency. *Nat. Chem.* **2022**, *14* (4), 441–449. 368
- (16) Gozem, S.; Luk, H. L.; Schapiro, I.; Olivucci, M. Theory and 369
Simulation of the Ultrafast Double-Bond Isomerization of Biological 370
Chromophores. *Chem. Rev.* **2017**, *117* (22), 13502–13565. 371
- (17) Bonačić-Koutecký, V.; Schöffel, K.; Michl, J. Critically 372
Heterosymmetric Biradicaloid Geometries of Protonated Schiff 373
Bases. *Theor. Chim. Acta* **1987**, *72* (5), 459–474. 374
- (18) Sugiura, M.; Ishikawa, K.; Katayama, K.; Sumii, Y.; Abe- 375
Yoshizumi, R.; Tsunoda, S. P.; Furutani, Y.; Shibata, N.; Brown, L. S.; 376
Kandori, H. Unusual Photoisomerization Pathway in a near-Infrared 377
Light Absorbing Enzymerhodopsin. *J. Phys. Chem. Lett.* **2022**, *13* (40), 378
9539–9543. 379
- (19) Buhrke, D.; Hildebrandt, P. Probing Structure and Reaction 380
Dynamics of Proteins Using Time-Resolved Resonance Raman 381
Spectroscopy. *Chem. Rev.* **2020**, *120* (7), 3577–3630. 382
- (20) Althaus, T.; Eisfeld, W.; Lohrmann, R.; Stockburger, M. 383
Application of Raman Spectroscopy to Retinal Proteins. *Isr. J. Chem.* 384
1995, *35* (3–4), 227–251. 385
- (21) Smith, S. O.; Lugtenburg, J.; Mathies, R. A. Determination of 386
Retinal Chromophore Structure in Bacteriorhodopsin with Resonance 387
Raman Spectroscopy. *J. Membr. Biol.* **1985**, *85* (2), 95–109. 388
- (22) Smith, S. O.; Braiman, M. S.; Myers, A. B.; Pardo, J. A.; 389
Courtin, J. M. L.; Winkel, C.; Lugtenburg, J.; Mathies, R. A. 390
Vibrational Analysis of the All-Trans-Retinal Chromophore in Light- 391
Adapted Bacteriorhodopsin. *J. Am. Chem. Soc.* **1987**, *109* (10), 3108– 392
3125. 393
- (23) Kloz, M.; Weissenborn, J.; Polívka, T.; Frank, H. A.; Kennis, J. 394
T. M. Spectral Watermarking in Femtosecond Stimulated Raman 395
Spectroscopy: Resolving the Nature of the Carotenoid S* State. *Phys.* 396
Chem. Chem. Phys. **2016**, *18* (21), 14619–14628. 397
- (24) Bruun, S.; Stoeppel, D.; Keidel, A.; Kuhlmann, U.; Luck, M.; 398
Diehl, A.; Geiger, M.-A.; Woodmansee, D.; Trauner, D.; Hegemann, 399
P.; et al. Light-Dark Adaptation of Channelrhodopsin Involves 400
Photoconversion between the All-Trans and 13-Cis Retinal Isomers. 401
Biochemistry **2015**, *54* (35), 5389–5400. 402
- (25) Doukas, A. G.; Aton, B.; Callender, R. H.; Ebrey, T. G. 403
Resonance Raman Studies of Bovine Metarhodopsin I and 404
Metarhodopsin II. *Biochemistry* **1978**, *17* (12), 2430–2435. 405
- (26) Kajimoto, K.; Kikukawa, T.; Nakashima, H.; Yamaryo, H.; 406
Saito, Y.; Fujisawa, T.; Demura, M.; Unno, M. Transient Resonance 407
Raman Spectroscopy of a Light-Driven Sodium-Ion-Pump Rhodopsin 408
from *Indibacter Alkaliphilus*. *J. Phys. Chem. B* **2017**, *121* (17), 4431– 409
4437. 410
- (27) Albrecht, A. C. On the Theory of Raman Intensities. *J. Chem.* 411
Phys. **1961**, *34* (5), 1476–1484. 412
- (28) Gozem, S.; Johnson, P. J. M.; Halpin, A.; Luk, H. L.; Morizumi, 413
T.; Prokhorenko, V. I.; Ernst, O. P.; Olivucci, M.; Miller, R. J. D. 414
Excited-State Vibronic Dynamics of Bacteriorhodopsin from Two- 415
Dimensional Electronic Photon Echo Spectroscopy and Multi- 416
configurational Quantum Chemistry. *J. Phys. Chem. Lett.* **2020**, *11* 417
(10), 3889–3896. 418
- (29) Babitzki, G.; Mathias, G.; Tavan, P. The Infrared Spectra of the 419
Retinal Chromophore in Bacteriorhodopsin Calculated by a Dft/Mm 420
Approach. *J. Phys. Chem. B* **2009**, *113* (30), 10496–10508. 421
- (30) Šali, A.; Blundell, T. L. Comparative Protein Modelling by 422
Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234* (3), 779–815. 423
- (31) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; 424
Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; 425
Potapenko, A.; et al. Highly Accurate Protein Structure Prediction 426
with Alphafold. *Nature* **2021**, *596* (7873), 583–589. 427

5.3 PAPER 3: Picosecond Quantum-classical Dynamics reveals that the Coexistence of Light-induced Microbial and Animal Chromophore Rotary Motion Modulates the Isomerization Quantum Yield of Heliorhodopsin

Authors: Riccardo Palombo, Leonardo Barneschi, Laura Pedraza-González, Xuchun Yang and Massimo Olivucci.

Contribution: Research design, calculations, manuscript writing and revision, data analysis and graphics generation.

Status: Published on March the 12th, 2024.



Cite this: *Phys. Chem. Chem. Phys.*,
2024, 26, 10343

Picosecond quantum-classical dynamics reveals that the coexistence of light-induced microbial and animal chromophore rotary motion modulates the isomerization quantum yield of heliorhodopsin†

Riccardo Palombo,^{ab} Leonardo Barneschi,^a Laura Pedraza-González,^c
Xuchun Yang^b and Massimo Olivucci^{*ab}

Rhodopsins are light-responsive proteins forming two vast and evolutionary distinct superfamilies whose functions are invariably triggered by the photoisomerization of a single retinal chromophore. In 2018 a third widespread superfamily of rhodopsins called heliorhodopsins was discovered using functional metagenomics. Heliorhodopsins, with their markedly different structural features with respect to the animal and microbial superfamilies, offer an opportunity to study how evolution has manipulated the chromophore photoisomerization to achieve adaptation. One question is related to the mechanism of such a reaction and how it differs from that of animal and microbial rhodopsins. To address this question, we use hundreds of quantum-classical trajectories to simulate the spectroscopically documented picosecond light-induced dynamics of a heliorhodopsin from the archaea thermoplasmatales archaeon (TaHeR). We show that, consistently with the observations, the trajectories reveal two excited state decay channels. However, inconsistently with previous hypotheses, only one channel is associated with the $-C13=C14-$ rotation of microbial rhodopsins while the second channel is characterized by the $-C11=C12-$ rotation typical of animal rhodopsins. The fact that such $-C11=C12-$ rotation is aborted upon decay and ground state relaxation, explains why illumination of TaHeR only produces the 13-*cis* isomer with a low quantum efficiency. We argue that the documented lack of regioselectivity in double-bond excited state twisting motion is the result of an “adaptation” that could be completely lost *via* specific residue substitutions modulating the steric hindrance experienced along the isomerization motion.

Received 16th January 2024,
Accepted 11th March 2024

DOI: 10.1039/d4cp00193a

rsc.li/pccp

Introduction

Rhodopsins are a class of transmembrane proteins constituted by a seven α -helices apoprotein (opsin) hosting a retinal protonated Schiff base chromophore (rPSB) covalently linked to a lysine residue. rPSB absorbs light and transduces it into nuclear motion leading to the breaking, isomerization and reconstitution of a specific C=C double-bond. The resulting rPSB isomer triggers the formation of opsin conformers that, ultimately, drive biological functions such as vision in

vertebrates; light-sensing, phototaxis, photoprotection, ion-pumping in eubacteria and archaea; ion-gating in algae and catalytic activity in fungi, algae, and choanoflagellates, just to cite a few.^{1–6} Such diverse functions emerge from the drastic differences in the protein amino acid sequence. The effect of such differences is already evident at the rPSB isomerization level that, in fact, displays different regioselectivity, stereoselectivity and quantum yield (φ_{iso}) in different rhodopsins.⁷ The study of the diversity of the rPSB isomerization and its efficiency is therefore an unavoidable milestone along the path to exploit rhodopsin functions. More specifically, it has become apparent that the investigation of the molecular-level factors controlling φ_{iso} ^{8,9} in diverse rhodopsins would have an impact on the design of effective optogenetic tools^{10,11} for neurosciences^{12–14} and of novel synthetic biology circuits allowing, for instance, for artificial CO₂ fixation.^{15,16} Furthermore, the investigation of such factors is needed to understand how in photic zones such as the upper ocean,¹⁷ specific microbial rhodopsins massively

^a Dipartimento di Biotecnologie, Chimica e Farmacia, Università di Siena, via A. Moro 2, I-53100 Siena, Siena, Italy. E-mail: olivucci@unisi.it

^b Department of Chemistry, Bowling Green State University, Bowling Green, Ohio 43403, USA. E-mail: molivuc@bgsu.edu

^c Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Giuseppe Moruzzi, 13, I-56124 Pisa, Italy

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d4cp00193a>

contribute to the solar energy capture.¹⁸ Despite these facts, only a limited number of studies have targeted the rPSB isomerization diversity in both animal^{19–21} and microbial^{22–27} rhodopsins. For this reason, we present a computational study of the photoisomerization mechanism and dynamics of a member of a recently discovered rhodopsin family and compare the results to those obtained for the evolutionarily distant dim-light visual pigment of vertebrates.

The discovery of heliorhodopsins (HeRs) in Lake Tiberias (Israel), then readily detected in eubacteria, archaea, algae and viruses in both freshwater and saltwater environments worldwide, dramatically expanded the known rhodopsin diversity.²⁸ HeRs have an inverted membrane orientation topology exposing the C-terminus at the extracellular side, rather than the N-terminus as is the case for animal and microbial rhodopsin homologues. Furthermore, they share no more than 15% of sequence identity with both the former and the latter. Different HeR functions have been proposed. For instance, although the slow HeR photocycle could indicate a potential light sensory activity,^{28,29} other novel metabolic functions such as light driven transport of amphiphilic molecules³⁰ and DNA protection against light-induced oxidative damage have been proposed.³¹ In addition, certain viral HeRs have been demonstrated to serve as proton channels, making them possible template for the development of optogenetic tools.³² Finally, the growing interest on HeRs is also motivated by their possible exploitation as light-driven redox enzymes possibly capable of carbon fixation.³³

The first crystallographic structure of one representative of HeRs family, the thermoplasmatales archaeon heliorhodopsin (TaHeR), was released in 2019 under the PDB code 6IS6.³⁴ Like canonical microbial (or Type 1) rhodopsins, the TaHeR function was shown to be initiated by the all-*trans* to 13-*cis* photoisomerization of the rPSB chromophore occurring on a sub-picosecond timescale.³⁵ This event triggers an internal proton transfer²⁸ event (*i.e.* the proton is never released from the protein) toward the counterion (Glu108) of the rPSB chromophore and then, according to a recent computational study³⁶ based on HeR 48C12 (43% of sequence identity with TaHeR), toward a nearby histidine pair, namely His23 and His80. The photoisomerization quantum yield (φ_{iso}) does not exceed 0.20.³⁵ Strikingly, this is a small value compared to 0.64 for the canonical proton-pump bacteriorhodopsin³⁷ as well as 0.67 for dim-light visual pigments.³⁸ While such differences point to a diverse photoisomerization mechanism in TaHeR and, possibly, in HeRs in general, it provides the opportunity to study how evolutionarily distant opsins modulate φ_{iso} and, therefore, an “elementary” chemical process such as the π -bond breaking.

Below we use quantum-classical population dynamics to study TaHeR. More specifically, we study the rPSB photoisomerization reaction by constructing, starting from the crystallographic structure, a hybrid multiconfigurational quantum mechanics/molecular mechanics (QM/MM) model of the protein and using it to propagate hundreds of Tully’s quantum-classical trajectories³⁹ for over two picoseconds. The results are then compared with those of the rPSB photoisomerization reaction

occurring in the dim-light visual rhodopsin expressed in the vertebrate *Bos taurus* (Rh) as an example of highly efficient 11-*cis* to all-*trans* photoisomerization of a canonical animal (*i.e.* Type 2) rhodopsin.

We show that, while in Rh the rPSB photoisomerization is accomplished through a single coordinate describing the counterclockwise (CCW) π -bond breaking of the C11=C12 double-bond typical of animal rhodopsins, in TaHeR the isomerization motion follows simultaneously two coordinates corresponding to the clockwise (CW) and CCW π -bond breaking of the C13=C14 and C11=C12 double-bonds, respectively. This result has two consequences; firstly, it provides an atomistic reinterpretation of recent transient absorption spectroscopy (TAS) measurements,³⁵ which pointed to the existence of two ultrafast decay channels, but both assigned to C13=C14 isomerization. Secondly, it reveals a previously unreported and effective way in which HeRs control φ_{iso} . In fact, the same quantum-classical simulations, show that the absence of the C11=C12 isomerization product in TaHeR is explained by a ground state mechanism that selectively inverts the animal-like isomerization motion initiated in the excited state and leads to reconstitution of the original all-*trans* rPSB reactant.

Results and discussion

Fig. 1 introduces the isomerization channels discussed for TaHeR and Rh in the following subsections. Light absorption promotes the rPSB in the all-*trans* (in TaHeR) or 11-*cis* (in Rh) configuration, from the ground electronic state (S_0) to the Franck–Condon (FC) region of the first excited (S_1) state. The $S_0 \rightarrow S_1$ change in electronic character induces an ultrafast relaxation that steers rPSB toward one or more S_1/S_0 conical intersections (CoIn’s) featuring a *ca.* 90° twisted double-bond. Each CoIn provides access to the S_0 potential energy surface (PES) and, ultimately, to photoproduct formation. The timescale of this process defines the S_1 lifetime while the branching along the isomerization coordinates associated with different skeletal C10–C11–C12–C13 and C12–C13–C14–C15 dihedral angles (α , α' in Fig. 1A and B) determines the isomerization regioselectivity. As we will discuss below, when compared to Rh that propagates exclusively along α , TaHeR is *non-regioselective* since the S_1 population splits in a bifurcation region (BF in Fig. 1C) initiating parallel α , α' propagations reaching the CoIn₁₁ and CoIn₁₃ channels, respectively. In Fig. 1B we also define τ , a function of the skeletal torsion α and substituents wag β proportional to the π -overlap of the isomerizing double bond. In the following, τ is used to discuss the bond-breaking and bond-making process. As schematically shown in Fig. 1C, from each CoIn, two S_0 relaxation paths can be accessed leading either to the photoproduct or to the reactant reconstitution. Thus, in TaHeR where only the C13=C14 isomerization product has been observed, the isomerization quantum efficiency ($\varphi_{\text{iso},13}$) must be a composite quantity that depends on both the S_1 population splitting at BF and the branching at CoIn₁₃. The description of the “inversion mechanism”

selectively blocking the C11=C12 isomerization (see dashed curly red arrow in Fig. 1C, left) and explaining its observed negligible $\varphi_{\text{iso},11}$, is one of the main target of the present research.

S_1 lifetime, reaction regioselectivity, $\varphi_{\text{iso},13}$ and $\varphi_{\text{iso},11}$ of TaHeR and Rh are calculated by simulating their light-induced dynamics starting from room-temperature S_0 populations (details are provided in Sections S1 and S2, ESI[†]). Based on a recent study by some of us, we assume that 200 trajectories are suitable for mechanistic investigations as well as for comparative studies of diverse systems.⁹ The results are documented and discussed in Sections A–D. In A we validate the QM/MM models of TaHeR and Rh by comparing a set of calculated and experimentally observed spectral quantities. In B we report on the dramatically different time progression of the two corresponding photoexcited populations, showing that, in TaHeR, two structurally adjacent isomerization channels are equally populated. In C, we provide evidence indicating that stereochemically opposite aborted bicycle-pedal isomerization mechanisms characterize the two channels in TaHeR. Finally, in D, we rationalize the computed $\varphi_{\text{iso},13}$ and $\varphi_{\text{iso},11}$ values

using the theoretical framework based on functions of τ : a parameter proportional to π -orbital overlap velocity of the reacting π -bond at the point of S_1 decay ($d\tau/dt^{\text{decay}}$).

A. Absorption spectroscopy and model validation

The QM/MM models of TaHeR and Rh were constructed using the automatic rhodopsin modelling protocol,^{40–43} α -ARM (see also Methods and Section S1, ESI[†]). A general overview of the two models is provided in Fig. 2A and B. The flexible part (in red) of the modelled chromophore cavities not only displays very different amino acidic compositions (see also Fig. S1, ESI[†]), but also a different three-dimensional shape. Notably, the cavity of TaHeR features a lateral fenestration visible above the β -ionone ring of the rPSB chromophore which exposes the chromophore itself to the external environment.³⁴ As shown in Fig. 2C and D, the constructed TaHeR and Rh models yielded FC vertical excitation energies (here expressed in terms of absorption maxima, λ_{max}^a) only 12 nm blue shifted (530 nm) and 1 nm red shifted (499 nm) relative to the observed values, respectively. Furthermore, by using the wild-type QM/MM models as starting point for constructing a set of mutant models,⁴⁴ it was possible

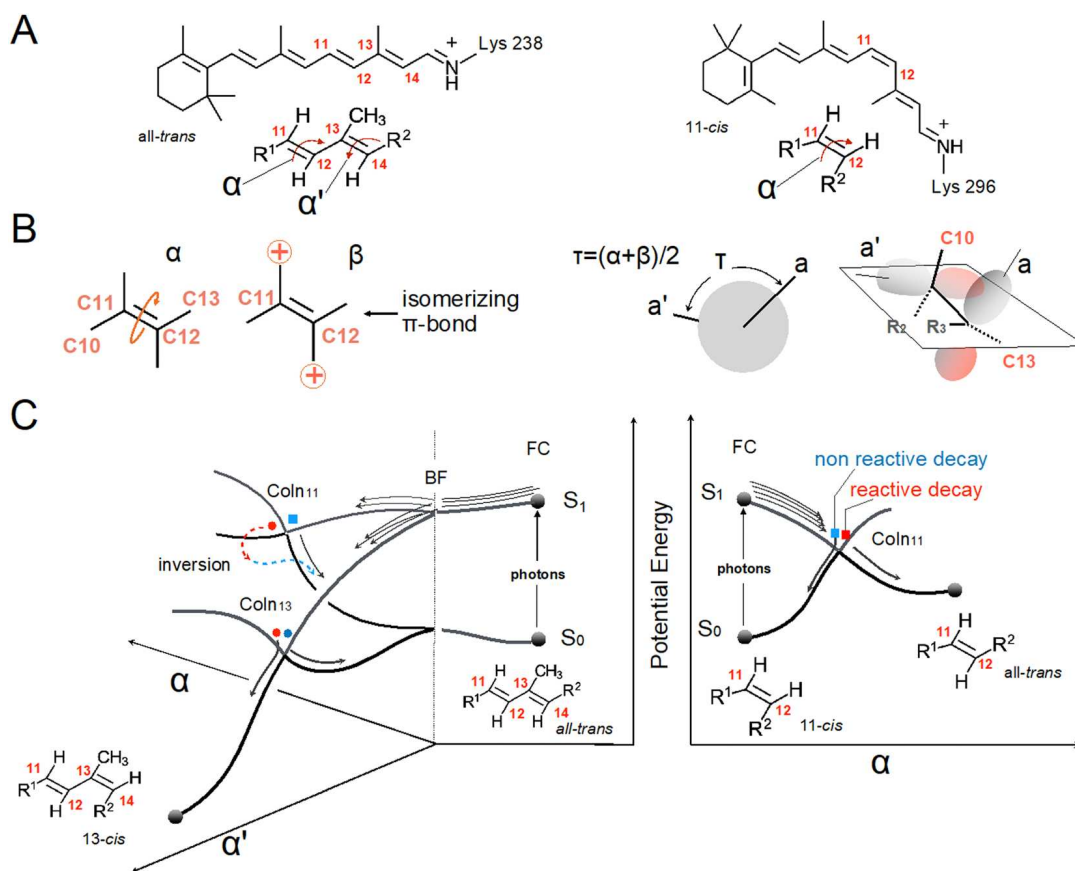


Fig. 1 S_1 isomerization channels in TaHeR and Rh. (A) Retinal chromophore in TaHeR (left) and Rh (right). The moiety at the bottom represents the reactive part of the two chromophores. The curly arrows indicate the S_1 reaction coordinates of interest and α and α' represents the rotation (or torsional or twisting) about C11=C12 and C13=C14 respectively. (B) Schematic representation of α , β and τ . The same representations apply to α' and β' and the corresponding τ' parameter. τ and τ' are geometrical parameters proportional to the π -overlap (roughly, the dihedral between the axes a and a' of the p -orbitals forming the π -bond) of the corresponding double bond. (C) Schematic representation of the photoisomerization channels in TaHeR (left) and Rh (right). The arrows in the FC region indicate the S_1 dynamics that, in TaHeR but not in Rh, leads to an ultrafast breaking (at BF) of the photoexcited population in two subpopulations reaching the regioselective Coln₁₁ and Coln₁₃ decay channels.

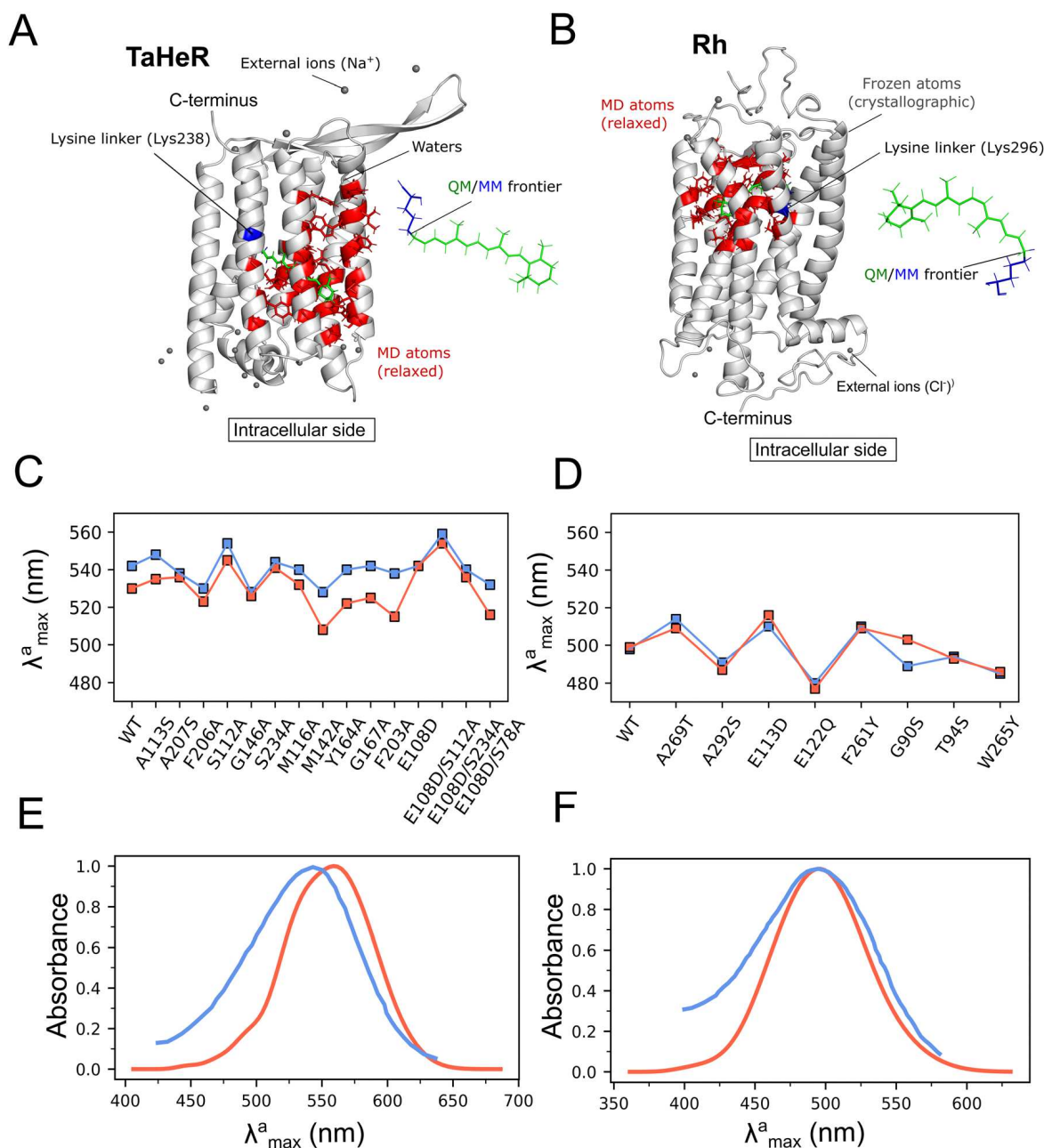


Fig. 2 TaHeR and Rh models validation. Overview of the QM/MM models for (A) TaHeR and (B) Rh. For each case, the protein environment fixed at the crystallographic position is colored in gray while the flexible binding cavity is in red. The lysine linker and rPSB chromophore are colored in blue and green respectively reflecting the QM/MM partition. Dark and light gray spheres indicate Na⁺ and Cl⁻ external counterions, respectively. The intracellular side of the protein is marked to show how the C-terminus of TaHeR is oriented oppositely with respect to Rh. (C) Comparison between computed (in orange) and observed (in blue) absorption maxima wavelengths λ^a_{\max} , calculated as vertical excitation energies for a set of TaHeR mutants. (D) Same as point C for Rh mutants reported in ref. 42. (E) Comparison between computed (in orange) and observed (in blue) absorption bands (in arbitrary units) for TaHeR. (F) Same as point E for Rh. The experimental bands were digitalized from literature data.^{34,47}

to reproduce the trend in the observed λ^a_{\max} displayed by a set of single mutants of Rh,⁴¹ and a set of single and double mutants of TaHeR (in most cases with a blue-shifted error typical of α -ARM models^{10,11,41,43–45}). The glutamic-to-aspartic acid mutation at the 108 counterion position of TaHeR yielded identical λ^a_{\max} value with respect to the wild-type, again in line with the observation.⁴⁶ Also, it is worthwhile mentioning that in our TaHeR model the protonation state of the two peculiar His23

and His82 residues is consistent with the proposed proton release pathway (see Section S1.1, ESI[†]). The agreement between computed and observed spectral data supports the use of the QM/MM models for simulating the corresponding room-temperature Boltzmann distributions and calculation of the absorption band. The computed absorption bands (see Fig. 2E and F) confirm the quality of the distribution represented by 200 geometries and velocities.

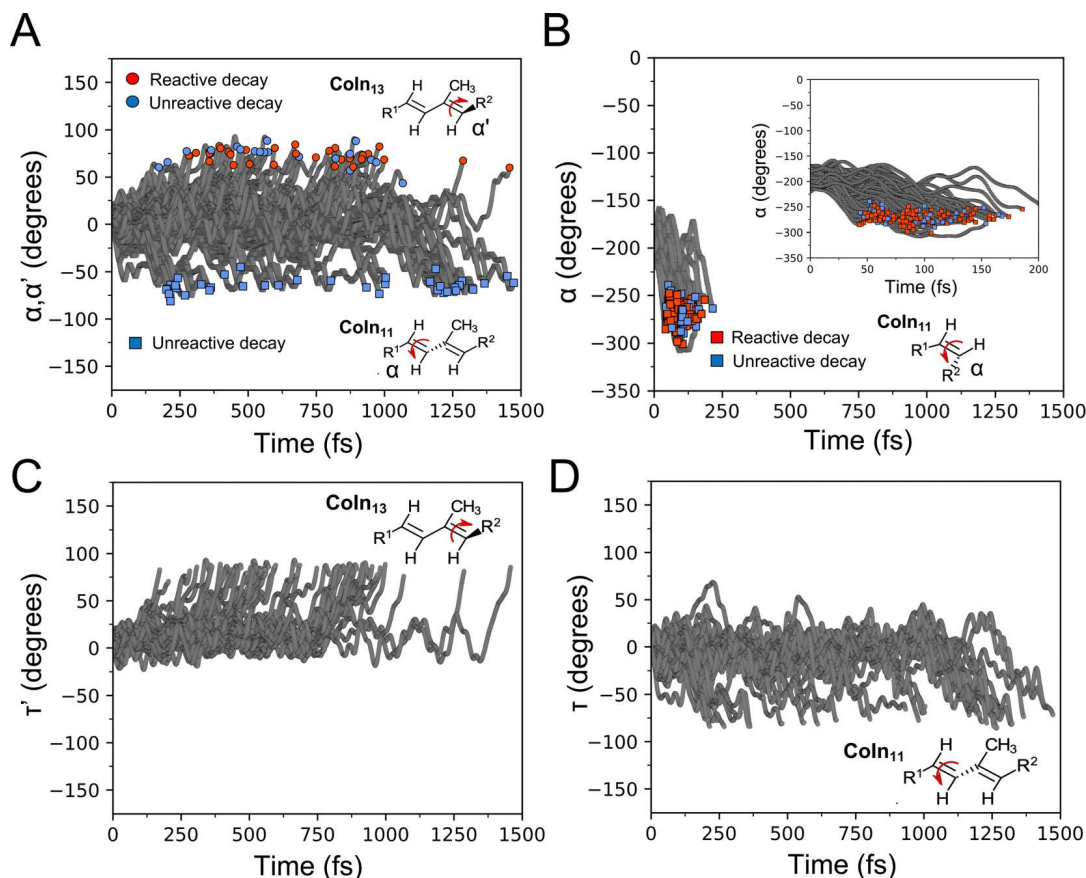


Fig. 3 Overview of the S_1 dynamics of (A) TaHeR and (B) Rh, simulated with 200 quantum-classical trajectories. Unreactive S_1 trajectories are marked by blue symbols placed at the time of the $S_1 \rightarrow S_0$ decay (hopping time). Similarly, reactive trajectories are marked by red symbols. We termed as reactive the trajectory that, after the hopping, continue to propagate on S_0 in the direction of the photoproduct. In contrast, unreactive trajectories revert their direction and propagate in the direction of the reactant (see also red and blue markers in Fig. 1C). In A, both α and α' coordinates (see Fig. 1A) are shown in order to illustrate the two coexisting isomerizations, however, a separate view is given in Fig. S5 (ESI \dagger). (C) Evolution of the overlap index τ' along the TaHeR trajectories undergoing C13=C14 isomerization (*i.e.* channel CoIn $_{13}$) until the $S_1 \rightarrow S_0$ decay point. (D) Same as C but for the index τ associated to C11=C12 isomerization (*i.e.* channel CoIn $_{11}$).

B. Excited state dynamics and TAS assignment

The datasets representing the Boltzmann distribution of TaHeR and Rh (see above) are used as initial conditions for the propagation of 200 quantum-classical trajectories (see methods and Section S2 for details, ESI \dagger). The room temperature geometries reveal a pre-twisting of the rPSB chromophore that is known⁷ to bias the S_1 motion in the CW or CCW directions. In fact, the average structure obtained from the initial conditions shows, in TaHeR, an average +9° CW pre-twisting of the C13=C14 double bond while, in Rh, an average CCW twisting of -13° of the C11=C12 double bond (these values are also consistent with those seen in the static α -ARM models of Fig. S3, ESI \dagger).

The extreme difference in amino acid composition between Rh and TaHeR cavities (see Fig. S1, ESI \dagger) is reflected in a qualitative, but drastic, different S_1 dynamics. This appears ballistic and substantially coherent in Rh but slower and more complex in TaHeR. This is shown in Fig. 3A and B where the time evolution of α and α' is displayed for each trajectory. By comparing the two figures, it becomes apparent that while the S_1 decay of TaHeR is not completed after 1.5 ps (only 93 out of 200 hopped to S_0), Rh exhibits a full S_1 depopulation in less

than 0.2 ps (see also the different population decay profiles in Fig. S4, ESI \dagger). This difference is already evident at <0.05 ps timescales, when most Rh trajectories exhibit a large CCW C11=C12 twist (*ca.* -70° on average) while no TaHeR double bond displays a twist larger than $\pm 20^\circ$. After this initial phase, all 200 Rh trajectories decay to S_0 at a single CoIn $_{11}$ channel and reconstitute the C11=C12 π -bond in either the CW or the CCW directions. This does not occur in TaHeR where both CoIn $_{11}$ and CoIn $_{13}$ channels, featuring a distinct regiochemistry and stereochemistry, are populated and coexist for the entire simulation time. CoIn $_{13}$ (see red and blue circles in Fig. 3A respectively) is dominated by the CW twisting of the C13=C14 double bond. CoIn $_{11}$ (see blue squares in Fig. 3A), on the other hand, is dominated by the CCW twisting of the C11=C12 double bond, in full analogy with Rh. Notably, the two TaHeR competing channels yield nearly identical decay counts (52 and 41 decays, respectively, after 1.5 ps of simulation time).

With the purpose of further characterizing the population of the TaHeR channels, we computed the S_1 minimum energy paths (MEPs) connecting the Franck-Condon (FC) point to the minimum energy conical intersections associated to the CoIn $_{11}$

and CoIn_{13} paths (see Section S3 for details, ESI[†]). The results show that the two channels are achieved through reaction coordinates spanning very flat S_1 PES regions from -180° to -220° along α and from -170° to -130° along α' . These MEPs are very different from the steep MEP reported for Rh⁴⁸ but consistent with the longer S_1 dynamics experimentally observed for TaHeR.³⁵ The analysis of the overlap coordinates τ and τ' in TaHeR indicates a $\text{C13}=\text{C14}$ and $\text{C11}=\text{C12}$ bond breaking process along a torsional or, as documented in the next section, bicycle-pedal coordinate that, in both cases, leads initially to an oscillatory $\text{C}=\text{C}$ torsional motion in the -35° to $+35^\circ$ range then followed by a decay event (Fig. 3C and D). In other words, all TaHeR trajectories display a nearly planar (*i.e.* with a nearly 0° average torsion) π -system that, eventually, escapes such an oscillatory region along the two isomerization paths that are randomly

accessed during the simulation time. Such random behaviour is further strengthened by the absence of any discernible correlation between early (< 30 fs) geometrical motions and the two isomerization directions (see Fig. S7–S9, ESI[†]). Therefore, the S_1 dynamics exhibited in the FC region (see Fig. 1C) seem to have a null or little impact on determining the isomerization direction.

In the past, flat S_1 PES have been proposed to be responsible for longer S_1 lifetimes of microbial rhodopsins with respect to Rh.^{49,50} An alternative reason for the slow S_1 dynamics has been reported to be the mixing of the S_1 and S_2 states.^{9,51} Indeed, compared to S_1 which is usually associated with a charge transfer state, S_2 is associated with a diradical state featuring a shorter and semi-locked $\text{C}=\text{C}$ double bonds. Thus, the mixing would be partly responsible for the increased excited state lifetime since isomerizations along these bonds became

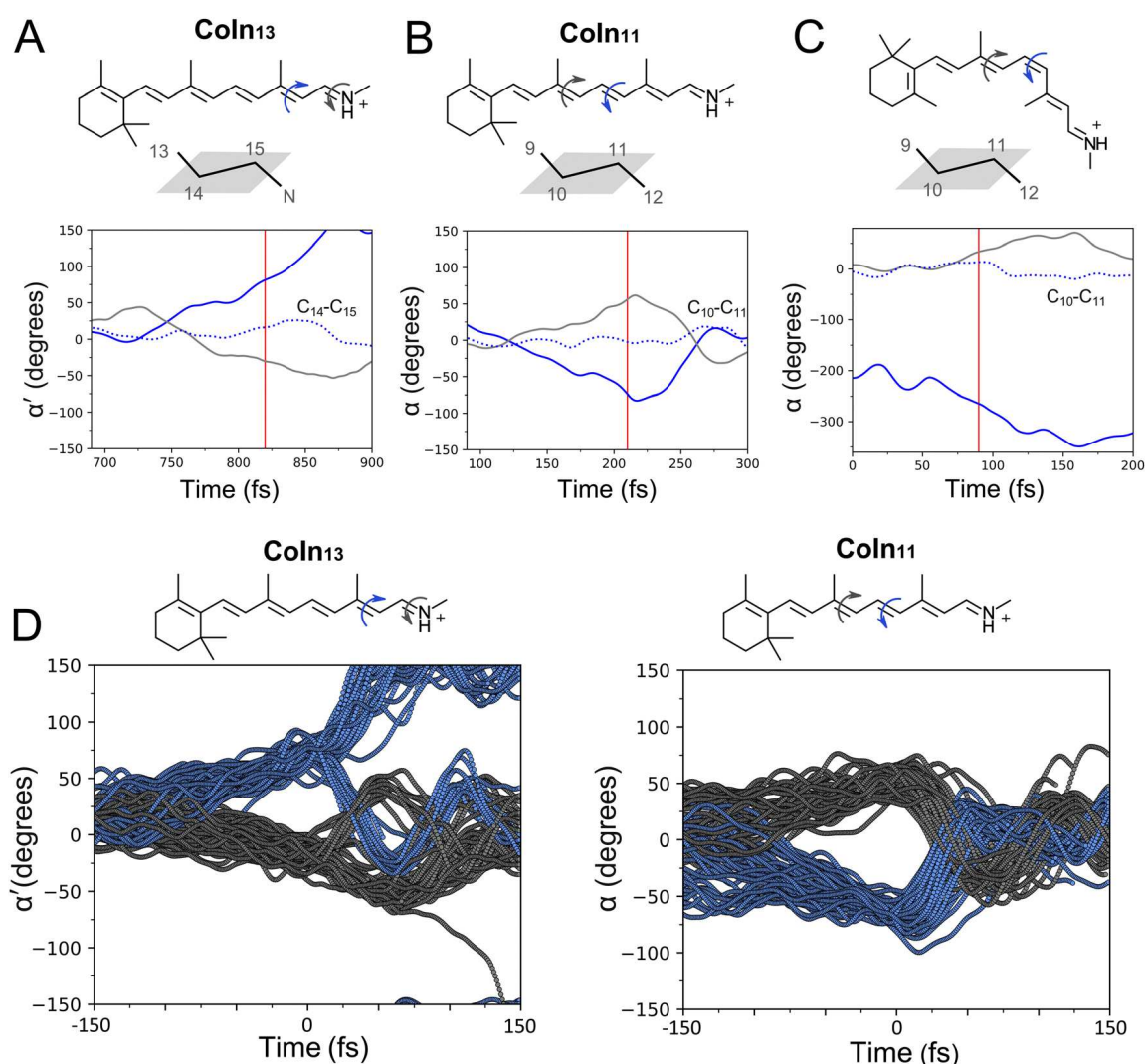


Fig. 4 Time evolution of the ABP photoisomerization coordinates of TaHeR and Rh. (A) A representative reactive trajectory achieving the Coln_{13} channel of TaHeR (coordinate α' in Fig. 1), (B) a representative trajectory (unreactive) entering the Coln_{11} channel of TaHeR (coordinate α in Fig. 1). (C) A reactive Rh trajectory. The main torsional deformations (see blue and gray curly arrows) are represented in terms of rotating planes as illustrated at the top. In all cases, the diagrams reported on the geometrical changes consistent with the bicycle-pedal motion and rotating planes. The progression of the $\text{C13}=\text{C14}-\text{C15}=\text{N}$ and $\text{C9}=\text{C10}-\text{C11}=\text{C12}$ dihedral angles is displayed as dotted lines. The vertical red lines correspond to $S_1 \rightarrow S_0$ decay events. (D) Same representation of above but including instead all the trajectories achieving the Coln_{13} (left panel) or the Coln_{11} (right panel) channels of TaHeR. The trajectories were aligned at the $S_1 \rightarrow S_0$ decay time ($t = 0$ fs). It is shown that while the Coln_{13} is partially reactive, the Coln_{11} is fully unreactive.

more impaired. As documented in the ESI† (see Section S4), the same level of S_2/S_1 mixing is detected in TaHeR trajectories leading to either the CoIn_{13} or CoIn_{11} channels and also the trajectories that do not decayed to S_0 (see Fig. S10–S12, ESI†).

The computed timescales are qualitatively consistent with the observed TAS dynamics for both TaHeR³⁵ and Rh⁵² thus providing a further validation of the constructed QM/MM models. For TaHeR, TAS has shown that the $S_1 \rightarrow S_0$ decay is completed after several ps (>8.5 ps) and exhibits a multi-exponential decay including fast (<1 ps) and slow (>1 ps) components. More specifically, the presence of those components has been drawn from the time evolution of the stimulated emission (SE), ground state bleaching (GSB) and excited state absorption (ESA) temporal profiles, with the latter shown to be a valid and reliable indicator of the S_1 depopulation.⁵³ However, these observations were interpreted by assuming that the C13=C14 rotation is the sole excited state photoisomerization process. Therefore, the fast as well as slower components were associated to a multi-exponential process where the sub-picosecond formation of the 13-*cis* photoproduct begins *ca.* 0.2 ps after photoexcitation, while slower decay components were attributed to “unreactive processes” generating less or no photoproduct. Our TaHeR simulation indicates that the observed fast and slower decay components are the result of the coexistence of the two S_1 isomerization channels described above. More specifically, our calculations associate the observed fast and slow decays of TaHeR with the presence of S_1 isomerizations with distinct dynamics, regioselectivity, stereoselectivity and efficiencies. Indeed, although in our trajectories (Fig. 3A) earlier decays are observed approximately after 0.2 ps, the majority of CoIn_{13} decays take place within 1 ps (only 3 out of 52 trajectories decayed after 1 ps), while along CoIn_{11} the trajectories decay more sparsely on a longer time-scale. In striking contrast, the simulated Rh dynamics that fully complies with TAS measurements,^{52,54} shows that the rPSB photoisomerization occurs, substantially, through a single mono-exponential decay (we disregard the very inefficient formation of isorhodopsin featuring a C9=C10 rotation⁵⁵). Thus, in TaHeR the low $\phi_{\text{iso},13}$ value would be the results of lack of isomerization regioselectivity during S_1 progression combined with an unreactive animal-like C11=C12 decay at CoIn_{11} . In fact, the 41 trajectories decaying through the CoIn_{11} channel, lead to reconstitution of the original all-*trans* rPSB exclusively (see blue squares, Fig. 3A). Conversely, the CoIn_{13} channel is highly “reactive”, as out of the 52 trajectories evolving along this path, 33 lead to photoproduct formation (see red circles, Fig. 3A) with 13-*cis*, 15-*trans* (32 trajectories) or 13-*cis*, 15-*cis* configurations (1 trajectory), as observed experimentally,³⁴ and the remaining 19 relax back to the all-*trans* rPSB configuration (see blue circles, Fig. 3A). Based on the assumption that the trajectories left on S_1 never decay productively (see above), we computed a 0.17 $\phi_{\text{iso},13}$ value (33 reactive trajectories over 200) of TaHeR, which is close to the experimentally observed ϕ_{iso} value of 0.19. Concerning Rh, all the trajectories displayed $S_1 \rightarrow S_0$ decays within *ca.* 0.1 ps (first 49 fs last 218 fs). The predicted ϕ_{iso} value is of 0.70, again in line with the observed 0.67.³⁸

Table 1 Summary of predicted (Pred.), computed (Comp.) and experimental (Exp.) ϕ_{iso} . Number of reactive trajectories is also shown. Values in parenthesis indicate the percentage of trajectories over the total number of trajectories decaying along each decay channel. Pred. ϕ_{iso} is calculated using the $d\tau/dt^{\text{decay}}$ sign (or $d\tau'/dt^{\text{decay}}$) as the index for predicting if a trajectory is reactive or unreactive (refer to main text for definition). Comp. ϕ_{iso} is calculated over 200 trajectories. It is important to note that, in TaHeR, only 93 trajectories decayed to S_0 (along CoIn_{11} or CoIn_{13}) and we consider the trajectories remained in S_1 as unreactive. N.D. refers to “not detected”

	Reactive	$d\tau/dt^{\text{decay}}$ or $d\tau'/dt^{\text{decay}}$ ^a	Pred. ϕ_{iso}	Comp. ϕ_{iso}	Exp. ϕ_{iso}
Rh, CoIn_{11}	139 (70%)	148 (74%)	0.74	0.70	0.67 ³⁸
TaHeR, CoIn_{13}	33 (63%)	39 (74%)	0.75	0.17	0.19 ³⁵
TaHeR, CoIn_{11}	0 (0%)	20 (49%)	0.49	0	N.D.

^a Number of Rh and TaHeR CoIn_{11} trajectories with $d\tau/dt^{\text{decay}} < 0$ and, for TaHeR CoIn_{13} , with $d\tau'/dt^{\text{decay}} > 0$.

C. Counterclockwise vs. clockwise rotation and aborted bicycle-pedal

Fig. 4 displays the time evolution of critical dihedral angles. In TaHeR the CW isomerization of the C13=C14 bond is strongly coupled with the CCW rotation of the C15=N bond leading to the CoIn_{13} channel (see Fig. 4A). In fact, the decay occurs when the dihedral angle at C13=C14 and C15=N bonds are $+80^\circ$ and -30° , respectively. Notice that the dihedral angle associated with the middle C14-C15 single bond is minimally changed. The same description can be applied to the C11=C12 isomerization that is coupled to an opposite C9=C10 motion (see Fig. 4B where, at decay, the dihedrals associated to C9=C10 and C11=C12 feature a twist of $+55^\circ$ and -73° , respectively). The described motions of the C13=C14-C15=N and C9=C10-C11=C12 moieties is consistent with the bicycle pedal (BP) isomerization mechanism. The BP mechanism has been originally proposed as “space saving” motion by Warshel for Rh.⁵⁶ More recently, it has been predicted that such a Rh mechanism is instead *aborted*⁴⁸ as one of the two adjacent double bonds of the C9=C10-C11=C12 moiety inverts its rotation direction at decay as here replicated in Fig. 4C. This aborted bicycle pedal (ABP) mechanism has also been found in microbial rhodopsins such as bacteriorhodopsin⁴⁹ and *Anabaena* Sensory rhodopsin⁵⁰ and similarly, in animal rhodopsins where the 11-*cis* C11=C12 bond isomerizes.^{55,57,58} Recently, the existence of computationally predicted ABP has been supported *via* time-resolved XFEL studies on Rh⁵⁹ and bacteriorhodopsin.⁶⁰

Our results support the conservation of the ABP mechanism in the HeR family. This is clearly shown at the population level in Fig. 4D, left diagram, where the CCW motion around C15=N reverts its rotation at decay to S_0 , then producing the isomerization at the C13=C14 exclusively. Notice, however, the presence of one trajectory that exhibits a full BP mechanism yielding the 13-*cis*, 15-*cis* photoproduct. The right diagram is related to the CoIn_{11} channel and demonstrates that, in contrast with CoIn_{13} , both C9=C10 and C11=C12 rotations invert directions immediately after decay to S_0 such that no net isomerization takes place.

D. Regioselectivity and quantum efficiency modulation in TaHeR

In this section we look at the mechanism responsible for the experimentally observed C13=C14 selectivity (*i.e.* only the

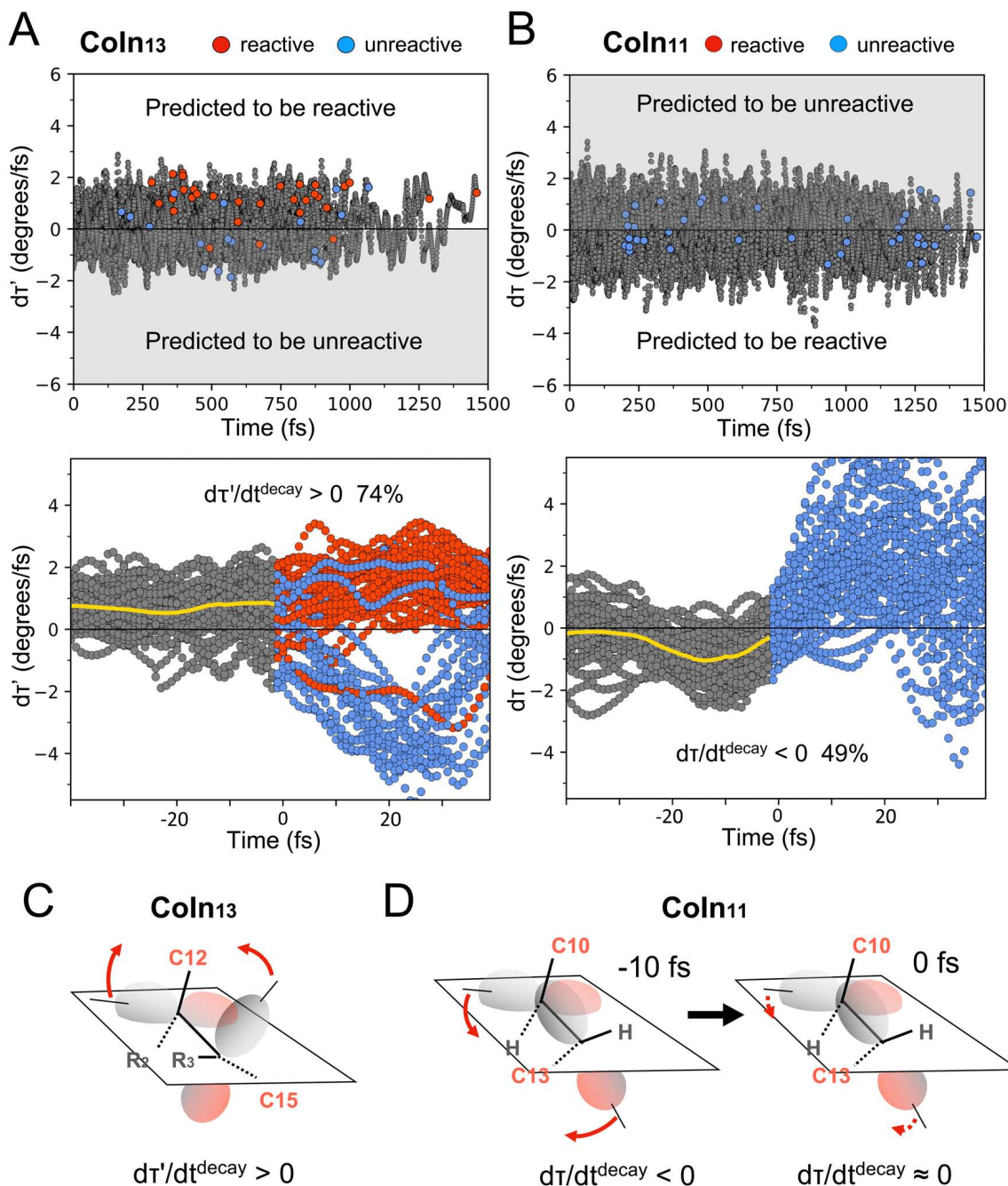


Fig. 5 Reactivity analysis of TaHeR trajectory sets. (A) Time evolution of the overlap index $d\tau'/dt$ towards the Coln₁₃ channel. The top panel shows the correlation between the distribution of reactive decays (red circles) and the $d\tau'/dt$ sign. The bottom panel displays the same data but with trajectories aligned at the $S_1 \rightarrow S_0$ decay time ($t = 0$ fs) and starting 40 fs before the decay. Gray data refers to the S_1 while red and blue data refers to S_0 depending on whether the trajectory is reactive (red) or unreactive (blue). The yellow curve represents the average value before the $S_1 \rightarrow S_0$ decay. The displayed numerical value refers to the percentage of trajectories decaying with $d\tau'/dt > 0$ (predicted quantum efficiency, see Table 1). To notice the behaviour of a few reactive and unreactive trajectories that do not correlate with, respectively, the $d\tau'/dt > 0$ and $d\tau'/dt < 0$ sign at the decay time. In these cases, $d\tau'/dt$ inverts the sign after the decay point. These trajectories can be seen as those that deviate from the average behaviour but still falling in the expected statistics provided by the τ velocity (see main text). (B) Same as A for the CCW motion leading to the Coln₁₁ channel. Notice that now the displayed numerical value refers to $d\tau'/dt < 0$. (C) Pictorial illustration of the τ' velocity in terms of decreasing overlap along the CW progression towards the Coln₁₃ channel. The curly arrow represents the rotational velocity of the orbital axes. (D) Same as C for the Coln₁₁ channel where, in contrast, the velocity of the CCW overlap decrease is decelerating during the progression towards the decay channel.

C13=C14 isomerization is detected) and low $\varphi_{iso,13}$ value of TaHeR in the presence of a non-regioselective S_1 isomerization dynamics. Past and recent studies on Rh^{8,9} have demonstrated the pivotal role of the geometrical parameter τ for the

mechanistic interpretation of reaction quantum efficiencies in Rh. More specifically, at decay one looks at the sign of the τ velocity ($d\tau/dt^{decay}$). Indeed, it has been established that in the CCW isomerization the $d\tau/dt^{decay} < 0$ provides a statistically

valid prediction of the corresponding computed $\phi_{\text{iso},11}$ value. When considering the stereochemical change in rotation direction, in a CW isomerization such a prediction would be based in the fraction of trajectories decaying with $d\tau/dt^{\text{decay}} > 0$. These “rules” are connected and justified by the canonical Landau–Zener (LZ) model for a reactive non-adiabatic transition.⁶¹ In fact, the LZ model relates the probability of non-adiabatic transitions leading to the product (*i.e.* the probability that after S_1 decay the system would generate a product) to the nuclear velocity along the “local” reaction coordinate. A comparison of Fig. 5A and B demonstrates that such sign rule is valid for the C13=C14 isomerization (*i.e.* for decay through the CoIn_{13} channel) of TaHeR but it is not valid for the C11=C12 isomerization (*i.e.* for decay through the CoIn_{11} channel) in deep contrast with the reported validity in the case of the CoIn_{11} channel of Rh. More specifically, in the top panel of Fig. 5A it can be seen that the vast majority of reactive trajectories (red circles) decay with $d\tau'/dt^{\text{decay}} > 0$ (upper half of the panel). In deep contrast, Fig. 5B shows that irrespective of the sign of $d\tau/dt^{\text{decay}}$ the trajectories are all unreactive.

The lower panels of Fig. 5A and B report an analysis of the $d\tau/dt$ in the vicinity of the TaHeR decay regions. By focusing on a time window of 40 fs centered on the decay event, it is shown

that τ velocities behave differently along the paths leading to the two distinct channels. In the case of trajectories pointing to the CoIn_{13} channel, the population displays a positive average velocity that appears to remain constant until the decay region is reached. This would translate into a constant velocity for the breaking of the π -bond as pictorially illustrated in Fig. 5C. In contrast, consistently with the illustration of Fig. 5D, the corresponding motion towards CoIn_{11} displays decreasing τ velocities suggesting formation of the 11-*cis* product. However, about 10 fs before decay, a deceleration occurs, and the trajectories reach the decay region with rapidly decreasing velocities and yielding only small negative values at decay.

Fig. 5A and B also provide information on the TaHeR π -bond making dynamics immediately after the decay. It can be seen that, at CoIn_{13} , the percentage of reactive trajectories generated at decay (63%) is relatively close to that of $d\tau'/dt^{\text{decay}} > 0$ (74%). This does not happen in the case of the unreactive CoIn_{11} channel (0%) where, despite a still large (49%) fraction of $d\tau/dt^{\text{decay}} < 0$, $d\tau/dt$ displays, immediately after the decay, large and positive values consistent with the resulting reactant reconstitution. This behavior suggests that in correspondence of the $d\tau/dt$ decrease seen about 10 fs before the hopping points, the C11=C12 rotation or, better, the bicycle pedal

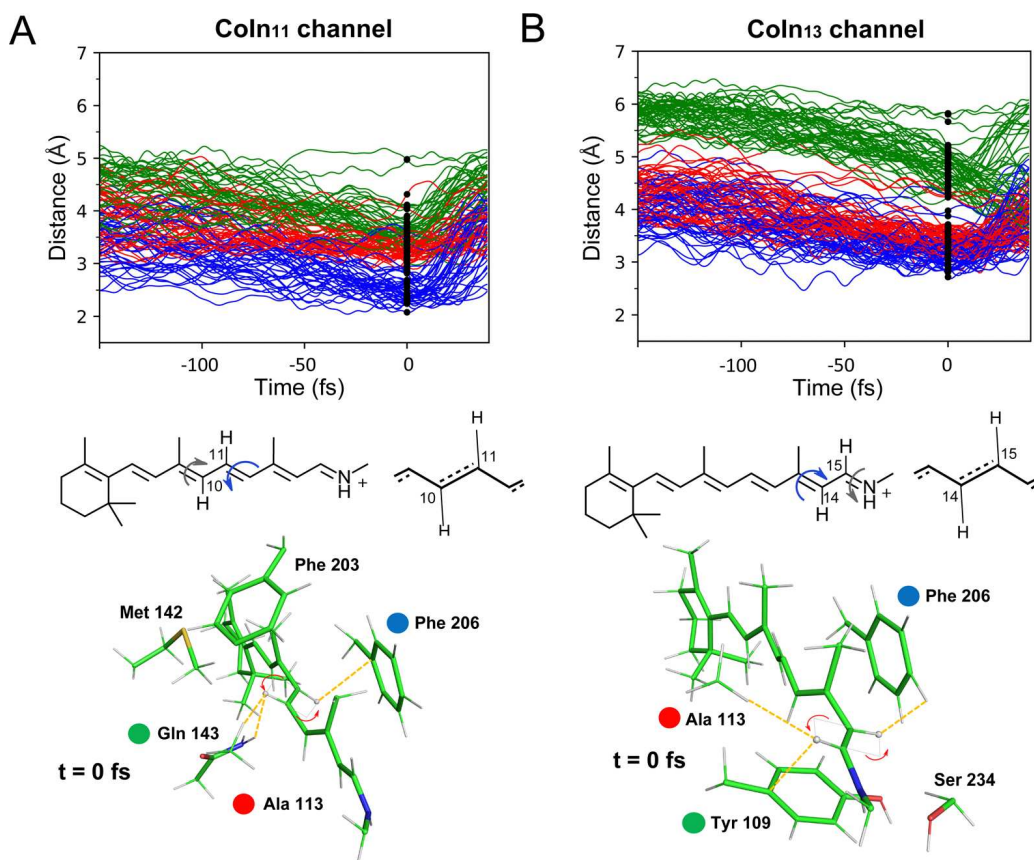


Fig. 6 Quantum efficiency quenching mechanism in TaHeR. (A) Population time evolution (trajectories aligned at the $S_1 \rightarrow S_0$ decay time) of the shortest distances between the rPSB chromophore atoms characterizing the CW rotation of the bicycle-pedal plane of the CoIn_{11} channel with neighboring amino acids indicated below. The vertical set of black dots indicate the distribution of distances at the hopping points. Below, the definition and color coding of the distances monitored in the top. The rotating bicycle pedal plane is also defined pictorially and within a snapshot of the reacting rPSB region at the $S_1 \rightarrow S_0$ decay time ($t = 0$ fs). (B) Same as A for the CoIn_{13} channel.

nuclear motion encounters a restraining force effectively inverting the motion from CCW to CW.

In order to provide a molecular-level understanding of the lack of reactivity of the CoIn₁₁ channel of TaHeR, we hypothesize that the restraining force mentioned above originate from steric and/or electrostatic interactions. Here the target is to provide a guidance to experimentalists that could eventually search for mutants capable of producing the 11-*cis* rPSB isomer. We start with the observation that the TaHeR cavity (see Fig. 6) accommodates the reactive C9=C10-C11=C12 moiety in the close vicinity of the Phe203, Phe206 and Tyr109 residues as well as of the polar Gln143 residue (Gln143 is absent in the cavities of most known microbial rhodopsins).³⁴ Inspection of the TaHeR trajectories reveals that the C11=C12 rotation leads to strongly decreased distances between the plane associated with the BP coordinate and the listed residue sidechains (see Fig. 6A). This effect appears to lead to the “collision” of the rotation of the substituents at C10 and C11 with the π -system of Phe206 and the -NH₂ aminic group of Gln143 causing competition of VdW radii⁶² and, thus, strong repulsion. The same distance decrease may lead to electrostatic repulsion between the positive charged nitrogen of Gln143 amidic group and the chromophore positive charge that becomes increasingly localized on the -C9-C10-C11- moiety along the S₁ isomerization path (see last 50 fs of charge evolution in Fig. S10, ESI[†]). As anticipated above, these steric and electrostatic effects may systematically invert the negative sign of $d\tau/dt$ before and after the decay pointing to π -bond reconstitution *via* CW, rather than CCW, rotation. On the other hand, the rotation of the C13=C14-C15=N BP plane is less impaired (compare Fig. 6A with B), as demonstrated by the >2.5 Å distance kept by the C14 and C15 substituents with the closest residue sidechains during the progression toward the CoIn₁₃ channel.

Finally, as shown in Fig. S14 (ESI[†]), further analysis has been carried out to disentangle the importance of the steric *vs.* electrostatic repulsive effects. The figure shows that after switching the charges of the protein cavity to zero values, the “repulsive” effect documented above is conserved suggesting a steric origin of the repulsion.

Conclusion

Above we have reported on the first atomistic description of the primary event in the photocycle of a member of the recently discovered HeR family. Since our TaHeR QM/MM model reproduces the observed spectroscopic parameters and measured ϕ_{iso} , we have used it to simulate the photoinduced dynamics and reinterpret the previously proposed mechanistic hypothesis.³⁵ The simulation reveals that TaHeR light absorption leads to an *equal population* of two regiochemically and stereochemically distinct S₁ isomerization channels. The first C13=C14 channel is found in substantially all microbial rhodopsins, while the second C11=C12 channel is related to the photoisomerization of animal rhodopsins but in the reversed all-*trans* to 11-*cis* direction. The coexistence of these two channels appears

connected with similar S₁ energy paths, which were found flat and substantially barrierless in both cases. These data may call TaHeR as promising candidate for isomerization selectivity control experiments. In fact, future studies may focus on the identification of variants that exhibit enhanced S₁ isomerization energy barriers along one of the two channels. This could ultimately lead to a different branching ratio compared to the original (wild-type) system. Since TaHeR is phylogenetically distant from both animal and microbial rhodopsins, the resulting promiscuous photoisomerization provides an example of a high isomerization diversity induced by diverse opsin cavity compositions. A second result is the finding of a new $\phi_{\text{iso},13}$ modulation mechanism in TaHeR, not based on a low $\phi_{\text{iso},13}$ but, rather, to the lack of S₁ regioselectivity with respect to an equally populated fully inefficient C11=C12 S₁ isomerization channel. The suppression of the C11=C12 channel would allow, in principle, an increase of ϕ_{iso} from 0.17 to 0.63. A third result is the confirmation of the universality, *i.e.* its validity also for a HeR member, of the space-saving ABP mechanism initially documented for Rh,⁴⁸ as well as of the validity of the theory based on τ velocity at the point of S₁ → S₀ decay.^{8,9}

Isomerization mechanisms featuring coexisting isomerization channels may operate in certain microbial rhodopsins. For instance, although the extreme distance between HeRs and microbial rhodopsins make any comparisons difficult, past and recent studies on chloride pump halorhodopsin,^{63,64} proteorhodopsin,^{23,65,66} sensory rhodopsin I (SRI),⁶⁷ channel-rhodopsin chimera C1C2,⁶⁸ sodium ion pump rhodopsin KR2⁶⁹ as well as bacteriorhodopsin under acidic conditions,²³ have investigated the factors that govern their observed multicomponent excited state decay. Similarly to TaHeR, these microbial rhodopsins display an S₁ dynamics well described by fast (subpicosecond) and reactive, and slow and unreactive decay components. Two types of mechanisms were proposed to rationalize the findings: (a) the potential heterogeneity of ground state species arising either from the acid-base equilibrium of the rPSB chromophore counterion or the presence of a mixture of chromophore isomers and (b) the possibility of reaction branching on S₁. Our results support the latter mechanism in the case of TaHeR. In fact, for TaHeR, the presence of different chromophore isomers in the dark-adapted state can be excluded on the basis of HPLC measurements that document a dominant presence of all-*trans* rPSB. Furthermore, the TAS measurements are carried out at neutral pH that should ensure the deprotonation of the rPSB counterion (Glu108).³⁵

A branched isomerization mechanism has been experimentally established for a microbial rhodopsin from *Haloquadratum walsbyi* called middle rhodopsin (MR)⁷⁰ where a mixture of 13-*cis* and 11-*cis* configurations is produced upon illumination starting from all-*trans* rPSB. It is possible that the isomerization mechanism of MR is essentially the same described above for TaHeR but having the all-*trans* to 11-*cis* isomerization channel corresponding to CoIn₁₁, at least partially reactive. The authors of such research propose that MR might represent “a missing link” in the evolution from microbial to animal rhodopsins. It is apparent that the presented TaHeR investigation indicates

that such hypothesis could also apply to HeRs. More specifically, we envision that a mutation of the Gln143 causing a reduction in steric hindrance and/or reverting the dipole moment of this residue (*i.e.* pointing the positive $-NH$ side of the dipole away from the isomerizing moiety) could make the two CoIn channels reactive providing a first example of HeRs variant capable of producing the 11-*cis* chromophore isomer typical of animal rhodopsins. Even more recently, the coexistence and interplay of C11=C12 and C13=C14 photoisomerizations was experimentally shown to be active in the rhodopsin tandem domains of the marine algae *Phaeocystis antarctica* bestrhodopsin.⁷¹

Methods

The computational models employed for bovine rhodopsin and thermoplasmatales archaeon rhodopsin were constructed with the α -ARM QM/MM model building protocol,^{40,41} implemented in the PyARM software package,^{11,42} starting from their crystallographic structures resolved at 2.4 Å (Protein Data Bank code: 1U19)⁷² and 2.2 Å (Protein Data Bank code: 6IS6),³⁴ respectively. Such models were used to simulate two room temperature Boltzmann-like distributions and extract 200 initial conditions for each rhodopsin at room temperature following a previously reported protocol.⁹ Finally, non-adiabatic population dynamics were computed at 2 root (S_0 and S_1) state average CASSCF(12,12)/6-31G*/Amber94 level of theory using the stochastic Tully surface-hop method.^{39,73} While the CASSCF method does not inherently incorporate dynamic correlation energy, benchmark studies^{74–77} have demonstrated that the impact of the CASPT2 correction on these systems primarily affects the timescale rather than the geometrical propagation described by the trajectory itself. Therefore, we are confident that the mechanistic theory presented in this work would be replicated, at least qualitatively, in a study based on trajectories computed at the CASPT2 level. All the calculation were run using the Molcas version 8.1⁷⁸/Tinker⁷⁹ interface. Further details of the α -ARM and initial condition protocols are provided in ESI† (Sections S1 and S2).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

M. O. is grateful to the NSF CHE-SDM A for Grant No. 2102619. R. P., L. B., X. Y. and M. O. are also grateful for partial support provided by EU funding within the MUR PNRR “National Center for Gene Therapy and Drugs based on RNA Technology” (Project no. CN00000041 CN3 RNA) - Spoke 6.

References

1 O. P. Ernst, D. T. Lodowski, M. Elstner, P. Hegemann, L. S. Brown and H. Kandori, *Microbial and Animal*

- Rhodopsins: Structures, Functions, and Molecular Mechanisms, *Chem. Rev.*, 2014, **114**(1), 126–163, DOI: [10.1021/cr4003769](https://doi.org/10.1021/cr4003769).
- 2 H. Kandori, *Retinal Proteins: Photochemistry and Optogenetics*, *Bull. Chem. Soc. Jpn.*, 2020, **93**(1), 76–85, DOI: [10.1246/bcsj.20190292](https://doi.org/10.1246/bcsj.20190292).
- 3 K.-H. Jung, V. D. Trivedi and J. L. Spudich, Demonstration of a Sensory Rhodopsin in Eubacteria, *Mol. Microbiol.*, 2003, **47**(6), 1513–1522, DOI: [10.1046/j.1365-2958.2003.03395.x](https://doi.org/10.1046/j.1365-2958.2003.03395.x).
- 4 G. Nagel, T. Szellas, W. Huhn, S. Kateriya, N. Adeishvili, P. Berthold, D. Ollig, P. Hegemann and E. Bamberg, Channelrhodopsin-2, a Directly Light-Gated Cation-Selective Membrane Channel, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**(24), 13940–13945, DOI: [10.1073/pnas.1936192100](https://doi.org/10.1073/pnas.1936192100).
- 5 E. G. Govorunova, O. A. Sineshchekov, H. Li and J. L. Spudich, Microbial Rhodopsins: Diversity, Mechanisms, and Optogenetic Applications, *Annu. Rev. Biochem.*, 2017, **86**(1), 845–872, DOI: [10.1146/annurev-biochem-101910-144233](https://doi.org/10.1146/annurev-biochem-101910-144233).
- 6 A. Rozenberg, K. Inoue, H. Kandori and O. Bèjà, Microbial Rhodopsins: The Last Two Decades, *Annu. Rev. Microbiol.*, 2021, **75**(1), 427–447, DOI: [10.1146/annurev-micro-031721-020452](https://doi.org/10.1146/annurev-micro-031721-020452).
- 7 S. Gozem, H. L. Luk, I. Schapiro and M. Olivucci, Theory and Simulation of the Ultrafast Double-Bond Isomerization of Biological Chromophores, *Chem. Rev.*, 2017, **117**(22), 13502–13565, DOI: [10.1021/acs.chemrev.7b00177](https://doi.org/10.1021/acs.chemrev.7b00177).
- 8 C. Schnedermann, X. Yang, M. Liebel, K. M. Spillane, J. Lugtenburg, I. Fernández, A. Valentini, I. Schapiro, M. Olivucci, P. Kukura and R. A. Mathies, Evidence for a Vibrational Phase-Dependent Isotope Effect on the Photochemistry of Vision, *Nat. Chem.*, 2018, **10**(4), 449–455, DOI: [10.1038/s41557-018-0014-y](https://doi.org/10.1038/s41557-018-0014-y).
- 9 X. Yang, M. Manathunga, S. Gozem, J. Léonard, T. Andruniów and M. Olivucci, Quantum–Classical Simulations of Rhodopsin Reveal Excited-State Population Splitting and Its Effects on Quantum Efficiency, *Nat. Chem.*, 2022, **14**(4), 441–449, DOI: [10.1038/s41557-022-00892-6](https://doi.org/10.1038/s41557-022-00892-6).
- 10 L. Barneschi, E. Marsili, L. Pedraza-González, D. Padula, L. De Vico, D. Kaliakin, A. Blanco-González, N. Ferré, M. Huix-Rotllant, M. Filatov and M. Olivucci, On the Fluorescence Enhancement of Arch Neuronal Optogenetic Reporters, *Nat. Commun.*, 2022, **13**(1), 6432, DOI: [10.1038/s41467-022-33993-4](https://doi.org/10.1038/s41467-022-33993-4).
- 11 L. Pedraza-González, L. Barneschi, M. Marszałek, D. Padula, L. De Vico and M. Olivucci, Automated QM/MM Screening of Rhodopsin Variants with Enhanced Fluorescence, *J. Chem. Theory Comput.*, 2023, **19**(1), 293–310, DOI: [10.1021/acs.jctc.2c00928](https://doi.org/10.1021/acs.jctc.2c00928).
- 12 E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel and K. Deisseroth, Millisecond-Timescale, Genetically Targeted Optical Control of Neural Activity, *Nat. Neurosci.*, 2005, **8**(9), 1263–1268, DOI: [10.1038/nn1525](https://doi.org/10.1038/nn1525).
- 13 F. Zhang, L.-P. Wang, M. Brauner, J. F. Liewald, K. Kay, N. Watzke, P. G. Wood, E. Bamberg, G. Nagel, A. Gottschalk and K. Deisseroth, Multimodal Fast Optical Interrogation of Neural Circuitry, *Nature*, 2007, **446**(7136), 633–639, DOI: [10.1038/nature05744](https://doi.org/10.1038/nature05744).

- 14 K. Deisseroth and P. Hegemann, The Form and Function of Channelrhodopsin, *Science*, 2017, **357**(6356), eaan5544, DOI: [10.1126/science.aan5544](https://doi.org/10.1126/science.aan5544).
- 15 P. A. Davison, W. Tu, J. Xu, S. Della Valle, I. P. Thompson, C. N. Hunter and W. E. Huang, Engineering a Rhodopsin-Based Photo-Electrosynthetic System in Bacteria for CO₂ Fixation, *ACS Synth. Biol.*, 2022, **11**(11), 3805–3816, DOI: [10.1021/acssynbio.2c00397](https://doi.org/10.1021/acssynbio.2c00397).
- 16 W. Tu, J. Xu, I. P. Thompson and W. E. Huang, Engineering Artificial Photosynthesis Based on Rhodopsin for CO₂ Fixation, *Nat. Commun.*, 2023, **14**(1), 8012, DOI: [10.1038/s41467-023-43524-4](https://doi.org/10.1038/s41467-023-43524-4).
- 17 O. Bèjà, 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 and E. F. DeLong, Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea, *Science*, 2000, **289**(5486), 1902–1906, DOI: [10.1126/science.289.5486.1902](https://doi.org/10.1126/science.289.5486.1902).
- 18 L. Gómez-Consarnau, J. A. Raven, N. M. Levine, L. S. Cutter, D. Wang, B. Seegers, J. Aristegui, J. A. Fuhrman, J. M. Gasol and S. A. Sañudo-Wilhelmy, Microbial Rhodopsins Are Major Contributors to the Solar Energy Captured in the Sea, *Sci. Adv.*, 2019, **5**(8), eaaw8855, DOI: [10.1126/sciadv.aaw8855](https://doi.org/10.1126/sciadv.aaw8855).
- 19 S. Inukai, K. Katayama, M. Koyanagi, A. Terakita and H. Kandori, Investigating the Mechanism of Photoisomerization in Jellyfish Rhodopsin with the Counterion at an Atypical Position, *J. Biol. Chem.*, 2023, **299**(6), 104726, DOI: [10.1016/j.jbc.2023.104726](https://doi.org/10.1016/j.jbc.2023.104726).
- 20 M. M. T. El-Tahawy, A. Nenov, O. Weingart, M. Olivucci and M. Garavelli, Relationship between Excited State Lifetime and Isomerization Quantum Yield in Animal Rhodopsins: Beyond the One-Dimensional Landau-Zener Model, *J. Phys. Chem. Lett.*, 2018, **9**(12), 3315–3322, DOI: [10.1021/acs.jpcclett.8b01062](https://doi.org/10.1021/acs.jpcclett.8b01062).
- 21 T. Mizukami, H. Kandori, Y. Shichida, A. H. Chen, F. Derguini, C. G. Caldwell, C. F. Biffe, K. Nakanishi and T. Yoshizawa, Photoisomerization Mechanism of the Rhodopsin Chromophore: Picosecond Photolysis of Pigment Containing 11-Cis-Locked Eight-Membered Ring Retinal, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**(9), 4072–4076, DOI: [10.1073/pnas.90.9.4072](https://doi.org/10.1073/pnas.90.9.4072).
- 22 K. Inoue, Photochemistry of the Retinal Chromophore in Microbial Rhodopsins, *J. Phys. Chem. B*, 2023, **127**(43), 9215–9222, DOI: [10.1021/acs.jpcc.3c05467](https://doi.org/10.1021/acs.jpcc.3c05467).
- 23 C. Chang, H. Kuramochi, M. Singh, R. Abe-Yoshizumi, T. Tsukuda, H. Kandori and T. Tahara, A Unified View on Varied Ultrafast Dynamics of the Primary Process in Microbial Rhodopsins, *Angew. Chem., Int. Ed.*, 2022, **61**(2), e202111930, DOI: [10.1002/anie.202111930](https://doi.org/10.1002/anie.202111930).
- 24 R. Govindjee, S. P. Balashov and T. G. Ebrey, Quantum Efficiency of the Photochemical Cycle of Bacteriorhodopsin, *Biophys. J.*, 1990, **58**(3), 597–608, DOI: [10.1016/S0006-3495\(90\)82403-6](https://doi.org/10.1016/S0006-3495(90)82403-6).
- 25 O. A. Smitienko, T. B. Feldman, L. E. Petrovskaya, O. V. Nekrasova, M. A. Yakovleva, I. V. Shelaev, F. E. Gostev, D. A. Cherepanov, I. B. Kolchugina, D. A. Dolgikh, V. A. Nadtochenko, M. P. Kirpichnikov and M. A. Ostrovsky, Comparative Femtosecond Spectroscopy of Primary Photoreactions of *Exiguobacterium Sibiricum* Rhodopsin and *Halobacterium Salinarum* Bacteriorhodopsin, *J. Phys. Chem. B*, 2021, **125**(4), 995–1008, DOI: [10.1021/acs.jpcc.0c07763](https://doi.org/10.1021/acs.jpcc.0c07763).
- 26 T. B. Feldman, O. A. Smitienko, I. V. Shelaev, F. E. Gostev, O. V. Nekrasova, D. A. Dolgikh, V. A. Nadtochenko, M. P. Kirpichnikov and M. A. Ostrovsky, Femtosecond Spectroscopic Study of Photochromic Reactions of Bacteriorhodopsin and Visual Rhodopsin, *J. Photochem. Photobiol., B*, 2016, **164**, 296–305, DOI: [10.1016/j.jphotobiol.2016.09.041](https://doi.org/10.1016/j.jphotobiol.2016.09.041).
- 27 C.-F. Chang, M. Konno, K. Inoue and T. Tahara, Effects of the Unique Chromophore-Protein Interactions on the Primary Photoreaction of Schizorhodopsin, *J. Phys. Chem. Lett.*, 2023, **14**(31), 7083–7091, DOI: [10.1021/acs.jpcclett.3c01133](https://doi.org/10.1021/acs.jpcclett.3c01133).
- 28 A. Pushkarev, K. Inoue, S. Larom, J. Flores-Urbe, M. Singh, M. Konno, S. Tomida, S. Ito, R. Nakamura, S. P. Tsunoda, A. Filosof, I. Sharon, N. Yutin, E. V. Koonin, H. Kandori and O. Bèjà, A Distinct Abundant Group of Microbial Rhodopsins Discovered Using Functional Metagenomics, *Nature*, 2018, **558**(7711), 595–599, DOI: [10.1038/s41586-018-0225-9](https://doi.org/10.1038/s41586-018-0225-9).
- 29 A. Shibukawa, K. Kojima, Y. Nakajima, Y. Nishimura, S. Yoshizawa and Y. Sudo, Photochemical Characterization of a New Heliorhodopsin from the Gram-negative Eubacterium *Bellilinea Caldifistulae* (BcHeR) and Comparison with Heliorhodopsin-48C12, *Biochemistry*, 2019, **58**(26), 2934–2943, DOI: [10.1021/acs.biochem.9b00257](https://doi.org/10.1021/acs.biochem.9b00257).
- 30 J. Flores-Urbe, G. Hevroni, R. Ghai, A. Pushkarev, K. Inoue, H. Kandori and O. Bèjà, Heliorhodopsins Are Absent in Diderm (Gram-negative) Bacteria: Some Thoughts and Possible Implications for Activity, *Environ. Microbiol. Rep.*, 2019, **11**(3), 419–424, DOI: [10.1111/1758-2229.12730](https://doi.org/10.1111/1758-2229.12730).
- 31 J. Shim, S.-G. Cho, S.-H. Kim, K. Chuon, S. Meas, A. Choi and K.-H. Jung, Heliorhodopsin Helps Photolyase to Enhance the DNA Repair Capacity, *Microbiol. Spectrosc.*, 2022, **10**(6), e02215–22, DOI: [10.1128/spectrum.02215-22](https://doi.org/10.1128/spectrum.02215-22).
- 32 S. Hososhima, R. Mizutori, R. Abe-Yoshizumi, A. Rozenberg, S. Shigemura, A. Pushkarev, M. Konno, K. Katayama, K. Inoue, S. P. Tsunoda, O. Bèjà and H. Kandori, Proton-Transporting Heliorhodopsins from Marine Giant Viruses, *eLife*, 2022, **11**, DOI: [10.7554/eLife.78416](https://doi.org/10.7554/eLife.78416).
- 33 K. Kovalev, D. Volkov, R. Astashkin, A. Alekseev, I. Gushchin, J. M. Haro-Moreno, I. Chizhov, S. Siletsky, M. Mamedov, A. Rogachev, T. Balandin, V. Borshchevskiy, A. Popov, G. Bourenkov, E. Bamberg, F. Rodriguez-Valera, G. Büldt and V. Gordelyi, High-Resolution Structural Insights into the Heliorhodopsin Family, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**(8), 4131–4141, DOI: [10.1073/pnas.1915888117](https://doi.org/10.1073/pnas.1915888117).
- 34 W. Shihoya, K. Inoue, M. Singh, M. Konno, S. Hososhima, K. Yamashita, K. Ikeda, A. Higuchi, T. Izume, S. Okazaki, M. Hashimoto, R. Mizutori, S. Tomida, Y. Yamauchi, R. Abe-Yoshizumi, K. Katayama, S. P. Tsunoda, M. Shibata, Y. Furutani, A. Pushkarev, O. Bèjà, T. Uchihashi, H. Kandori and O. Nureki, Crystal Structure of

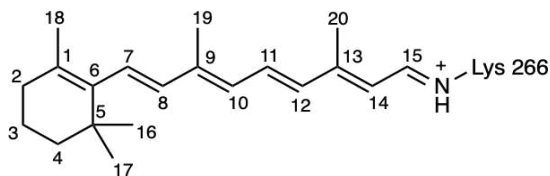
- Heliorhodopsin, *Nature*, 2019, **574**(7776), 132–136, DOI: [10.1038/s41586-019-1604-6](https://doi.org/10.1038/s41586-019-1604-6).
- 35 S. Tahara, M. Singh, H. Kuramochi, W. Shihoya, K. Inoue, O. Nureki, O. Bèjà, Y. Mizutani, H. Kandori and T. Tahara, Ultrafast Dynamics of Heliorhodopsins, *J. Phys. Chem. B*, 2019, **123**(11), 2507–2512, DOI: [10.1021/acs.jpcc.9b00887](https://doi.org/10.1021/acs.jpcc.9b00887).
- 36 M. Tsujimura, Y. Chiba, K. Saito and H. Ishikita, Proton Transfer and Conformational Changes along the Hydrogen Bond Network in Heliorhodopsin, *Commun. Biol.*, 2022, **5**(1), 1336, DOI: [10.1038/s42003-022-04311-x](https://doi.org/10.1038/s42003-022-04311-x).
- 37 J. Tittor and D. Oesterhelt, The Quantum Yield of Bacteriorhodopsin, *FEBS Lett.*, 1990, **263**(2), 269–273, DOI: [10.1016/0014-5793\(90\)81390-A](https://doi.org/10.1016/0014-5793(90)81390-A).
- 38 J. E. Kim, M. J. Tauber and R. A. Mathies, Wavelength Dependent Cis-Trans Isomerization in Vision, *Biochemistry*, 2001, **40**(46), 13774–13778, DOI: [10.1021/bi0116137](https://doi.org/10.1021/bi0116137).
- 39 J. C. Tully, Molecular Dynamics with Electronic Transitions, *J. Chem. Phys.*, 1990, **93**(2), 1061–1071, DOI: [10.1063/1.459170](https://doi.org/10.1063/1.459170).
- 40 F. Melaccio, M. del Carmen Marín, A. Valentini, F. Montisci, S. Rinaldi, M. Cherubini, X. Yang, Y. Kato, M. Stenrup, Y. Orozco-Gonzalez, N. Ferré, H. L. Luk, H. Kandori and M. Olivucci, Toward Automatic Rhodopsin Modeling as a Tool for High-Throughput Computational Photobiology, *J. Chem. Theory Comput.*, 2016, **12**(12), 6020–6034, DOI: [10.1021/acs.jctc.6b00367](https://doi.org/10.1021/acs.jctc.6b00367).
- 41 L. Pedraza-González, L. De Vico, M. Marín del C., F. Fanelli and M. Olivucci, *a*-ARM: Automatic Rhodopsin Modeling with Chromophore Cavity Generation, Ionization State Selection, and External Counterion Placement, *J. Chem. Theory Comput.*, 2019, **15**(5), 3134–3152, DOI: [10.1021/acs.jctc.9b00061](https://doi.org/10.1021/acs.jctc.9b00061).
- 42 L. Pedraza-González, L. Barneschi, D. Padula, L. De Vico and M. Olivucci, Evolution of the Automatic Rhodopsin Modeling (ARM) Protocol, *Top. Curr. Chem.*, 2022, **380**(3), 21, DOI: [10.1007/s41061-022-00374-w](https://doi.org/10.1007/s41061-022-00374-w).
- 43 L. Pedraza-González, M. D. C. Marín, A. N. Jorge, T. D. Ruck, X. Yang, A. Valentini, M. Olivucci and L. De Vico, Web-ARM: A Web-Based Interface for the Automatic Construction of QM/MM Models of Rhodopsins, *J. Chem. Inf. Model.*, 2020, **60**(3), 1481–1493, DOI: [10.1021/acs.jcim.9b00615](https://doi.org/10.1021/acs.jcim.9b00615).
- 44 Y. Nakajima, L. Pedraza-González, L. Barneschi, K. Inoue, M. Olivucci and H. Kandori, Pro219 is an electrostatic color determinant in the light-driven sodium pump KR2, *Commun. Biol.*, 2021, **4**(1), 1185, DOI: [10.1038/s42003-021-02684-z](https://doi.org/10.1038/s42003-021-02684-z).
- 45 R. Palombo, L. Barneschi, L. Pedraza-González, D. Padula, I. Schapiro and M. Olivucci, Retinal chromophore charge delocalization and confinement explain the extreme photophysics of Neorhodopsin, *Nat. Commun.*, 2022, **13**(1), 6652, DOI: [10.1038/s41467-022-33953-y](https://doi.org/10.1038/s41467-022-33953-y).
- 46 T. Tanaka, M. Singh, W. Shihoya, K. Yamashita, H. Kandori and O. Nureki, Structural Basis for Unique Color Tuning Mechanism in Heliorhodopsin, *Biochem. Biophys. Res. Commun.*, 2020, **533**(3), 262–267, DOI: [10.1016/j.bbrc.2020.06.124](https://doi.org/10.1016/j.bbrc.2020.06.124).
- 47 H. Kandori, Y. Shichida and T. Yoshizawa, Photoisomerization in Rhodopsin, *Biochemistry*, 2001, **66**(11), 1197–1209, DOI: [10.1023/A:1013123016803](https://doi.org/10.1023/A:1013123016803).
- 48 L. M. Frutos, T. Andruniów, F. Santoro, N. Ferré and M. Olivucci, Tracking the Excited-State Time Evolution of the Visual Pigment with Multiconfigurational Quantum Chemistry, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**(19), 7764–7769, DOI: [10.1073/pnas.0701732104](https://doi.org/10.1073/pnas.0701732104).
- 49 P. Altoè, A. Cembran, M. Olivucci and M. Garavelli, Aborted Double Bicycle-Pedal Isomerization with Hydrogen Bond Breaking Is the Primary Event of Bacteriorhodopsin Proton Pumping, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**(47), 20172–20177, DOI: [10.1073/pnas.1007000107](https://doi.org/10.1073/pnas.1007000107).
- 50 A. Strambi, B. Durbeej, N. Ferré and M. Olivucci, *Anabaena* Sensory Rhodopsin Is a Light-Driven Unidirectional Rotor, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**(50), 21322–21326, DOI: [10.1073/pnas.1015085107](https://doi.org/10.1073/pnas.1015085107).
- 51 Y. Boeijs and M. Olivucci, From a One-Mode to a Multi-Mode Understanding of Conical Intersection Mediated Ultrafast Organic Photochemical Reactions, *Chem. Soc. Rev.*, 2023, **52**(8), 2643–2687, DOI: [10.1039/D2CS00719C](https://doi.org/10.1039/D2CS00719C).
- 52 R. W. Schoenlein, L. A. Peteanu, R. A. Mathies and C. V. Shank, The First Step in Vision: Femtosecond Isomerization of Rhodopsin, *Science*, 1991, **254**(5030), 412–415, DOI: [10.1126/science.1925597](https://doi.org/10.1126/science.1925597).
- 53 C. Xu, K. Lin, D. Hu, F. L. Gu, M. F. Gelin and Z. Lan, Ultrafast Internal Conversion Dynamics through the On-the-Fly Simulation of Transient Absorption Pump-Probe Spectra with Different Electronic Structure Methods, *J. Phys. Chem. Lett.*, 2022, **13**(2), 661–668, DOI: [10.1021/acs.jpcclett.1c03373](https://doi.org/10.1021/acs.jpcclett.1c03373).
- 54 Q. Wang, R. W. Schoenlein, L. A. Peteanu, R. A. Mathies and C. V. Shank, Vibrationally Coherent Photochemistry in the Femtosecond Primary Event of Vision, *Science*, 1994, **266**(5184), 422–424, DOI: [10.1126/science.7939680](https://doi.org/10.1126/science.7939680).
- 55 D. Polli, O. Weingart, D. Brida, E. Poli, M. Maiuri, K. M. Spillane, A. Bottoni, P. Kukura, R. A. Mathies, G. Cerullo and M. Garavelli, Wavepacket Splitting and Two-Pathway Deactivation in the Photoexcited Visual Pigment Isorhodopsin, *Angew. Chem., Int. Ed.*, 2014, **53**(9), 2504–2507, DOI: [10.1002/anie.201309867](https://doi.org/10.1002/anie.201309867).
- 56 A. Warshel, Bicycle-Pedal Model for the First Step in the Vision Process, *Nature*, 1976, **260**(5553), 679–683, DOI: [10.1038/260679a0](https://doi.org/10.1038/260679a0).
- 57 H. L. Luk, F. Melaccio, S. Rinaldi, S. Gozem and M. Olivucci, Molecular Bases for the Selection of the Chromophore of Animal Rhodopsins, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**(50), 15297–15302, DOI: [10.1073/pnas.1510262112](https://doi.org/10.1073/pnas.1510262112).
- 58 S. Rinaldi, F. Melaccio, S. Gozem, F. Fanelli and M. Olivucci, Comparison of the Isomerization Mechanisms of Human Melanopsin and Invertebrate and Vertebrate Rhodopsins, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**(5), 1714–1719, DOI: [10.1073/pnas.1309508111](https://doi.org/10.1073/pnas.1309508111).
- 59 M. Schmidt and E. A. Stojković, Earliest Molecular Events of Vision Revealed, *Nature*, 2023, **615**(7954), 802–803, DOI: [10.1038/d41586-023-00504-4](https://doi.org/10.1038/d41586-023-00504-4).
- 60 P. Nogly, T. Weinert, D. James, S. Carbajo, D. Ozerov, A. Furrer, D. Gashi, V. Borin, P. Skopintsev, K. Jaeger, K. Nass, P. Båth, R. Bosman, J. Koglin, M. Seaberg,

- T. Lane, D. Kekilli, S. Brünle, T. Tanaka, W. Wu, C. Milne, T. White, A. Barty, U. Weierstall, V. Panneels, E. Nango, S. Iwata, M. Hunter, I. Schapiro, G. Schertler, R. Neutze and J. Standfuss, Retinal isomerization in bacteriorhodopsin captured by a femtosecond X-ray laser, *Science*, 2018, **361**(6398), DOI: [10.1126/science.aat0094](https://doi.org/10.1126/science.aat0094).
- 61 C. Wittig, The Landau–Zener Formula, *J. Phys. Chem. B*, 2005, **109**(17), 8428–8430, DOI: [10.1021/jp040627u](https://doi.org/10.1021/jp040627u).
- 62 S. S. Batsanov, van der Waals Radii of Elements, *Inorg. Mater.*, 2001, **37**(9), 871–885, DOI: [10.1023/A:1011625728803](https://doi.org/10.1023/A:1011625728803).
- 63 T. Arlt, S. Schmidt, W. Zinth, U. Haupts and D. Oesterhelt, The Initial Reaction Dynamics of the Light-Driven Chloride Pump Halorhodopsin, *Chem. Phys. Lett.*, 1995, **241**(5–6), 559–565, DOI: [10.1016/0009-2614\(95\)00664-P](https://doi.org/10.1016/0009-2614(95)00664-P).
- 64 D. Oesterhelt, P. Hegemann and J. Tittor, The Photocycle of the Chloride Pump Halorhodopsin. II: Quantum Yields and a Kinetic Model, *EMBO J.*, 1985, **4**(9), 2351–2356, DOI: [10.1002/j.1460-2075.1985.tb03938.x](https://doi.org/10.1002/j.1460-2075.1985.tb03938.x).
- 65 M. O. Lenz, R. Huber, B. Schmidt, P. Gilch, R. Kalmbach, M. Engelhard and J. Wachtveitl, First Steps of Retinal Photoisomerization in Proteorhodopsin, *Biophys. J.*, 2006, **91**(1), 255–262, DOI: [10.1529/biophysj.105.074690](https://doi.org/10.1529/biophysj.105.074690).
- 66 K. Neumann, M.-K. Verhoefen, I. Weber, C. Glaubitz and J. Wachtveitl, Initial Reaction Dynamics of Proteorhodopsin Observed by Femtosecond Infrared and Visible Spectroscopy, *Biophys. J.*, 2008, **94**(12), 4796–4807, DOI: [10.1529/biophysj.107.125484](https://doi.org/10.1529/biophysj.107.125484).
- 67 Y. Sudo, M. Mizuno, Z. Wei, S. Takeuchi, T. Tahara and Y. Mizutani, The Early Steps in the Photocycle of a Photosensor Protein Sensory Rhodopsin I from *Salinibacter Ruber*, *J. Phys. Chem. B*, 2014, **118**(6), 1510–1518, DOI: [10.1021/jp4112662](https://doi.org/10.1021/jp4112662).
- 68 Y. Hontani, M. Marazzi, K. Stehfest, T. Mathes, I. H. M. van Stokkum, M. Elstner, P. Hegemann and J. T. M. Kennis, Reaction Dynamics of the Chimeric Channelrhodopsin C1C2, *Sci. Rep.*, 2017, **7**(1), 7217, DOI: [10.1038/s41598-017-07363-w](https://doi.org/10.1038/s41598-017-07363-w).
- 69 S. Tahara, S. Takeuchi, R. Abe-Yoshizumi, K. Inoue, H. Ohtani, H. Kandori and T. Tahara, Origin of the Reactive and Nonreactive Excited States in the Primary Reaction of Rhodopsins: PH Dependence of Femtosecond Absorption of Light-Driven Sodium Ion Pump Rhodopsin KR2, *J. Phys. Chem. B*, 2018, **122**(18), 4784–4792, DOI: [10.1021/acs.jpcc.8b01934](https://doi.org/10.1021/acs.jpcc.8b01934).
- 70 Y. Sudo, K. Ihara, S. Kobayashi, D. Suzuki, H. Irieda, T. Kikukawa, H. Kandori and M. Homma, A Microbial Rhodopsin with a Unique Retinal Composition Shows Both Sensory Rhodopsin II and Bacteriorhodopsin-like Properties, *J. Biol. Chem.*, 2011, **286**(8), 5967–5976, DOI: [10.1074/jbc.M110.190058](https://doi.org/10.1074/jbc.M110.190058).
- 71 S. Kaziannis, M. Broser, I. H. M. van Stokkum, J. Dostal, W. Busse, A. Munhoven, C. Bernardo, M. Kloz, P. Hegemann and J. T. M. Kennis, Multiple retinal isomerizations during the early phase of the bacteriorhodopsin photoreaction, *Proc. Natl. Acad. Sci. U. S. A.*, 2024, **121**(12), e2318996121, DOI: [10.1073/pnas.2318996121](https://doi.org/10.1073/pnas.2318996121).
- 72 T. Okada, M. Sugihara, A.-N. Bondar, M. Elstner, P. Entel and V. Buss, The Retinal Conformation and Its Environment in Rhodopsin in Light of a New 2.2 Å Crystal Structure, *J. Mol. Biol.*, 2004, **342**(2), 571–583, DOI: [10.1016/j.jmb.2004.07.044](https://doi.org/10.1016/j.jmb.2004.07.044).
- 73 G. Granucci and M. Persico, Critical Appraisal of the Fewest Switches Algorithm for Surface Hopping, *J. Chem. Phys.*, 2007, **126**(13), 134114, DOI: [10.1063/1.2715585](https://doi.org/10.1063/1.2715585).
- 74 S. Gozem, M. Huntress, I. Schapiro, R. Lindh, A. A. Granovsky, C. Angeli and M. Olivucci, Dynamic Electron Correlation Effects on the Ground State Potential Energy Surface of a Retinal Chromophore Model, *J. Chem. Theory Comput.*, 2012, **8**(11), 4069–4080, DOI: [10.1021/ct3003139](https://doi.org/10.1021/ct3003139).
- 75 L. Barneschi, D. Kaliakin, M. Huix-Rotllant, N. Ferré, M. Filatov(Gulak) and M. Olivucci, Assessment of the Electron Correlation Treatment on the Quantum-Classical Dynamics of Retinal Protonated Schiff Base Models: XMS-CASPT2, RMS-CASPT2, and REKS Methods, *J. Chem. Theory Comput.*, 2023, **19**(22), 8189–8200, DOI: [10.1021/acs.jctc.3c00879](https://doi.org/10.1021/acs.jctc.3c00879).
- 76 S. Sen and I. Schapiro, A Comprehensive Benchmark of the XMS-CASPT2 Method for the Photochemistry of a Retinal Chromophore Model, *Mol. Phys.*, 2018, **116**(19–20), 2571–2582, DOI: [10.1080/00268976.2018.1501112](https://doi.org/10.1080/00268976.2018.1501112).
- 77 S. Gozem, M. Huntress, I. Schapiro, R. Lindh, A. A. Granovsky, C. Angeli and M. Olivucci, Dynamic Electron Correlation Effects on the Ground State Potential Energy Surface of a Retinal Chromophore Model, *J. Chem. Theory Comput.*, 2012, **8**(11), 4069–4080, DOI: [10.1021/ct3003139](https://doi.org/10.1021/ct3003139).
- 78 F. Aquilante, J. Autschbach, R. K. Carlson, L. F. Chibotaru, M. G. Delcey, L. De Vico, I. F. Galván, N. Ferré, L. M. Frutos, L. Gagliardi, M. Garavelli, A. Giussani, C. E. Hoyer, G. Li Manni, H. Lischka, D. Ma, P. Å. Malmqvist, T. Müller, A. Nenov, M. Olivucci, T. B. Pedersen, D. Peng, F. Plasser, B. Pritchard, M. Reiher, I. Rivalta, I. Schapiro, J. Segarra-Martí, M. Stenrup, D. G. Truhlar, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, V. P. Vysotskiy, O. Weingart, F. Zapata and R. Lindh, MOLCAS 8: New Capabilities for Multiconfigurational Quantum Chemical Calculations across the Periodic Table, *J. Comput. Chem.*, 2016, **37**(5), 506–541, DOI: [10.1002/jcc.24221](https://doi.org/10.1002/jcc.24221).
- 79 J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren and J. W. Ponder, Tinker 8: Software Tools for Molecular Design, *J. Chem. Theory Comput.*, 2018, **14**(10), 5273–5289, DOI: [10.1021/acs.jctc.8b00529](https://doi.org/10.1021/acs.jctc.8b00529).

Chapter 6

Appendix

Normal mode vibrations of the rPSB chromophore in the NeoR QM/MM model presented and discussed in Section 3.1. Frequencies are given in cm^{-1} , and relative intensities (I_{rel}) are related to the strongest mode (= 1.0).



ν	I_{rel}	Normal mode description (only contributions > 0.50 are reported)
880	0.0	Lysine vibrations
886	0.0	$(\text{C12-C11-C10-H}) + 0.874(\text{C8-C9-C10-H}) - 0.775(\text{C9-C10-C11-H}[-\text{C10}])_{\text{outofplane}} + 0.667(\text{C9-C10-C11-H}) + 0.615(\text{C19-C9-C10-H}) - 0.516(\text{C13-C12-C11-H})$
918	0.007	$(\text{C5-C16-H}) - 0.789(\text{C5-C17-H}) + 0.743(\text{C5-C17}) - 0.627(\text{C5-C16-H}) - 0.541(\text{C5-C16})$
921	0.0	Lysine vibrations
927	0.005	Lysine vibrations
929	0.0	$(\text{C2-C3}) - 0.735(\text{C4-C5}) + 0.703(\text{C5-C16}) + 0.681(\text{C5-C17-H}) + 0.611(\text{C3-C4-H}) - 0.541(\text{C5-C17-H})$
939	0.041	$(\text{H-C14-C15-H}) + 0.971(\text{H-N-C15-H}) - 0.860(\text{C13-C14-C15-H}) - 0.716(\text{C14-C15-N-H}[-\text{C15}])_{\text{outofplane}} + \text{lysine vibrations}$
955	0.001	$(\text{H-C7-C8-H}) - 0.728(\text{H-C11-C12-H}) - 0.570(\text{C1-C6-C7-H}) + 0.559(\text{C9-C10-C11-H}) + 0.548(\text{H-C10-C11-H}) - 0.507(\text{H-C8-C9-C10})$
973	0.009	$(\text{H-C7-C8-H}) + 0.901(\text{H-C11-C12-H}) - 0.805(\text{H-C10-C11-H}) - 0.652(\text{H-C8-C9-C10}) + 0.636(\text{H-C11-C12-C13}) - 0.629(\text{C9-C10-C11-H}) + 0.520(\text{C6-C7-C8-H})$
981	0.006	$-(\text{C4-C3-H}) - 0.722(\text{C2-C3-H}) + 0.686(\text{C4-C3-H}) - 0.683(\text{C1-C18-H}) - 0.577(\text{C5-C16-H}) + 0.554(\text{C2-C3-H}) + 0.552(\text{C5-C17-H}) + 0.503(\text{C5-C4-H})$
986	0.0	$(\text{C1-C18-H}) - 0.915(\text{C1-C18-H}) - 0.561(\text{C1-C2-H}) - 0.552(\text{C5-C4-H}) + 0.541(\text{C5-C4-H})$
995	0.285	$-(\text{H-C14-C15-H}) + 0.983(\text{H-N-C15-H}) - 0.836(\text{C14-C14-C15-H}) - 0.772(\text{H}[-\text{C15}]-\text{N-C15-C14})_{\text{outofplane}} + 0.646(\text{C9-C19-H}) - 0.634(\text{C13-C20-H}) - 0.619(\text{C9-C19-H}) + 0.578(\text{C13-C20-H}) + \text{lysine vibrations}$
1013	0.009	$(\text{C1-C18-H}) + 0.998(\text{C9-C19-H}) + 0.968(\text{C5-C16-H}) - 0.918(\text{C5-C16-H}) - 0.918(\text{C5-C17-H}) + 0.901(\text{C5-C17-H}) - 0.857(\text{C9-C19-H}) - 0.593(\text{C1-C18-H}) - 0.512(\text{C7-C6-C1-C18})$
1014	0.003	Lysine vibrations
1017	0.008	$(\text{C9-C19-H}) - 0.669(\text{C13-C20-H}) - 0.577(\text{C9-C19-H}) + 0.572(\text{H-C7-C8-H}) + 0.534(\text{C19-C9-C10-H})$
1023	0.086	$-(\text{C9-C19-H}) - 0.685(\text{C13-C20-H}) + 0.664(\text{H-C11-C12-H}) - 0.664(\text{H-C10-C11-H}) + 0.647(\text{C13-C20-H}) + 0.619(\text{C9-C19-H}) - 0.593(\text{C9-C10-C11-H}) + 0.589(\text{C13-C12-C11-H}) - 0.502(\text{H-C7-C8-H})$

1034	0.022	$(C13-C20-H) - 0.769(C13-C20-H) + 0.753(C13-C14-C15-H) + 0.731(H-C14-C15-H) - 0.657(H-N-C15-H) + 0.555(H[-C15]-N-C14-C15)_{outofplane} + \text{lysine vibrations}$
1042	0.155	$(C5-C17-H) - 0.918(C2-C3) + 0.800(C18-C1) + 0.773(C3-C4) + 0.765(C4-C3-H) + 0.749(C3-C4-H) - 0.731(C5-C4-H) - 0.621(C5-C17-H) + 0.543(C1-C18-H) + 0.542(C9-C19-H) - 0.537(C13-C20-H) - 0.523(C2-C3-H)$
1044	0.008	$-(C1-C18-H) + 0.874(C1-C18-H) + 0.778(C5-C16-H) + 0.686(C4-C3-H) - 0.593(C19-C1) - 0.544(C4-C3-H) - 0.526(C5-C17-H)$
1061	0.0	$(C2-C3-H) - 0.921(C1-C18-H) - 0.825(C5-C17-H) + 0.758(C3-C2-H) + 0.728(C5-C4-H) - 0.713(C3-C4) + 0.648(C5-C17-H) - 0.645(C1-C2-H) + 0.625(C1-C18-H) - 0.555(C4-C3-H)$
1091	0.412	$-(C6-C7) - 0.609(C1-C18) + 0.598(C1-C2) + 0.586(C1-C18-H) + 0.503(C3-C4-H)$
1097	0.012	Lysine vibrations
1109	0.021	Lysine vibrations
1133	0.004	$(C1-C2-H) - 0.802(C2-C3-H) - 0.743(C1-C2-H) + 0.677(C3-C4-H) - 0.608(C2-C3-H)$
1141	0.006	$(C10-C11) - 0.919(H-C14-C15-H) - 0.810(C13-C20-H) + 0.711(H-N-C15-H) + 0.637(C13-C20-H) - 0.599(C13-C14-C15-H) - 0.572(C15-N) - 0.562(H[-C15]-N-C14-C15)_{outofplane} - 0.505(C13-C14) + \text{lysine vibrations}$
1147	0.0	Lysine vibrations
1177	0.139	$-(C9-C19-H) + 0.941(C8-C9) - 0.887(C5-C16-H) - 0.872(C10-C11) - 0.806(C11-C12) + 0.767(C3-C4-H) + 0.756(C9-C8-H) - 0.702(C9-C10-H) + 0.701(C17-C5) + 0.685(C1-C2-H) - 0.665(C5-C4-H) + 0.637(C10-C11-C12) + 0.598(C14-C15) - 0.594(C16-C5-C17) - 0.579(C6-C5) - 0.562(C10-C11-H) + 0.557(C1-C18-H) + 0.551(C9-C19-H) - 0.549(C1-C18-H) - 0.547(C2-C3-H) + 0.537(C14-C15-H) + 0.537(C5-C16-H) + 0.512(H-C19-H) + 0.510(C8-C7-H)$
1181	0.012	$-(C5-C16-H) + 0.833(C5-C16-H) - 0.794(C18-C1) - 0.771(C3-C4-H) - 0.712(C17-C5) - 0.710(C5-C17-H) + 0.705(C2-C3-H) + 0.614(C1-C2) - 0.605(C6-C5-C16) + 0.536(C6-C5) + 0.508(C4-C5-C17) + 0.503(C5-C6-C1)$
1191	0.004	$-(C5-C4-H) + 0.935(C5-C16-H) + 0.811(C5-C17-H) + 0.742(C1-C2-H) - 0.719(C5-C16-H) + 0.715(C6-C7) - 0.711(C3-C2-H) - 0.647(C4-C5) + 0.556(C4-C5-C16) + 0.533(C11-C12) + 0.505(C17-C5)$
1213	0.004	$-(C12-C11-H) + 0.919(C10-C11-H) + 0.508(C14-C15-H)$
1219	0.001	$-(C4-C3-H) + 0.939(C2-C3-H) - 0.738(C1-C2-H) + 0.656(C5-C4-H) + 0.629(C3-C2-H) - 0.534(C4-C5) + 0.504(C5-C16-H)$
1223	0.013	$(C4-C3-H) + 0.990(C5-C4-H) + 0.757(C5-C17-H) + 0.736(C3-C2-H) - 0.713(C1-C2-H) - 0.709(C16-C5) - 0.709(C4-C3-H) - 0.553(C4-C5-C17) - 0.550(C5-C17-H) - 0.509(C3-C4-H)$
1227	0.002	Lysine vibrations
1250	0.037	$-(C11-C10-H) + 0.674(C15-C14-H) + 0.663(C9-C10-H) - 0.592(C10-C11) - 0.517(C8-C9) + 0.502(C19-C9)$
1258	0.0	Lysine vibrations
1281	0.179	$-(C7-C8-H) + 0.806(C9-C8-H) + 0.657(C10-C11-H) - 0.642(C12-C11-H) + 0.544(C11-C12-H)$

1287	0.0	$-(C3-C2-H) + 0.759(C1-C2-H) - 0.669(C3-C4-H) + 0.660(C7-C8-H) + 0.600(C3-C4-H) + 0.569(C3-C2-H)$
1294	0.048	$(C12-C11-H) - 0.895(C7-C8-H) + 0.869(C14-C15-H) + 0.836(C11-C12) + 0.812(C11-C12-H) + 0.719(C9-C8-H) + 0.699(C14-C15) - 0.593(C10-C11-H) - 0.579(C12-C13-C14) - 0.570(N-C15-H) - 0.557(C11-C12-C13) + 0.517(C11-C10-H) + \text{lysine vibrations}$
1297	0.0	Lysine vibrations
1307	0.013	$(C11-C12-H) - 0.788(C13-C12-H) + 0.741(C8-C9) + 0.616(C7-C8-H) + 0.577(C9-C10-H) + 0.519(C10-C11-H)$
1310	0.056	$-(C5-C4-H) + 0.864(C6-C5) + 0.816(C5-C4-H) - 0.638(C7-C8-H) + 0.619(C3-C4-H) - 0.546(C3-C4-H) + 0.545(C3-C2-H) + 0.516(C4-C3-H) + \text{lysine vibrations}$
1312	0.010	Lysine vibrations
1321	0.053	Lysine vibrations
1336	0.264	$(C8-C7-H) - 0.736(C6-C7-H) + 0.501(C7-C8)$
1345	0.002	Lysine vibrations
1362	0.007	$(C9-C10-H) - 0.794(C11-C10-H) - 0.760(C13-C14) + 0.704(H-N-C15) + 0.610(H-C19-H) - 0.579(N-C15-H) - 0.559(C12-C11-H) - 0.552(C14-C15-H) + \text{lysine vibrations}$
1370	0.005	$-(C4-C3-H) + 0.978(C3-C4-H) - 0.863(C5-C4-H) + 0.851(C2-C3-H) - 0.797(C5-C4-H) + 0.711(C3-C4-H) - 0.577(C3-C2-H) + 0.513(C2-C3-H)$
1370	0.016	$(H-C20-H) + 0.863(C13-C14-H) + 0.829(C12-C13) - 0.707(C13-C20-H) + 0.703(H-N-C15) + 0.696(C13-C12-H) - 0.561(C20-C13) + \text{lysine vibrations}$
1391	0.005	$(C2-C3-H) - 0.919(C1-C2-H) + 0.843(C3-C2-H) - 0.823(C4-C3-H) - 0.744(C1-C2-H) - 0.626(C2-C3) + 0.503(C3-C2-H)$
1391	0.012	$(H-C19-H) + 0.986(C13-C12-H) - 0.884(C9-C19-H) + 0.848(H-C19-H) - 0.844(C2-C3-H) - 0.839(C9-C19-H) + 0.777(C1-C2-H) - 0.743(C11-C12-H) - 0.687(C3-C2-H) + 0.680(C9-C10-H) + 0.651(C13-C14) + 0.673(C4-C3-H) + 0.604(C9-C10) - 0.566(H-N-C15) - 0.565(C11-C10-H) + 0.554(C9-C8-H) + 0.525(C12-C11-H) + 0.512(C1-C2-H) - 0.507(C9-C19-H) + \text{lysine vibrations}$
1393	0.002	Lysine vibrations
1396	0.010	$-(C4-C3-H) - 0.867(C3-C2-H) - 0.865(C3-C4-H) + 0.859(C1-C2-H) + 0.809(C5-C4-H) + 0.764(H-C16-H) + 0.759(C2-C3-H) + 0.726(C2-C3-H) + 0.644(C5-C4-H) - 0.618(C5-C16-H) + 0.602(H-C16-H) - 0.597(H-C17-H) + 0.557(C3-C4) - 0.548(C5-C16-H) + 0.532(C1-C2-H)$
1402	0.006	$(H-C16-H) - 0.935(H-C17-H) - 0.660(H-C17-H) + 0.637(H-C16-H) - 0.636(C5-C16-H) - 0.632(C5-C16-H) + 0.627(C5-C17-H) - 0.581(C5-C4-H) + 0.579(C5-C17-H) + 0.554(C3-C4-H) - 0.553(C5-C16-H) + 0.533(C5-C17-H) + 0.500(C4-C3-H)$
1410	0.002	$-(H-C19-H) + 0.675(H-C20-H) + \text{lysine vibrations}$
1412	0.003	$(H-C19-H)$
1416	0.040	$(H-C18-H) - 0.598(C1-C18-H) + 0.589(H-C18-H) - 0.540(C1-C18-H) - 0.537(C1-C18-H)$
1419	0.020	$(H-C20-H) + 0.933(H-C20-H) - 0.691(C13-C20-H) - 0.559(C13-C20-H) - 0.501(C13-C20-H)$

1423	0.009	$(\text{H-C17-H}) + 0.774(\text{H-C16-H}) + 0.689(\text{H-C18-H}) - 0.635(\text{C5-C17-H}) - 0.612(\text{C5-C17-H}) + 0.570(\text{H-C16-H}) + 0.540(\text{H-C17-H}) - 0.532(\text{C5-C16-H}) - 0.529(\text{C5-C16-H}) - 0.509(\text{C5-C17-H})$
1441	0.003	$(\text{H-C20-H}) + 0.635(\text{H-C20-H}) - 0.519(\text{C15-C14-H}) + 0.510(\text{C13-C14-H}) + \text{lysine vibrations}$
1451	0.002	$(\text{H-C19-H}) + 0.645(\text{H-C19-H}) + 0.516(\text{H-C20-H})$
1455	0.006	$-(\text{H-C2-H}) + 0.738(\text{H-C19-H}) - 0.670(\text{H-C20-H}) - 0.647(\text{H-C19-H}) + 0.586(\text{H-C20-H})$
1456	0.010	(H-C2-H)
1465	0.002	$(\text{H-C18-H}) - 0.914(\text{H-C18-H}) - 0.500(\text{H-C-H})$
1470	0.006	$(\text{H-C16-H}) - 0.815(\text{H-C16-H}) - 0.691(\text{H-C3-H}) + 0.586(\text{H-C4-H})$
1475	0.0	$-(\text{H-C4-H}) - 0.742(\text{H-C16-H}) + 0.541(\text{H-C17-H})$
1477	0.008	$(\text{H-C19-H}) - 0.837(\text{H-C19-H}) + 0.599(\text{H-C19-H}) + \text{lysine vibrations}$
1482	0.016	$(\text{H-C3-H}) - 0.676(\text{H-C17-H}) + 0.665(\text{H-C17-H})$
1486	0.021	$(\text{H-C3-H}) + 0.690(\text{H-C19-H}) + 0.656(\text{N-C15-H}) + 0.638(\text{H-N-C15}) + 0.602(\text{H-C17-H}) + 0.531(\text{C12-C11-H}) - 0.524(\text{H-C16-H}) - 0.523(\text{C14-C15-H47}) + \text{lysine vibrations}$
1488	0.012	$-(\text{H-C16-H}) + 0.873(\text{H-C20-H}) + 0.792(\text{H-C17-H}) + 0.781(\text{H-C16-H}) - 0.599(\text{H-C17-H}) + 0.551(\text{H-C4-H}) + 0.541(\text{H-C3-H})$
1492	0.016	$(\text{H-C17-H}) - 0.643(\text{H-C17-H}) - 0.529(\text{H-C4-H})$
1494	0.005	(H-C20-H)
1495	0.028	$-(\text{H-C16-H}) - 0.889(\text{H-C4-H}) + 0.781(\text{H-C3-H}) + \text{lysine vibrations}$
1499	0.002	$(\text{H-C16-H}) + 0.931(\text{H-C3-H}) - 0.852(\text{H-C4-H}) + 0.665(\text{H-C17-H}) + 0.660(\text{H-C20-H}) - 0.650(\text{H-C16-H}) - 0.560(\text{H-C17-H}) + \text{lysine vibrations}$
1506	0.080	$-(\text{H-C18-H})$
1530	0.006	$-(\text{H-N-C15}) - 0.891(\text{C11-C12-H}) + 0.843(\text{C13-C12-H}) + 0.671(\text{C11-C12}) - 0.572(\text{C12-C11-H}) - 0.562(\text{C15-C14-H}) + 0.513(\text{C13-C14-H}) + \text{lysine vibrations}$
1554	0.822	$(\text{C1-C6}) + 0.921(\text{C9-C10}) + 0.844(\text{C11-C10-H}) - 0.806(\text{C9-C10-H}) + 0.736(\text{C12-C13}) - 0.719(\text{C13-C12-H}) + 0.659(\text{C11-C12-H}) - 0.550(\text{H-N-C15}) - 0.545(\text{C13-C14}) + \text{lysine vibrations}$
1567	0.090	$(\text{C1-C6}) + \text{lys}$
1569	1.00	$-(\text{C6-C7}) + 0.981(\text{C9-C8-H}) + 0.939(\text{C7-C8}) - 0.843(\text{C8-C7-H}) + 0.688(\text{C6-C7-H}) - 0.540(\text{C8-C9}) + \text{lysine vibrations}$
1594	0.542	$(\text{H-N-C15}) + 0.995(\text{C6-C7-H}) + 0.974(\text{C1-C6}) - 0.933(\text{C8-C7-H}) - 0.842(\text{C11-C10-H}) + 0.815(\text{C9-C10-H}) - 0.754(\text{C9-C10}) + 0.739(\text{C7-C8}) + 0.643(\text{C10-C11}) - 0.633(\text{C6-C7}) - 0.553(\text{C10-C11-H}) + \text{lysine vibrations}$
1646	0.0	$(\text{H-N-C15}) + 0.797(\text{N-C15-H}) - 0.706(\text{C14-C15-H}) - 0.676(\text{C15-N}) + 0.625(\text{C14-C15}) + \text{lysine vibrations}$
1759	0.008	Lysine vibrations

Bibliography

- [1] Oliver P Ernst, David T Lodowski, Marcus Elstner, Peter Hegemann, Leonid S Brown, and Hideki Kandori. Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. *Chemical reviews*, 114(1):126–163, 2014.
- [2] Maria del Carmen Marin, Damianos Agathangelou, Yoelvis Orozco-Gonzalez, Alessio Valentini, Yoshitaka Kato, Rei Abe-Yoshizumi, Hideki Kandori, Ahreum Choi, Kwang-Hwan Jung, Stefan Haacke, et al. Fluorescence enhancement of a microbial rhodopsin via electronic reprogramming. *Journal of the American Chemical Society*, 141(1):262–271, 2018.
- [3] Matthias Broser, Wayne Busse, Anika Spreen, Maila Reh, Yinth Andrea Bernal Sierra, Songhwan Hwang, Tillmann Utesch, Han Sun, and Peter Hegemann. Diversity of rhodopsin cyclases in zoospore-forming fungi. *Proceedings of the National Academy of Sciences*, 120(44):e2310600120, 2023.
- [4] Matthias Broser, Anika Spreen, Patrick E Konold, Enrico Peter, Suliman Adam, Veniamin Borin, Igor Schapiro, Reinhard Seifert, John TM Kennis, Yinth Andrea Bernal Sierra, et al. Neor, a near-infrared absorbing rhodopsin. *Nature communications*, 11(1):5682, 2020.
- [5] Judy E Kim, Michael J Tauber, and Richard A Mathies. Wavelength dependent cis-trans isomerization in vision. *Biochemistry*, 40(46):13774–13778, 2001.
- [6] Wataru Shihoya, Keiichi Inoue, Manish Singh, Masae Konno, Shoko Hososhima, Keitaro Yamashita, Kento Ikeda, Akimitsu Higuchi, Tamaki Izume, Sae Okazaki, et al. Crystal structure of heliorhodopsin. *Nature*, 574(7776):132–136, 2019.
- [7] Michael A van der Horst and Klaas J Hellingwerf. Photoreceptor proteins, “star actors of modern times”: a review of the functional dynamics in the structure of representative members of six different photoreceptor families. *Accounts of chemical research*, 37(1):13–20, 2004.
- [8] Karl Deisseroth. Optogenetics. *Nature methods*, 8(1):26–29, 2011.

- [9] Weiming Tu, Jiabao Xu, Ian P Thompson, and Wei E Huang. Engineering artificial photosynthesis based on rhodopsin for co2 fixation. *Nature Communications*, 14(1):8012, 2023.
- [10] Alina Pushkarev, Keiichi Inoue, Shirley Larom, José Flores-Uribe, Manish Singh, Masae Konno, Sahoko Tomida, Shota Ito, Ryoko Nakamura, Satoshi P Tsunoda, et al. A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. *Nature*, 558(7711):595–599, 2018.
- [11] Samer Gozem, Hoi Ling Luk, Igor Schapiro, and Massimo Olivucci. Theory and simulation of the ultrafast double-bond isomerization of biological chromophores. *Chemical reviews*, 117(22):13502–13565, 2017.
- [12] Hideki Kandori. Retinal proteins: photochemistry and optogenetics. *Bulletin of the Chemical Society of Japan*, 93(1):76–85, 2020.
- [13] Andrey Rozenberg, Keiichi Inoue, Hideki Kandori, and Oded Bèjà. Microbial rhodopsins: the last two decades. *Annual review of microbiology*, 75:427–447, 2021.
- [14] John M Christie, Lisa Blackwood, Jan Petersen, and Stuart Sullivan. Plant flavoprotein photoreceptors. *Plant and Cell Physiology*, 56(3):401–413, 2015.
- [15] Georg Nagel, Tanjef Szellas, Wolfram Huhn, Suneel Kateriya, Nona Adeishvili, Peter Berthold, Doris Ollig, Peter Hegemann, and Ernst Bamberg. Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proceedings of the National Academy of Sciences*, 100(24):13940–13945, 2003.
- [16] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic acids research*, 28(1):235–242, 2000.
- [17] Elisabeth P Carpenter, Konstantinos Beis, Alexander D Cameron, and So Iwata. Overcoming the challenges of membrane protein crystallography. *Current opinion in structural biology*, 18(5):581–586, 2008.
- [18] Daria N Shalaeva, Michael Y Galperin, and Armen Y Mulkidjanian. Eukaryotic g protein-coupled receptors as descendants of prokaryotic sodium-translocating rhodopsins. *Biology direct*, 10(1):1–12, 2015.
- [19] Keiichi Kojima and Yuki Sudo. Convergent evolution of animal and microbial rhodopsins. *RSC advances*, 13(8):5367–5381, 2023.
- [20] Dieter Oesterhelt and Walther Stoeckenius. Rhodopsin-like protein from the purple membrane of halobacterium halobium. *Nature new biology*, 233(39):149–152, 1971.

- [21] Akemi Matsuno-Yagi and Yasuo Mukohata. Two possible roles of bacteriorhodopsin; a comparative study of strains of halobacterium halobium differing in pigmentation. *Biochemical and biophysical research communications*, 78(1):237–243, 1977.
- [22] Kwang-Hwan Jung, Vishwa D Trivedi, and John L Spudich. Demonstration of a sensory rhodopsin in eubacteria. *Molecular microbiology*, 47(6):1513–1522, 2003.
- [23] John L Spudich and Roberto A Bogomolni. Mechanism of colour discrimination by a bacterial sensory rhodopsin. *Nature*, 312(5994):509–513, 1984.
- [24] Hartmann Harz and Peter Hegemann. Rhodopsin-regulated calcium currents in chlamydomonas. *Nature*, 351(6326):489–491, 1991.
- [25] Karl Deisseroth and Peter Hegemann. The form and function of channelrhodopsin. *Science*, 357(6356):eaan5544, 2017.
- [26] Hjalte V Kiefer, Elisabeth Gruber, Jeppe Langeland, Pavel A Kusocek, Anastasia V Bochenkova, and Lars H Andersen. Intrinsic photoisomerization dynamics of protonated schiff-base retinal. *Nature communications*, 10(1):1210, 2019.
- [27] P Hamm, M Zurek, T Röschinger, H Patzelt, D Oesterhelt, and W Zinth. Femtosecond spectroscopy of the photoisomerisation of the protonated schiff base of all-trans retinal. *Chemical physics letters*, 263(5):613–621, 1996.
- [28] Stephan L Logunov, Li Song, and Mostafa A El-Sayed. Excited-state dynamics of a protonated retinal schiff base in solution. *The Journal of Physical Chemistry*, 100(47):18586–18591, 1996.
- [29] Goran Zgrablic, Anna Maria Novello, and Fulvio Parmigiani. Population branching in the conical intersection of the retinal chromophore revealed by multipulse ultrafast optical spectroscopy. *Journal of the American Chemical Society*, 134(2):955–961, 2012.
- [30] Jörg Tittor and Dieter Oesterhelt. The quantum yield of bacteriorhodopsin. *FEBS letters*, 263(2):269–273, 1990.
- [31] Hoi Ling Luk, Nihar Bhattacharyya, Fabio Montisci, James M Morrow, Federico Melaccio, Akimori Wada, Mudi Sheves, Francesca Fanelli, Belinda SW Chang, and Massimo Olivucci. Modulation of thermal noise and spectral sensitivity in lake baikal cottoid fish rhodopsins. *Scientific Reports*, 6(1):38425, 2016.
- [32] Ying Guo, Sivakumar Sekharan, Jian Liu, Victor S Batista, John C Tully, and Elsa CY Yan. Unusual kinetics of thermal decay of dim-light photoreceptors in vertebrate vision. *Proceedings of the National Academy of Sciences*, 111(29):10438–10443, 2014.

- [33] JK Bowmaker, VI Govardovskii, SA Shukolyukov, J LV Zueva, DM Hunt, VG Sideleva, and OG Smirnova. Visual pigments and the photic environment: the cottoid fish of lake baikal. *Vision research*, 34(5):591–605, 1994.
- [34] Zuzana Musilova, Fabio Cortesi, Michael Matschiner, Wayne IL Davies, Jagdish Suresh Patel, Sara M Stieb, Fanny de Busserolles, Martin Malmstrøm, Ole K Tørresen, Celeste J Brown, et al. Vision using multiple distinct rod opsins in deep-sea fishes. *Science*, 364(6440):588–592, 2019.
- [35] Wenjing Wang, James H Geiger, and Babak Borhan. The photochemical determinants of color vision: revealing how opsins tune their chromophore’s absorption wavelength. *Bioessays*, 36(1):65–74, 2014.
- [36] Ramkumar Rajamani, Yen-Lin Lin, and Jiali Gao. The opsin shift and mechanism of spectral tuning in rhodopsin. *Journal of computational chemistry*, 32(5):854–865, 2011.
- [37] Kota Katayama, Sivakumar Sekharan, and Yuki Sudo. Color tuning in retinylidene proteins. *Optogenetics: Light-sensing proteins and their applications*, pages 89–107, 2015.
- [38] Michael Hoffmann, Marius Wanko, Paul Strodel, Peter H König, Thomas Frauenheim, Klaus Schulten, Walter Thiel, Emad Tajkhorshid, and Marcus Elstner. Color tuning in rhodopsins: the mechanism for the spectral shift between bacteriorhodopsin and sensory rhodopsin ii. *Journal of the American Chemical Society*, 128(33):10808–10818, 2006.
- [39] Mikhail N Ryazantsev, Ahmet Altun, and Keiji Morokuma. Color tuning in rhodopsins: the origin of the spectral shift between the chloride-bound and anion-free forms of halorhodopsin. *Journal of the American Chemical Society*, 134(12):5520–5523, 2012.
- [40] Barry Honig, Uri Dinur, Koji Nakanishi, Valeria Balogh-Nair, Mary Ann Gawinowicz, Maria Arnaboldi, and Michael G Motto. An external point-charge model for wavelength regulation in visual pigments. *Journal of the American Chemical Society*, 101(23):7084–7086, 1979.
- [41] Wenjing Wang, Zahra Nossoni, Tetyana Berbasova, Camille T Watson, Ipek Yapici, Kin Sing Stephen Lee, Chrysoula Vasileiou, James H Geiger, and Babak Borhan. Tuning the electronic absorption of protein-embedded all-trans-retinal. *Science*, 338(6112):1340–1343, 2012.
- [42] C Schnedermann, X Yang, M Liebel, KM Spillane, J Lugtenburg, I Fernández, A Valentini, I Schapiro, Massimo Olivucci, P Kukura, et al. Evidence for a vibrational phase-dependent isotope effect on the photochemistry of vision. *Nature chemistry*, 10(4):449–455, 2018.

- [43] R Scott McIsaac, Martin KM Engqvist, Timothy Wannier, Adam Z Rosenthal, Lukas Herwig, Nicholas C Flytzanis, Eleonora S Imasheva, Janos K Lanyi, Sergei P Balashov, Viviana Gradinaru, et al. Directed evolution of a far-red fluorescent rhodopsin. *Proceedings of the National Academy of Sciences*, 111(36):13034–13039, 2014.
- [44] Riccardo Palombo, Leonardo Barneschi, Laura Pedraza-González, Daniele Padula, Igor Schapiro, and Massimo Olivucci. Retinal chromophore charge delocalization and confinement explain the extreme photophysics of neorhodopsin. *Nature Communications*, 13(1):6652, 2022.
- [45] Zhenzhong Yu and Reinhard Fischer. Light sensing and responses in fungi. *Nature Reviews Microbiology*, 17(1):25–36, 2019.
- [46] Elena G Govorunova, Oleg A Sineshchekov, Hai Li, Yumei Wang, Leonid S Brown, and John L Spudich. Rubyacrs, nonalgal anion channelrhodopsins with highly red-shifted absorption. *Proceedings of the National Academy of Sciences*, 117(37):22833–22840, 2020.
- [47] Luis Javier Galindo, David S Milner, Suely Lopes Gomes, and Thomas A Richards. A light-sensing system in the common ancestor of the fungi. *Current Biology*, 32(14):3146–3153, 2022.
- [48] Tadeusz Andruniów and Massimo Olivucci. *QM/MM studies of light-responsive biological systems*. Springer, 2021.
- [49] AG Doukas, MR Junnarkar, RR Alfano, RH Callender, Toshiaki Kakitani, and Barry Honig. Fluorescence quantum yield of visual pigments: evidence for subpicosecond isomerization rates. *Proceedings of the National Academy of Sciences*, 81(15):4790–4794, 1984.
- [50] Joel M Kralj, Adam D Douglass, Daniel R Hochbaum, Dougal Maclaurin, and Adam E Cohen. Optical recording of action potentials in mammalian neurons using a microbial rhodopsin. *Nature methods*, 9(1):90–95, 2012.
- [51] T Kouyama, K Kinoshita, and A Ikegami. Excited-state dynamics of bacteriorhodopsin. *Biophysical journal*, 47(1):43–54, 1985.
- [52] Martin KM Engqvist, R Scott McIsaac, Peter Dollinger, Nicholas C Flytzanis, Michael Abrams, Stanford Schor, and Frances H Arnold. Directed evolution of *gloeobacter violaceus* rhodopsin spectral properties. *Journal of molecular biology*, 427(1):205–220, 2015.
- [53] Benjamin R Rost, Jonas Wietek, Ofer Yizhar, and Dietmar Schmitz. Optogenetics at the presynapse. *Nature neuroscience*, 25(8):984–998, 2022.

- [54] Yongxian Xu, Peng Zou, and Adam E Cohen. Voltage imaging with genetically encoded indicators. *Current Opinion in Chemical Biology*, 39:1–10, 2017.
- [55] Shoko Hososhima, Ritsu Mizutori, Rei Abe-Yoshizumi, Andrey Rozenberg, Shunta Shigemura, Alina Pushkarev, Masae Konno, Kota Katayama, Keiichi Inoue, Satoshi P Tsunoda, et al. Proton-transporting heliorhodopsins from marine giant viruses. *Elife*, 11:e78416, 2022.
- [56] Jin-gon Shim, Shin-Gyu Cho, Se-Hwan Kim, Kimleng Chuon, Seanghun Meas, Ahreum Choi, and Kwang-Hwan Jung. Heliorhodopsin helps photolyase to enhance the dna repair capacity. *Microbiology Spectrum*, 10(6):e02215–22, 2022.
- [57] José Flores-Uribe, Gur Hevroni, Rohit Ghai, Alina Pushkarev, Keiichi Inoue, Hideki Kandori, and Oded Bèjà. Heliorhodopsins are absent in diderm (gram-negative) bacteria: Some thoughts and possible implications for activity. *Environmental microbiology reports*, 11(3):419–424, 2019.
- [58] Kirill Kovalev, D Volkov, Roman Astashkin, Alexey Alekseev, Ivan Gushchin, Jose M Haro-Moreno, Igor Chizhov, Sergey Siletsky, M Mamedov, A Rogachev, et al. High-resolution structural insights into the heliorhodopsin family. *Proceedings of the National Academy of Sciences*, 117(8):4131–4141, 2020.
- [59] Frank Jensen. *Introduction to computational chemistry*. John wiley & sons, 2017.
- [60] Attila Szabo and Neil S Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.
- [61] Basile FE Curchod and Todd J Martínez. Ab initio nonadiabatic quantum molecular dynamics. *Chemical reviews*, 118(7):3305–3336, 2018.
- [62] Björn O Roos, Peter R Taylor, and Per EM Sigbahn. A complete active space scf method (casscf) using a density matrix formulated super-ci approach. *Chemical Physics*, 48(2):157–173, 1980.
- [63] Per-Åke Malmqvist and Björn O Roos. The casscf state interaction method. *Chemical physics letters*, 155(2):189–194, 1989.
- [64] Kerstin Andersson, Per Aake Malmqvist, Bjoern O Roos, Andrzej J Sadlej, and Krzysztof Wolinski. Second-order perturbation theory with a casscf reference function. *Journal of Physical Chemistry*, 94(14):5483–5488, 1990.
- [65] James Finley, Per-Åke Malmqvist, Björn O Roos, and Luis Serrano-Andrés. The multi-state caspt2 method. *Chemical physics letters*, 288(2-4):299–306, 1998.

- [66] Giovanni Ghigo, Björn O Roos, and Per-Åke Malmqvist. A modified definition of the zeroth-order hamiltonian in multiconfigurational perturbation theory (caspt2). *Chemical physics letters*, 396(1-3):142–149, 2004.
- [67] S Gozem, Federico Melaccio, HL Luk, Silvia Rinaldi, and Massimo Olivucci. Learning from photobiology how to design molecular devices using a computer. *Chemical Society Reviews*, 43(12):4019–4036, 2014.
- [68] David A Case, Thomas E Cheatham III, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M Merz Jr, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J Woods. The amber biomolecular simulation programs. *Journal of computational chemistry*, 26(16):1668–1688, 2005.
- [69] Arieh Warshel and Michael Levitt. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology*, 103(2):227–249, 1976.
- [70] Martin J Field, Paul A Bash, and Martin Karplus. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of computational chemistry*, 11(6):700–733, 1990.
- [71] U Chandra Singh and Peter A Kollman. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the $\text{ch}_3\text{cl} + \text{cl}^-$ exchange reaction and gas phase protonation of polyethers. *Journal of Computational Chemistry*, 7(6):718–730, 1986.
- [72] Federico Melaccio, María del Carmen Marín, Alessio Valentini, Fabio Montisci, Silvia Rinaldi, Marco Cherubini, Xuchun Yang, Yoshitaka Kato, Michael Stenrup, Yoelvis Orozco-Gonzalez, et al. Toward automatic rhodopsin modeling as a tool for high-throughput computational photobiology. *Journal of Chemical Theory and Computation*, 12(12):6020–6034, 2016.
- [73] Laura Pedraza-González, Luca De Vico, Maria del Carmen Marín, Francesca Fanelli, and Massimo Olivucci. a-arm: automatic rhodopsin modeling with chromophore cavity generation, ionization state selection, and external counterion placement. *Journal of chemical theory and computation*, 15(5):3134–3152, 2019.
- [74] Samira Gholami, Laura Pedraza-González, Xuchun Yang, Alexander A Granovsky, Ilya N Ioffe, and Massimo Olivucci. Multistate multiconfiguration quantum chemical computation of the two-photon absorption spectra of bovine rhodopsin. *The journal of physical chemistry letters*, 10(20):6293–6300, 2019.
- [75] Yuta Nakajima, Laura Pedraza-González, Leonardo Barneschi, Keiichi Inoue, Massimo Olivucci, and Hideki Kandori. Pro219 is an electrostatic color determinant in the light-driven sodium pump kr2. *Communications Biology*, 4(1):1185, 2021.

- [76] Maria-Andrea Mroginski, Suliman Adam, Gil S Amoyal, Avishai Barnoy, Ana-Nicoleta Bondar, Veniamin A Borin, Jonathan R Church, Tatiana Domratcheva, Bernd Ensing, Francesca Fanelli, et al. Frontiers in multiscale modeling of photoreceptor proteins. *Photochemistry and photobiology*, 97(2):243–269, 2021.
- [77] Francesco Aquilante, Jochen Autschbach, Alberto Baiardi, Stefano Battaglia, Veniamin A Borin, Liviu F Chibotaru, Irene Conti, Luca De Vico, Mickaël Delcey, Nicolas Ferré, et al. Modern quantum chemistry with [open] molcas. *The Journal of chemical physics*, 152(21), 2020.
- [78] Peter Schmidtke, Axel Bidon-Chanal, F Javier Luque, and Xavier Barril. Mdpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics*, 27(23):3276–3285, 2011.
- [79] Chresten R Søndergaard, Mats HM Olsson, Michał Rostkowski, and Jan H Jensen. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p k a values. *Journal of chemical theory and computation*, 7(7):2284–2295, 2011.
- [80] Benjamin Webb and Andrej Sali. Comparative protein structure modeling using modeller. *Current protocols in bioinformatics*, 54(1):5–6, 2016.
- [81] Li Zhang and Jan Hermans. Hydrophilicity of cavities in proteins. *Proteins: Structure, Function, and Bioinformatics*, 24(4):433–438, 1996.
- [82] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R Shirts, Jeremy C Smith, Peter M Kasson, David Van Der Spoel, et al. Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 2013.
- [83] Gunnar Karlström, Roland Lindh, Per-Åke Malmqvist, Björn O Roos, Ulf Ryde, Valera Veryazov, Per-Olof Widmark, Maurizio Cossi, Bernd Schimmelpfennig, Pavel Neogradý, et al. Molcas: a program package for computational chemistry. *Computational Materials Science*, 28(2):222–239, 2003.
- [84] Jay W Ponder et al. Tinker: Software tools for molecular design. *Washington University School of Medicine, Saint Louis, MO*, 3, 2004.
- [85] John C Tully. Molecular dynamics with electronic transitions. *The Journal of Chemical Physics*, 93(2):1061–1071, 1990.
- [86] Xuchun Yang, Madushanka Manathunga, Samer Gozem, Jérémie Léonard, Tadeusz Andruniów, and Massimo Olivucci. Quantum–classical simulations of rhodopsin reveal excited-state population splitting and its effects on quantum efficiency. *Nature Chemistry*, 14(4):441–449, 2022.

- [87] Alessandro Cembran, Fernando Bernardi, Massimo Olivucci, and Marco Garavelli. Counterion controlled photoisomerization of retinal chromophore models: a computational investigation. *Journal of the American Chemical Society*, 126(49):16018–16037, 2004.
- [88] Mohsen MT El-Tahawy, Artur Nenov, and Marco Garavelli. Photoelectrochromism in the retinal protonated schiff base chromophore: photoisomerization speed and selectivity under a homogeneous electric field at different operational regimes. *Journal of chemical theory and computation*, 12(9):4460–4475, 2016.
- [89] Mohsen MT El-Tahawy, Irene Conti, Matteo Bonfanti, Artur Nenov, and Marco Garavelli. Tailoring spectral and photochemical properties of bioinspired retinal mimics by in silico engineering. *Angewandte Chemie International Edition*, 59(46):20619–20627, 2020.
- [90] Timothy Y James, Peter M Letcher, Joyce E Longcore, Sharon E Mozley-Standridge, David Porter, Martha J Powell, Gareth W Griffith, and Rytas Vilgalys. A molecular phylogeny of the flagellated fungi (chytridiomycota) and description of a new phylum (blastocladiomycota). *Mycologia*, 98(6):860–871, 2006.
- [91] Chi-Yun Lin, Matthew G Romei, Luke M Oltrogge, Irimpan I Mathews, and Steven G Boxer. Unified model for photophysical and electro-optical properties of green fluorescent proteins. *Journal of the American Chemical Society*, 141(38):15250–15265, 2019.
- [92] Chi-Yun Lin, Matthew G Romei, Irimpan I Mathews, and Steven G Boxer. Energetic basis and design of enzyme function demonstrated using gfp, an excited-state enzyme. *Journal of the American Chemical Society*, 144(9):3968–3978, 2022.
- [93] Daniel R Hochbaum, Yongxin Zhao, Samouil L Farhi, Nathan Klapoetke, Christopher A Werley, Vikrant Kapoor, Peng Zou, Joel M Kralj, Dougal Maclaurin, Niklas Smedemark-Margulies, et al. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nature methods*, 11(8):825–833, 2014.
- [94] Kiryl D Piatkevich, Erica E Jung, Christoph Straub, Changyang Linghu, Demian Park, Ho-Jun Suk, Daniel R Hochbaum, Daniel Goodwin, Eftychios Pnevmatikakis, Nikita Pak, et al. A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nature chemical biology*, 14(4):352–360, 2018.
- [95] Andrej Šali and Tom L Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*, 234(3):779–815, 1993.
- [96] Laura Pedraza-González, Leonardo Barneschi, Michał Marszałek, Daniele Padula, Luca De Vico, and Massimo Olivucci. Automated qm/mm screening of rhodopsin variants with enhanced fluorescence. *Journal of chemical theory and computation*, 19(1):293–310, 2022.

- [97] Laura Pedraza-González, María del Carmen Marín, Luca De Vico, Xuchun Yang, and Massimo Olivucci. On the automatic construction of qm/mm models for biological photoreceptors: rhodopsins as model systems. *QM/MM studies of light-responsive biological systems*, pages 1–75, 2021.
- [98] Samer Gozem, Igor Schapiro, Nicolas Ferré, and Massimo Olivucci. The molecular mechanism of thermal noise in rod photoreceptors. *Science*, 337(6099):1225–1228, 2012.
- [99] Piero Altoè, Alessandro Cembran, Massimo Olivucci, and Marco Garavelli. Aborted double bicycle-pedal isomerization with hydrogen bond breaking is the primary event of bacteriorhodopsin proton pumping. *Proceedings of the National Academy of Sciences*, 107(47):20172–20177, 2010.
- [100] Ruibin Liang, Jimmy K Yu, Jan Meisner, Fang Liu, and Todd J Martinez. Electrostatic control of photoisomerization in channelrhodopsin 2. *Journal of the American Chemical Society*, 143(14):5425–5437, 2021.
- [101] Elena N Laricheva, Samer Gozem, Silvia Rinaldi, Federico Melaccio, Alessio Valentini, and Massimo Olivucci. Origin of fluorescence in 11-cis locked bovine rhodopsin. *Journal of Chemical Theory and Computation*, 8(8):2559–2563, 2012.
- [102] Masahiro Sugiura, Kazuki Ishikawa, Kota Katayama, Yuji Sumii, Rei Abe-Yoshizumi, Satoshi P Tsunoda, Yuji Furutani, Norio Shibata, Leonid S Brown, and Hideki Kandori. Unusual photoisomerization pathway in a near-infrared light absorbing enzymehodopsin. *The Journal of Physical Chemistry Letters*, 13(40):9539–9543, 2022.
- [103] T Althaus, W Einfeld, R Lohrmann, and M Stockburger. Application of raman spectroscopy to retinal proteins. *Israel journal of chemistry*, 35(3-4):227–251, 1995.
- [104] Steven O Smith, Johan Lugtenburg, and Richard A Mathies. Determination of retinal chromophore structure in bacteriorhodopsin with resonance raman spectroscopy. *The Journal of membrane biology*, 85:95–109, 1985.
- [105] Steven O Smith, Mark S Braiman, Anne B Myers, Johannes A Pardoën, Jacques ML Courtin, Chris Winkel, Johan Lugtenburg, and Richard A Mathies. Vibrational analysis of the all-trans-retinal chromophore in light-adapted bacteriorhodopsin. *Journal of the American Chemical Society*, 109(10):3108–3125, 1987.
- [106] Miroslav Kloz, Jörn Weißenborn, Tomáš Polívka, Harry A Frank, and John TM Kennis. Spectral watermarking in femtosecond stimulated raman spectroscopy: resolving the nature of the carotenoid s* state. *Physical Chemistry Chemical Physics*, 18(21):14619–14628, 2016.
- [107] Sara Bruun. Resonance raman spectroscopy on microbial rhodopsins. 2013.

- [108] Kousuke Kajimoto, Takashi Kikukawa, Hiroki Nakashima, Haruki Yamaryo, Yuta Saito, Tomotsumi Fujisawa, Makoto Demura, and Masashi Unno. Transient resonance raman spectroscopy of a light-driven sodium-ion-pump rhodopsin from *indibacter alkaliphilus*. *The journal of physical chemistry B*, 121(17):4431–4437, 2017.
- [109] Sylvia IE Touw, Huub JM de Groot, and Francesco Buda. Ab initio modeling of the spatial, electronic, and vibrational structure of schiff base models for visual photoreceptors. *The Journal of Physical Chemistry B*, 108(35):13560–13572, 2004.
- [110] AG Doukas, B Aton, RH Callender, and TG Ebrey. Resonance raman studies of bovine metarhodopsin i and metarhodopsin ii. *Biochemistry*, 17(12):2430–2435, 1978.
- [111] David Buhrke and Peter Hildebrandt. Probing structure and reaction dynamics of proteins using time-resolved resonance raman spectroscopy. *Chemical reviews*, 120(7):3577–3630, 2019.
- [112] Andreas C Albrecht. On the theory of raman intensities. *The Journal of chemical physics*, 34(5):1476–1484, 1961.
- [113] Seth Olsen and Ross H McKenzie. Conical intersections, charge localization, and photoisomerization pathway selection in a minimal model of a degenerate monomethine dye. *The Journal of chemical physics*, 131(23), 2009.
- [114] Leonardo Barneschi, Emanuele Marsili, Laura Pedraza-González, Daniele Padula, Luca De Vico, Danil Kaliakin, Alejandro Blanco-González, Nicolas Ferré, Miquel Huix-Rotllant, Michael Filatov, et al. On the fluorescence enhancement of arch neuronal optogenetic reporters. *Nature Communications*, 13(1):6432, 2022.
- [115] Edward S Boyden, Feng Zhang, Ernst Bamberg, Georg Nagel, and Karl Deisseroth. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature neuroscience*, 8(9):1263–1268, 2005.
- [116] Feng Zhang, Li-Ping Wang, Martin Brauner, Jana F Liewald, Kenneth Kay, Natalie Watzke, Phillip G Wood, Ernst Bamberg, Georg Nagel, Alexander Gottschalk, et al. Multimodal fast optical interrogation of neural circuitry. *Nature*, 446(7136):633–639, 2007.
- [117] Paul A Davison, Weiming Tu, Jiabao Xu, Simona Della Valle, Ian P Thompson, C Neil Hunter, and Wei E Huang. Engineering a rhodopsin-based photo-electrosynthetic system in bacteria for co2 fixation. *ACS Synthetic Biology*, 11(11):3805–3816, 2022.
- [118] Oded Béja, L Aravind, Eugene V Koonin, Marcelino T Suzuki, Andrew Hadd, Linh P Nguyen, Stevan B Jovanovich, Christian M Gates, Robert A Feldman, John L Spudich, et al. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*, 289(5486):1902–1906, 2000.

- [119] Laura Gómez-Consarnau, John A Raven, Naomi M Levine, Lynda S Cutter, Deli Wang, Brian Seegers, Javier Arístegui, Jed A Fuhrman, Josep M Gasol, and Sergio A Sañudo-Wilhelmy. Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Science advances*, 5(8):eaaw8855, 2019.
- [120] Shino Inukai, Kota Katayama, Mitsumasa Koyanagi, Akihisa Terakita, and Hideki Kandori. Investigating the mechanism of photoisomerization in jellyfish rhodopsin with the counterion at an atypical position. *Journal of Biological Chemistry*, 299(6), 2023.
- [121] Mohsen MT El-Tahawy, Artur Nenov, Oliver Weingart, Massimo Olivucci, and Marco Garavelli. Relationship between excited state lifetime and isomerization quantum yield in animal rhodopsins: Beyond the one-dimensional landau–zener model. *The journal of physical chemistry letters*, 9(12):3315–3322, 2018.
- [122] Taku Mizukami, Hideki Kandori, Yoshinori Shichida, Arh-Hwang Chen, Fadila Derguini, Charles G Caldwell, CF Biffe, Koji Nakanishi, and Toru Yoshizawa. Photoisomerization mechanism of the rhodopsin chromophore: picosecond photolysis of pigment containing 11-cis-locked eight-membered ring retinal. *Proceedings of the National Academy of Sciences*, 90(9):4072–4076, 1993.
- [123] Keiichi Inoue. Photochemistry of the retinal chromophore in microbial rhodopsins. *The Journal of Physical Chemistry B*, 127(43):9215–9222, 2023.
- [124] Chun-Fu Chang, Hikaru Kuramochi, Manish Singh, Rei Abe-Yoshizumi, Tatsuya Tsukuda, Hideki Kandori, and Tahei Tahara. A unified view on varied ultrafast dynamics of the primary process in microbial rhodopsins. *Angewandte Chemie International Edition*, 61(2):e202111930, 2022.
- [125] R Govindjee, SP Balashov, and TG Ebrey. Quantum efficiency of the photochemical cycle of bacteriorhodopsin. *Biophysical journal*, 58(3):597–608, 1990.
- [126] Olga A Smitienko, Tatiana B Feldman, Lada E Petrovskaya, Oksana V Nekrasova, Marina A Yakovleva, Ivan V Shelaev, Fedor E Gostev, Dmitry A Cherepanov, Irina B Kolchugina, Dmitry A Dolgikh, et al. Comparative femtosecond spectroscopy of primary photoreactions of *exiguobacterium sibiricum* rhodopsin and *halobacterium salinarum* bacteriorhodopsin. *The Journal of Physical Chemistry B*, 125(4):995–1008, 2021.
- [127] Tatiana B Feldman, Olga A Smitienko, Ivan V Shelaev, Fedor E Gostev, Oksana V Nekrasova, Dmitriy A Dolgikh, Victor A Nadtochenko, Mikhail P Kirpichnikov, and Mikhail A Ostrovsky. Femtosecond spectroscopic study of photochromic reactions of bacteriorhodopsin and visual rhodopsin. *Journal of Photochemistry and Photobiology B: Biology*, 164:296–305, 2016.

- [128] Chun-Fu Chang, Masae Konno, Keiichi Inoue, and Tahei Tahara. Effects of the unique chromophore–protein interactions on the primary photoreaction of schizorhodopsin. *The Journal of Physical Chemistry Letters*, 14(31):7083–7091, 2023.
- [129] Shinya Tahara, Manish Singh, Hikaru Kuramochi, Wataru Shihoya, Keiichi Inoue, Osamu Nureki, Oded Bèjà, Yasuhisa Mizutani, Hideki Kandori, and Tahei Tahara. Ultrafast dynamics of heliorhodopsins. *The Journal of Physical Chemistry B*, 123(11):2507–2512, 2019.
- [130] Tatsuki Tanaka, Manish Singh, Wataru Shihoya, Keitaro Yamashita, Hideki Kandori, and Osamu Nureki. Structural basis for unique color tuning mechanism in heliorhodopsin. *Biochemical and Biophysical Research Communications*, 533(3):262–267, 2020.
- [131] RW Schoenlein, LA Peteanu, RA Mathies, and CV Shank. The first step in vision: femtosecond isomerization of rhodopsin. *Science*, 254(5030):412–415, 1991.
- [132] Qing Wang, Robert W Schoenlein, Linda A Peteanu, Richard A Mathies, and Charles V Shank. Vibrationally coherent photochemistry in the femtosecond primary event of vision. *Science*, 266(5184):422–424, 1994.
- [133] Dario Polli, Oliver Weingart, Daniele Brida, Emiliano Poli, Margherita Maiuri, Katelyn M Spillane, Andrea Bottoni, Philipp Kukura, Richard A Mathies, Giulio Cerullo, et al. Wavepacket splitting and two-pathway deactivation in the photoexcited visual pigment isorhodopsin. *Angewandte Chemie International Edition*, 53(9):2504–2507, 2014.
- [134] Luis Manuel Frutos, Tadeusz Andruniów, Fabrizio Santoro, Nicolas Ferré, and Massimo Olivucci. Tracking the excited-state time evolution of the visual pigment with multi-configurational quantum chemistry. *Proceedings of the National Academy of Sciences*, 104(19):7764–7769, 2007.
- [135] Arie Warshel. Bicycle-pedal model for the first step in the vision process. *Nature*, 260(5553):679–683, 1976.
- [136] Hoi Ling Luk, Federico Melaccio, Silvia Rinaldi, Samer Gozem, and Massimo Olivucci. Molecular bases for the selection of the chromophore of animal rhodopsins. *Proceedings of the National Academy of Sciences*, 112(50):15297–15302, 2015.
- [137] Marius Schmidt and Emina A Stojković. Earliest molecular events of vision revealed, 2023.
- [138] Curt Wittig. The landau- zener formula. *The Journal of Physical Chemistry B*, 109(17):8428–8430, 2005.

- [139] Stepan S Batsanov. Van der waals radii of elements. *Inorganic materials*, 37(9):871–885, 2001.
- [140] D Oesterhelt, P Hegemann, and J Tittor. The photocycle of the chloride pump halorhodopsin. ii: Quantum yields and a kinetic model. *The EMBO journal*, 4(9):2351–2356, 1985.
- [141] T Arlt, S Schmidt, W Zinth, U Haupts, and D Oesterhelt. The initial reaction dynamics of the light-driven chloride pump halorhodopsin. *Chemical physics letters*, 241(5-6):559–565, 1995.
- [142] Martin O Lenz, Robert Huber, Bernhard Schmidt, Peter Gilch, Rolf Kalmbach, Martin Engelhard, and Josef Wachtveitl. First steps of retinal photoisomerization in proteorhodopsin. *Biophysical journal*, 91(1):255–262, 2006.
- [143] Karsten Neumann, Mirka-Kristin Verhoefen, Ingrid Weber, Clemens Glaubitz, and Josef Wachtveitl. Initial reaction dynamics of proteorhodopsin observed by femtosecond infrared and visible spectroscopy. *Biophysical journal*, 94(12):4796–4807, 2008.
- [144] Yuki Sudo, Kunio Ihara, Shiori Kobayashi, Daisuke Suzuki, Hiroki Irieda, Takashi Kikukawa, Hideki Kandori, and Michio Homma. A microbial rhodopsin with a unique retinal composition shows both sensory rhodopsin ii and bacteriorhodopsin-like properties. *Journal of Biological Chemistry*, 286(8):5967–5976, 2011.
- [145] Yusaku Hontani, Marco Marazzi, Katja Stehfest, Tilo Mathes, Ivo HM van Stokkum, Marcus Elstner, Peter Hegemann, and John TM Kennis. Reaction dynamics of the chimeric channelrhodopsin c1c2. *Scientific Reports*, 7(1):7217, 2017.
- [146] Shinya Tahara, Satoshi Takeuchi, Rei Abe-Yoshizumi, Keiichi Inoue, Hiroyuki Ohtani, Hideki Kandori, and Tahei Tahara. Origin of the reactive and nonreactive excited states in the primary reaction of rhodopsins: ph dependence of femtosecond absorption of light-driven sodium ion pump rhodopsin kr2. *The Journal of Physical Chemistry B*, 122(18):4784–4792, 2018.