UNIVERSITÀ
DI SIENA
1240

DIPARTIMENTO DI BIOTECNOLOGIE, CHMICA E FARMACIA

DOTTORATO DI RICERCA IN CHEMICAL AND PHARMACEUTICAL SCIENCES

CICLO XXXV

COORDINATORE: PROF. Maurizio TADDEI

Computational studies of the mechanism of fluorescence enhancement and spectral tuning in arch neuronal optogenetic reporters

SETTORE SCIENTIFICO-DISCIPLINARE: CHIM/06

Leonardo BARNESCHI
Dottorando

Chiar.mo Prof. Massimo OLIVUCCI
Supervisore

ANNO ACCADEMICO 2019/2022

UNIVERSITÀ
DI SIENA
1240

DIPARTIMENTO DI BIOTECNOLOGIE, CHMICA E FARMACIA

DOTTORATO DI RICERCA IN CHEMICAL AND PHARMACEUTICAL SCIENCES

CICLO XXXV

Candidates's Declaration

I hereby declare that the work presented in the thesis entitled "**On the fluorescence enhancement of Arch neuronal optogenetic reporters**", 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 a period from October 2019 to November 2022 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/University.

Leonardo Barneschi

Leonardo Barneschi
Ph.D. candidate

This is to certify that the above statement made by the candidate is true to the best of our knowledge and belief.
Siena, Italy, December 2022.

Massimo Olivucci

Prof. Massimo OLIVUCCI
Supervisor

Abstract

Several biological systems have evolved structures that can interact with light and exploit the photons energy for their biological functions. In fact, thanks to specialized organic molecules spread across all life kingdoms, life forms use light to trigger and regulate an incredible variety of different activities, including vision, photosynthesis, circadian rhythm, bioluminescence and in general a wide spectrum of biochemical processes. The common thread to all these activities is that they are invariably initiated by a photochemical reaction. Among these specialized structures, rhodopsins, membrane proteins harbouring a covalently bound retinal chromophore, play a preeminent role. In fact, while the first member of the rhodopsin family identified in the 1930s was the visual pigment of the animal retina, it was then found that this class of light-triggered proteins populates the whole spectrum of living organisms, where it promotes an astounding variety of biochemical processes, ranging from vision to cellular metabolism and sensorial function. More recently, rhodopsin became especially important in Optogenetics, a collection of neurobiology techniques where photo-sensitive proteins are used to control and visualize neural activity. The imaging of networks of interacting neurons is a particularly challenging task, since common fluorescence microscopy techniques require bright fluorescent probes localized in the neuron membrane. Ideally, such probes would also display a red-shifted maximum absorption wavelength ($\lambda^{\mathrm{a}}_{\mathrm{max}}$), allowing to use excitation lasers with higher tissue penetration and reduced phototoxicity.

Since rhodopsins are light-responsive membrane proteins, they are ideal candidates to report on action potentials. Archaerhodopsin-3 (Arch3) a microbial rhodopsin from *Halorubrum Sodomense,* was the first proposed rhodopsin-based fluorescent reporter. However, Arch3 has a fluorescence quantum yield (FQY) of ca. $\sim 1 \cdot 10^{-4}$, which impairs the possibility to efficiently visualize the activity of neuronal populations. Throughout the years, several experiments of random and site-directed mutagenesis on Arch3 culminated in the discovery of variants with increased FQY, such as the Archers, the Archons, Arch5 and Arch7. These variants feature a brighter fluorescence enabling applications in imaging of acute brain slices, but also in living mammals and invertebrates. However, the FQY value are still in the range 10^{-3}-10^{-2} and therefore not yet as bright as desirable.

Therefore, it would be highly beneficial to develop computational tools for the high-throughput rational-design of rhodopsins with desired photochemical and/or photophysical properties.

In my thesis, we provide the necessary theoretical foundations to envision the development of such tools. By constructing multiconfigurational quantum chemistry (MCQC) model of Arch3 and six of is variants, we establish a theory connecting the amino acid sequence to the increase of fluorescent brightness. We show that the observed experimental FQY trend correlates with the decrease in energy difference between the planar fluorescent emitting state and a newly characterized exotic diradical intermediate intercepted by a nearby photoisomerization channel. Investigating the molecular-level factors modulating this critical quantity, we show that the electronic structure of the retinal chromophore at the two minima is substantially different and that this is reflected by their different charge distributions. This is important because it indicates that a variation in the protein electrostatic potential that simultaneously stabilize the fluorescent state and destabilize the region of the decay channel would dramatically increase the FQY, suggesting an ideal target for microbial rhodopsins fluorescence engineering.

Since all the reported Arch3 variants with increased FQY display also red-shifted absorption maximum ($\lambda^{\mathrm{a}}_{\mathrm{max}}$), another appealing feature in *Optogenetics*, we used Arch3 as template to develop a new computational tool which allowed us to identify the critical opsin electrostatic variations that contribute to the spectral tuning. By designing an optimization procedure based on a variational protocol, we show that the "effective delocalization" of the counterion is the critical determinant of Arch3 $\lambda^{\mathrm{a}}_{\mathrm{max}}$. This theoretical framework agrees with the experimental findings, which heuristically identified in the counterion complex the most sensible replacements necessary to shift to the red the absorption of Arch3. Finally, using the same tool, we show that, in this family of rhodopsins, there is indeed a first-order relationship between $\lambda^{\mathrm{a}}_{\mathrm{max}}$ and FQY.

List of Figures

List of Tables

List of Acronyms and Abbreviations

ΔE^a_{S1-S0} : Vertical excitation energy of absorption
ΔE^f_{S1-S0} : Vertical excitation energy of emission
$\Delta E_{FS-TIDIR}$: Isomerization energy
λ^a_{max} : Maximum absorption wavelength
λ^f_{max} : Maximum emission wavelength
$1A_g$: Covalent character
$1B_u$: Charge Transfer character
$2A_g$: Diradical character
Arch1 : Archaerhodopsin-1
Arch2 : Archaerhodopsin-2
Arch3 : Archaerhodopsin-3
Arch5 : Archaerhodopsin-5 (=Arch3 D95E/T99C/V59A/P60L/P196S)
Arch7 : Archaerhodopsin-7 (=Arch3 D95E/T99C/V59A/P60L/P196S/D222S/A225C)
Archon2 : (=Arch3 T56P/P60S/T80P/D95H/T99S/T116I/F161V/T183I/L197I/A225C)
ARM : Automatic Rhodopsin Modelling protocol
ASR : Anabaena sensory rhodopsin
BLA : Bond Length Alternation
BR : Bacteriorhodopsin
BO : Born-Oppenheimer
CASPT2 : Complete Active-Space second-order perturbation theory
CASSCF : Complete Active Space Self-Consistent field
ChR : Channelrhodopsin
ChR1 : Channelrhodopsin-1
ChR2 : Channelrhodopsin-2
CI : Configuration-Interaction
CISD : Configuration-Interaction single double
CoIn : Conical intersection
CT : charge-transfer
COV : covalent
DIR : diradical
DA : Dark-Adapted
DOPE : Discrete optimized protein energy
E^f_{S1} : photoisomerization barrier
ESL : Excited state lifetime
ESP : Electrostatic potential
FC : Franck-Condon
fOsc : Transition oscillator strength
FS : Fluorescent State
FQY : Fluorescence Quantum Yield
GEVI : Genetically Encoded Voltage Indicator
GFP : Green Fluorescent Protein
GR : Gloeobacter violaceus
HBN : Hydrogen-bond networks
hCRBPII : human cellular retinol-binding protein II
HF : Hartree-Fock

HLA : hydrogen link atom
HR : Halorhodopsin
KR2 : Krokinobacter rhodopsin 2
LA: Light-Adapted
MAE : mean absolute error
MC : Main Counterion
MCQC : Multi-configurational quantum chemistry
MD : Molecular Dynamics
MM : Molecular Mechanics
NeoR : Neorhodopsin
OS : extracellular surface
PDB : Protein Data Bank
PES : Potential Energy Surface
PR : Proteorhodopsin
PSB : Protonated Schiff-base
QC : Quantum Chemistry
QM : Quantum Mechanics
QM/MM : Quantum Mechanics/Molecular Mechanics
QuasAr1 : QuasAr1 (=Arch3 P60S/T80S/D95H/D106H/F161V)
QuasAr2 : QuasAr2 (=QuasAr1 H95Q)
QY : Photoisomerization quantum yield
Rh : Bovine rhodopsin
rPSB : Retinal protonated Schiff base
rPSBAT : Retinal protonated Schiff base all-trans
rSB : Retinal Schiff base
S_0 : Ground state
S_1 : First excited state
S_2 : Second excited state
SRI : sensory rhodopsin I
SRII : sensory rhodopsin II
TICT : twisted intermediate charge-transfer
TIDIR : twisted intermediate diradical
TISE : time independent Schrödinger equation
TDSE : time dependent Schrödinger equation
TMAE : trend mean absolute error
TS_{S1} : excited state transition state
WT : wild type
X^{f}_{S1} : excited state transition state position
XMS-CASPT2 : extended multi-state Complete Active-Space 2nd-order perturbation theory

List of Publications

Published:

- Barneschi, L., Marsili, E., Pedraza-González, L., Padula, D., De Vico, L., Kaliakin, D., Blanco-González, A., Ferré, N., Huix-Rotllant, M., Filatov, M. and Olivucci, M., 2022. On the fluorescence enhancement of arch neuronal optogenetic reporters. *Nature communications*, *13*(1), pp.1-9.

- Nakajima, Y., Pedraza-González, L., Barneschi, L., Inoue, K., Olivucci, M. and Kandori, H., 2021. Pro219 is an electrostatic color determinant in the light-driven sodium pump KR2. *Communications biology*, *4*(1), pp.1-15.

- Palombo, R., Barneschi, L., Pedraza-González, L., Padula, D., Schapiro, I. and Olivucci, M., 2022. Retinal chromophore charge delocalization and confinement explain the extreme photophysics of Neorhodopsin. *Nature communications*, *13*(1), pp.1-9.

- Pedraza-González, L., Barneschi, L., Padula, D., De Vico, L. and Olivucci, M., 2022. Evolution of the Automatic Rhodopsin Modeling (ARM) Protocol. *Topics in Current Chemistry*, *380*(3), pp.1-48.

Accepted:

- Pedraza-González, L ., Barneschi, L., *Marszałek*, M., Valentini, A., Padula, D., De Vico L., and Olivucci, M., 2022. Automated QM/MM Model Screening of Rhodopsin Variants Displaying Enhanced Fluorescence*. Journal of Chemical Theory and Computation.*

Contents

Chapter 1

Introduction

1.1 Applications of Natural and Engineered Rhodopsins: Optogenetics

The main topic of the present dissertation is the investigation of the photophysical and photochemical properties of rhodopsins which are currently used in the rapidly expanding field of Optogenetics[1–3]. Optogenetics is a collection of techniques based on the optical control of genetically encodable proteins, which may be used to target and monitor the electrical activity in living tissues, such as neural systems[4,5]. The conceptual basis for this technique were established in 1979 by Francis Crick, who speculated about the possibilty to use light to control the activity of a specific cellular population[6]. Before then, Bacteriorhodopsin (BR) was already found to be a visible light-induced proton pump[7], but it has not been until several years later that the use of microbial opsin based gene expression has started to be routinely used to target neurons and control their activity (alternatively turn on and off their action potentials)[8–10]. Nowadays, the optogenetical control has grown to such an extent that in neuroscience it is possible to deliver spatially and temporally resolved brain cells control in behaving mammals[11–13].

Optogenetics expresses its greatest capabilities by combining the potential of two classes of molecules, namely actuators and sensors. The actuators are optically controllable molecules which can influence the transmembrane voltage acting as ion-channels or ion pumps[11,14,15]. In other words, action potentials may be triggered or suppressed by using light of a specific wavelength. On the other hand, sensors are typically fluorescent molecules which senses the action potentials by modulating the light output (in terms of wavelength or brightness), depending on the electrical field experienced[16,17]. At least four of the molecules investigated for this dissertation, namely, Arch3, QuasAr1, QuasAr2 and Archon2, are currently used in Optogenetics as genetically encoded voltage indicators (GEVIs). In order to visualize neural dynamics and at the same time being able to deliver spatially and temporally resolved stimuli, it is highly desirable to identify actuators and sensors which are spectroscopically orthogonal (**Figure 1.1.1**)[18–20]. It is indeed important that the actuators are not triggered at the excitation wavelength used to provoke the fluorescence emission of the sensor[20]. The ideal sensor should then display high fluorescence quantum yield (FQY), fast kinetics (the fluorescence response time to voltage changes), high fluorescence sensibility to voltage changes, and excitation wavelength compatible with the actuators used. To this aim, scientists have recently started to look for rhodopsins variants showing red-shifted absorption/emission wavelength and high FQY. Indeed, when used to excite the GEVIs, light with longer wavelength is less phototoxic and penetrates more easily the tissues[21,22]. Among the rhodopsin based GEVIs reported thus far, the family of microbial rhodopsins is prominent; in particular, the wild type microbial rhodopsins Arch3 and, more recently, several of its artificially engineered variants discussed later in this chapter, have successfully been employed to visualize neural activity in brain slices and even in behaving mammals[23,24].

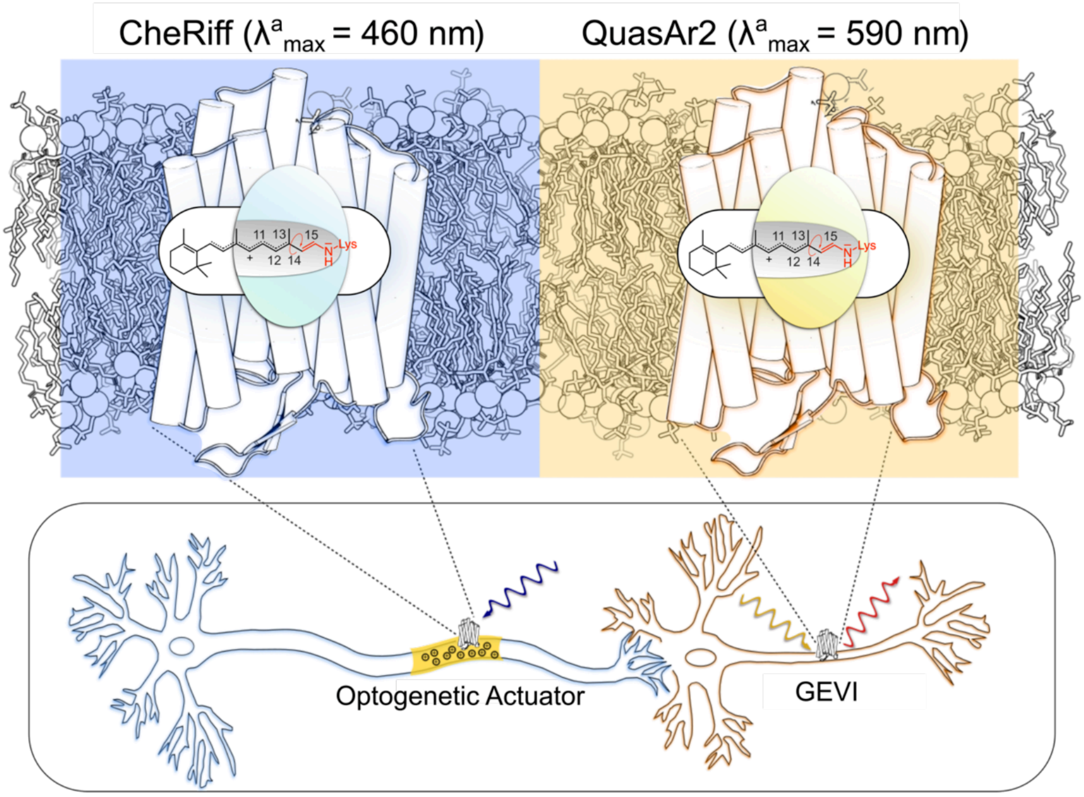


Figure 1.1.1 Pictorial representation of rhodopsins used as actuator (left) and sensor (right). Actuators may promote or suppress a neuronal action potential upon illumination with light of a specific wavelength. On the other hand, Sensors are usually fluorescent proteins whose fluorescent is highly sensitive in response to membrane voltage.

1.2 Rhodopsins, a versatile protein family

In order to exploit the energy of light, life forms to be provided with tools that are able to sense it and trigger responses to it. The proteins of the Rhodopsin family are exactly one of these instruments and their widespread diffusion in all of the different life kingdoms testifies their unique efficiency and versatility[25]. For instance, in vertebrates and invertebrates animals, rhodopsins are found as visual pigments, at the heart of the vision process. In prokaryotes, the light sensitivity of rhodopsins is exploited for an astounding variety of biological functions, which however, may be summarized in two major classes; (i) *photoenergy transducers* create electrochemical gradients across a membrane to store energy for their metabolism (e.g., light-driven proton pumps [H+], inward chloride transporters [Cl−][26], and outward sodium transporter [Na+]) and (ii) *photosensory receptors* may use photons for phototaxis, or photoadaptation, where light regulates gene expression or protein function[27,28].

Although the amino acid composition of rhodopsins can greatly vary among different organisms, the general architecture is conserved and is made of two fundamental building blocks; (i) a retinal chromophore covalently linked to a lysine residue and (ii) the opsin, organized in seven transmembrane α-helices with a cavity binding pocket to host the chromophore[29]. The retinal is the portion of the photoreceptor which is able to interact with photons of light and its binding to the opsin is realized via a Schiff base (rSB), usually protonated (rPSB), with the amino group of the Lysine side chain (see **Figure 1.2.1**). Interestingly, the reaction of double-bond isomerization of the retinal fragment is a necessary condition for all the functions displayed by rhodopsins in nature. The main contributor to this process is the chromophore's π framework, constituted by six conjugated double bonds and deeply affected by the local protein environment. It is not by chance, for instance, that the configuration of the retinal and the products of the light-induced photoisomerization are distinct in microbial and animal rhodopsins (see **Figure 1.2.1**). Hence, to get further insight into the origin of the different functions related with these structural differences, it is imperative to analyze the actual structure of the retinal in the rhodopsin protein, before and after the process of photoisomerization. In this regard, several efforts have made recently to develop experimental techniques that allow the elucidation of the crystallographic structure of rhodopsins in the unphotolyzed dark-adapted (DA) state and light-adapted

(LA) states (photo intermediates). As a result, in the Protein Data Bank (PDB)[30] a considerable number of X-ray structures for rhodopsins from different organisms is available, with microbial rhodopsins generally being more represented than the animal counterpart. However, crystallization of membrane proteins poses several important challenges with respect to soluble proteins[31]. For this reason, while a very well-resolved structure of wild-type Arch3 (PDB ID: 6GUX) has been reported recently Moraes and coworkers[32], none of the Arch3 variants investigated in this dissertation and currently applied in optogenetics are available. To generate resolved structures of rhodopsins that are still not attainable with the current experimental technologies, it is mandatory take advantage of computational methods that allow their in-silico design via comparative modelling, using as template rhodopsins similar in sequence and having available an X-ray structure. To this aim, in **Section 3.3**, we discuss the comparative modelling protocol which allowed us to simulate the proteins of the Arch set at the atomistic level, considering the possible limitations of employing comparative models rather than X-ray structures.

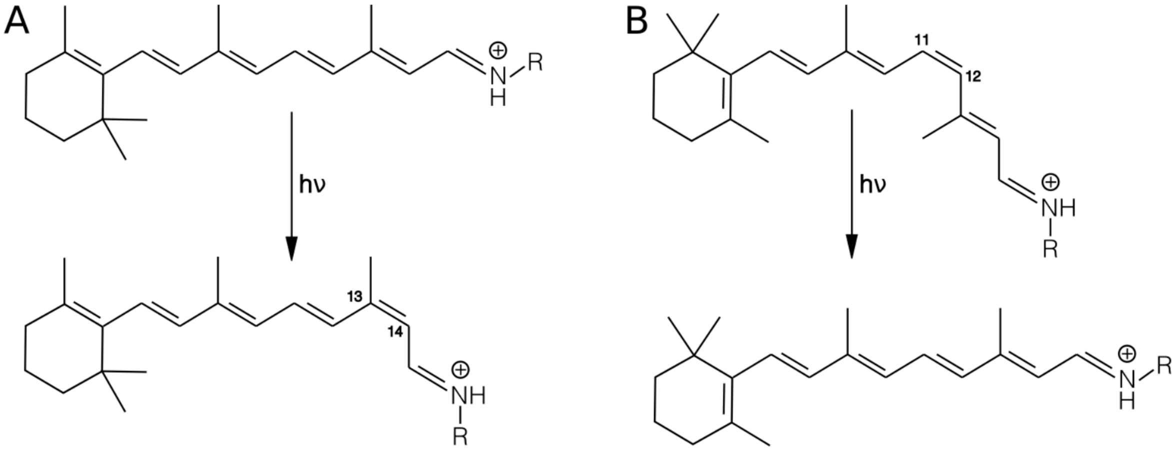


Figure 1.2.1 (A) Photoisomerization all-trans to 13-*cis* in microbial rhodopsins and (B) from 11-*cis* to all-*trans* in animal rhodopsins. Atom labels are given for the isomerizing double bonds.

Studying experimental or computer-modelled rhodopsin structures is imperative to provide insights into the mechanisms behind the tuning of reactivity, maximum absorption wavelength (λ^{a}_{max}), fluorescence quantum yield (FQY) occurring in natural or engineered rhodopsin. Indeed, a better understanding of these processes at the molecular level is the first step towards the prediction, enhancement or inhibition of specific functions (such as ion pumping for optogenetic actuators or fluorescence emission for GEVIs) useful for several technological applications of rhodopsins. In general, the scientific community is showing increasing interest in the development and implementation of computational tools in photochemistry and computational photobiology and their application in the rhodopsin field. Indeed, being able to provide theoretical insights into specific chemical and photochemical features is the basis for building more sophisticated methods with predictive capabilities (at least in a statistical sense), such as the a-ARM protocol[33,34] discussed in **Section 2.7**.

1.3 Microbial rhodopsins

As mentioned above, microorganisms have proven to be able to use light energy with incredible creativity. The first reported microbial rhodopsins were archeal rhodopsins, found in *Halobacterium salinarum*. Among them, halorhodopsin (HR) and bacteriorhodopsin (BR) have been widely studied since both of them exhibit vectorial ion pumping function, albeit with different specificity[35]; whereas HR acts as an inward chloride pump, BR acts as an outward proton pump[36]. In addition, a large number of microbial rhodopsins coming from other organisms have been discovered and characterized during the last few decades (see **Figure 1.3.1**)[37]. Meaningful examples are the Channelrhodopsins (ChRs)[8,38], which work as light- gated ion channels (i.e. ChR1 is selectively permeable to protons); the sensory rhodopsin I (SRI) and II (SRII)[28], where the phototaxis is regulated by visible light wavelength; or the eubacterial Anabaena sensory rhodopsin (ASR), reported to be a photosensory receptor that transmits the signal through a soluble transducer[39]. Most of the functions inherent to microbial rhodopsins can be translated into technological and medical applications, motivating an increase on both the in vivo and in silico studies of wild-type and mutants, generated by the substitution of one or more amino acids in the protein sequence.

In my thesis, I have mainly investigated a set of seven microbial rhodopsins, most of whom are either already used as GEVIs in Optogenetics or have shown great potential for future applications, one of them, Arch3, being also used to suppress neuronal action potentials or more recently to promote phototriggered apoptotic cell death in human HeLa cells[40]. These seven proteins, namely the wild types Archaerodopsin-2 (Arch2), Archaerhodopsin-3 (Arch3) and five artificially engineered Arch3 mutants, QuasAr1, QuasAr2, Archon2, Arch7 and Arch5, will be from now on referred to as the

"Arch set". In the following (**Section 1.4**), I will provide an overlook of the Arch set, discussing how the photochemical and photophysical properties of Arch3 were manipulated to obtain protein variants with useful applications in optogenetics.

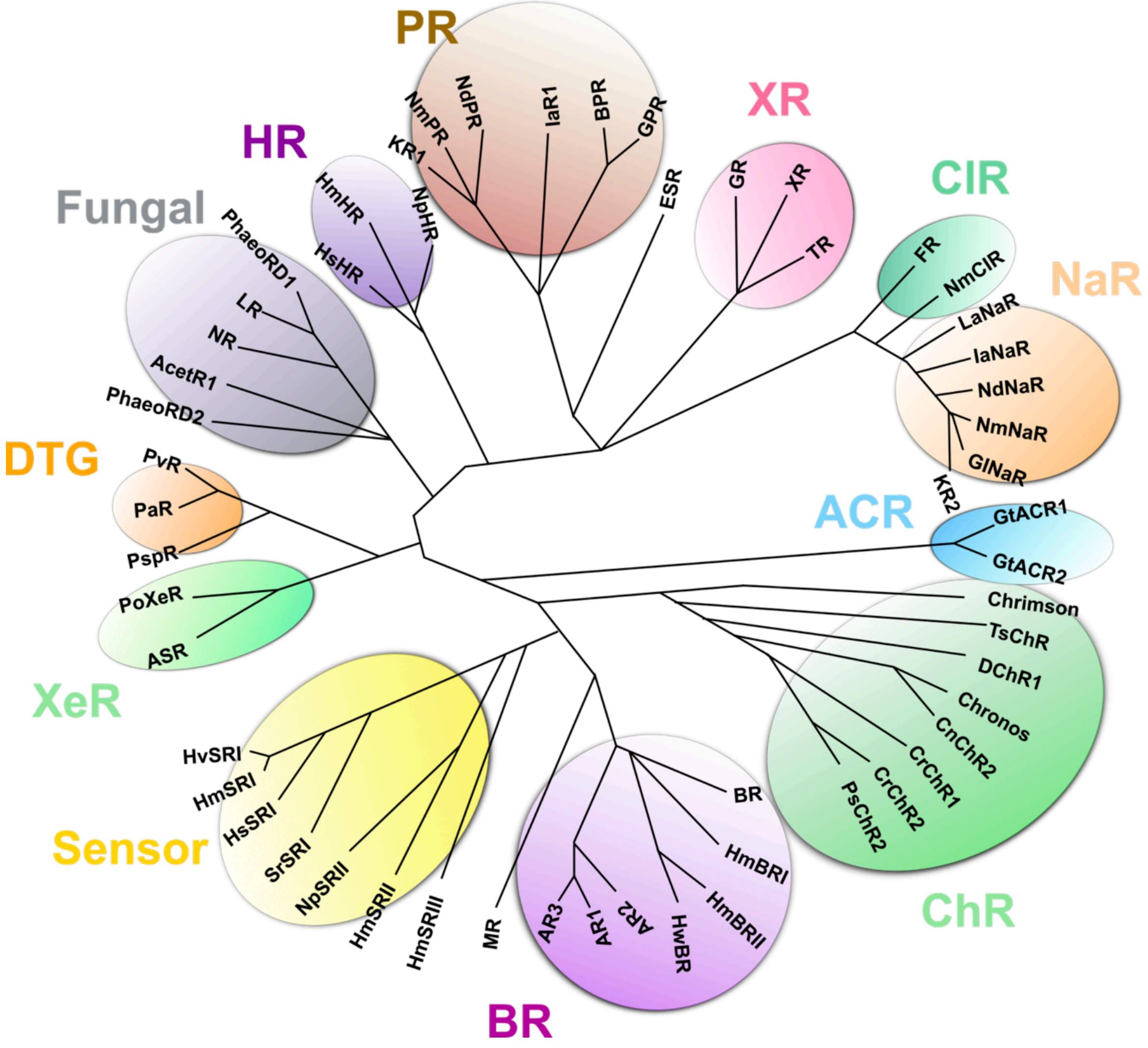


Figure 1.3.1 Adapted from [41]. Pictorial representation of known microbial rhodopsins phylogenetic tree.

1.4 The Arch family; journey towards enhanced fluorescence

Arch3 (PDB ID: 6GUX), as its name implies, is an archeal rhodopsin phylogenetically related to BR and found in the archaebacterium *Halorubrum Sodomense*. Although its sequence similarity with BR is not particularly high (~74%), the residues in the hydrophobic core are conserved, and the outward proton pumping function is preserved[42]. The event driving the proton pumping mechanism, shared among all the archaeal rhodopsins (including BR and Arch3), is the light-induced isomerization of the retinal C_{13}=C_{14} double bond .This event is the origin of a chain process, a photocycle initiated by light of a specific wavelength and followed by a number of chemical reactions that finally result in the net translocation of a proton across the membrane[42].

Although the proton pumping mechanism of Arch3 can be exploited in optogenetics to enable individual neurons to be silenced or triggered[40], a lot of effort has been put by scientists into identifying Arch3-mutants to be used as GEVIs. Indeed, Arch3 has been found to be fluorescent and exhibiting a significant voltage-sensitive fluorescence and fast kinetics[23,43]. However, the FQY of the DA state of Arch3 is estimated to be as low as ~0.0001. It was proposed that a higher FQY in Arch3 is reached through a three-photon process, where the first photon initiates a photocycle (typical of microbial rhodopsins), and the second photon absorbed by the N-state ultimately brings the molecule to the fluorescent Q-state, which upon absorption of a third photon may finally emit fluorescence[44,45]. Despite this increase of FQY, which in BR is estimated to be 0.007 in the Q state, the 3-photon process is a non-desirable technological feature, which hampers the use of Arch3 as a GEVI due to the limited response time of the fluorescence to voltage changes and the high intensity laser required to excite the molecules[20].

Among the first experiments to increase the DA FQY of Arch3 were performed by Hochbaum and coworkers[20]. Here, a hierarchical screening approach combined with random mutagenesis has allowed to identify two Arch3 variants with increased FQY named QuasAr1 (FQY ca. 0.080) and QuasAr2 (FQY ca. 0.065). On the other hand, Flytzanis et al. identified two new mutants, namely Archer1 and Archer2, using the site-saturation mutagenesis library of GR rhodopsin, which was originally designed to evolve GR for spectral shifts[46,47]. Here, they used the same library transferring the mutations to the homologue positions in Arch3 and screening the mutants for fluorescence intensity rather than λ^{a}_{max}. Archer1 features an FQY of 0.033, while the FQY of Archer2 was not measured. Later, Arnold and coworkers used a similar approach to identify Arch3 variants; whilst still transferring mutations from GR to Arch3, they resorted to multiple rounds of random and targeted site-saturation mutagenesis to accumulate mutations improving fluorescence over multiple generations[46]. With this approach, they found that several Arch3 mutants were showing increased FQY, the most remarkable being Arch5 (five point mutations to the wild-type Arch3 and Arch7 (seven mutations), with measured FQYs of, respectively, 0.087 and 1.2. More recently, Boyden and coworkers designed a directed evolution strategy to screen hundreds of thousands of proteins and evaluate each one for specific properties, culminated in the identification of multiple opsin-based fluorescent reporters, such as Archon2 (FQY ~1.2)[48]. Importantly, these increase in FQY are accompanied, for all the Arch3 variants, by a red shift in λ_{max}^{a}. Red-shifted absorption is a desired property in optogenetics, allowing for spectral orthogonality with actuators, low tissue penetration of the light excitation wavelength and reduced trigger of tissues auto-fluorescence.

A summary of the most relevant fluorescent Arch variants is reported in **Table 1.4.1**.

Table 1.4.1 Summary of relevant available spectral data for Arch2, Arch3 and the engineered Arch3 variants studied in this thesis.

Protein	$\Delta E_{S1\text{-}S0}^{a,Exp}$ (kcal mol^{-1})	$\lambda_{max}^{a,Exp}$ (nm)	$\Delta E_{S1\text{-}S0}^{f,Exp}$ (kcal mol^{-1})	$\lambda_{max}^{f,Exp}$ (nm)	FQY
Arch2	51.5	555	NR[a]	NR[a]	NR[a]
Arch3	51.4	556	41.6 [b]	687[b]	$1–9 \cdot 10^{-4}$
QuasAr1	49.3	580	40.0	715	$6.5 \cdot 10^{-3}$
Archon2	48.8	586	38.9	735	$1.1 \cdot 10^{-2}$
QuasAr2	48.5	590	40.0	715	$4.0 \cdot 10^{-3}$
Arch7	46.4	616	39.3	727	$1.2 \cdot 10^{-2}$
Arch5	46.0	622	39.1	731	$8.7 \cdot 10^{-3}$

[a] Not Reported

[b] The λ^{f}_{max} reported is likely coming from the Q-intermediate rather than from the DA state (see ref. [49])

Interestingly, two different studies seem to support the hypothesis that in these Arch3-mutants with increased FQY, the fluorescence is a one-photon process (i.e., originating directly from the DA state) as opposed to the parent three-photons process[43,50–52].

Although there are many biological and technological features that a rhodopsin based GEVI should ideally display to enable its in-vivo application in optogenetics, it is mandatory that the rhodopsin candidate should display a significant fluorescence in its DA state. For this reason, a large part of my dissertation and of my research, discussed in Chapter 3, provides a mechanistic picture which should answer to two questions; (i) whether the fluorescence in Arch3 variants is a one or three-photon process, as speculated for the wild-type and (ii) how limited modifications in Arch3 gene sequence reflect in new interactions between the opsin and the retinal chromophore in order to enhance fluorescence. In the next paragraph, I will provide to the reader the basis for understanding the main results of my research. I will discuss (i) how chromophore-protein interactions modify photophysical and photochemical properties and (ii) the currently available theoretical framework to explain spectral shift and fluorescence tuning from the standpoint of electronic structure theory.

1.5 The photochemistry of microbial rhodopsins

1.5.1 Opsin shift and spectral tuning

It is clear, at this point, that microbial rhodopsins are light-responsive proteins evolved to capture photons of light and exploit their energy to exert all sorts of different biological functions. The way this conversion is carried out is via a photochemical reaction, which, in microbial rhodopsins, consists in the stereo-selective isomerization of the all-*trans* isoform of the rPSB chromophore to yield the 13-*cis* isoform (see **Figure 1.2.1**). For different natural rhodopsins, which have adjusted the absorption spectra of the rPSB chromophore to adjust to different photic environments, the maximum likelihood of initiating the photoisomerization reaction will occur if the wavelength of the incoming photon will correspond to the λ_{max}^{a} of that rhodopsin. Obviously, this property has extreme biological importance and contributes to the regulation of fundamental processes such as the human vision; here, in the retina, rod cells are sensible to light with λ_{max}^{a} = ~500 nm, allowing for vision in dim conditions, whereas bright color vision relies on cones containing three different rhodopsins with λ_{max}^{a} at 440 nm, 530 nm and 580 nm, that make up for the perception of blue, green and red color respectively[53,54].

Spectral tuning plays a fundamental role also in microorganisms. Different organisms expressing different rhodopsin variants occupy a variety of different niches in earth's ecosystems. For instance, halo-archaea express rhodopsin chloride-pumps (Ugalde et al. 2011[55]) with λ^{a}_{max} in the yellow, the optimum to capture the photons of light which can get through the surface of salt ponds where they are found. Similarly, sensory rhodopsin II (SRII) and proteorhodopsin (PR)[56] inhabit shallow-water photic zones where peak spectral irradiance is attenuated by water transmissivity and feature blue-shifted $\lambda^{\mathrm{a}}_{\mathrm{max}}$[55]. In general, however, rhodopsins feature a strongly red-shifted λ^{a}_{max} with respect to the rPSB in solution (λ^{a}_{max}=440 nm in methanol), resulting from differential interactions between the protein and the chromophore. This phenomenon goes under the name of "opsin shift" and depends on how steric and more importantly electrostatics interaction between the rPSB chromophore and the protein environment modulate the energy difference between S_0 and S_1 (ΔE^{a}_{S1-S0}). The most important electrostatic interaction is summarized in **Figure 1.5.1**. In general, the following factor regulate the S_1-S_0 energy gap and affect the spectral shift: (i) chromophore planarity and distortion, (ii) the electrostatic interaction induced by the charged, polar, and polarizable amino acids in the retinal binding pocket[57]. It should be noted, however, that it has been found on several occasions that also residues far from the chromophore can modify $\Delta \mathrm{E}^{\mathrm{a}}_{\mathrm{S1-S0}}$ [58].

1.5.2 The Nakanishi two-point charges model

As discussed in the previous section, the determinants of λ^{a}_{max} modulation can be mapped to the macro-categories of (i) steric and (ii) electrostatic interactions of the opsin environment with the rPSB chromophore. However, another way to look at the phenomenon, is identifying to which specific residues in the proteins the λ^{a}_{max} value is more sensible to. With this regard, it is clear that a major role is played by the main counterion complex, as reported in several contributions[59–62]. Indeed, a key mechanism for rhodopsins to achieve sensitivity to light in the visible spectrum (the "opsin shift" discussed in the introduction of the paragraph), is the protonation of retinal Schiff base (rSB). Since a positive iminium cation would not be stable in the highly hydrophobic environment of the retinal binding pocket, in most rhodopsins, evolution selected a negatively charged amino acid residue (i.e., an aspartic or glutamic acid) near the rPSB (hence the name "counterion"), to stabilize it[63]. The most relevant theoretical framework to discuss the interaction of the counterion with the rPSB chromophore is the external two-point charges models proposed in the 1980 by Nakanishi and coworkers for Rh and BR[64,65]. According to this model, the interaction of the retinal with the counterion can be modeled as follows: In the electronic ground state (S_0), the positive charge of the retinal chromophore is mostly located on the PSB moiety. Upon photon absorption, the chromophore is promoted to its first electronically excited state (S_1), which features a strong charge transfer (CT) character, characterized by an inversion of the single and double bond character (bond length alternation or BLA), with the positive charge being displaced towards to the β-ionone ring end of the chromophore (**Figure 1.5.1A**). Within this framework, the interaction of negatively charged counterion(s) positioned in proximity of the rPSB determines a strong energy stabilization of S_0, which leads to an increased S_1-S_0 energy gap resulting in a blue-shifted $\lambda^{\mathrm{a}}_{\mathrm{max}}$ (**Figure 1.5.1D**). On the contrary, if a negative charge where to be located closer to the β-ionone moiety, S_1 would benefit from a larger stabilization with respect to S_0, ultimately diminishing $\Delta\mathrm{E}^{\mathrm{a}}_{\mathrm{S1-S0}}$ and red-shifting $\lambda^{\mathrm{a}}_{\mathrm{max}}$ (**Figure 1.5.1C and Figure 1.5.1D**). A more complex spectral tuning picture, outlined by Borhan et. al. consists in a rhodopsin model not featuring a charged counterion (**Figure 1.5.1B**); in this situation, $\Delta\mathrm{E}^{\mathrm{a}}_{\mathrm{S1-S0}}$ has been shown to become more sensitive to subtle changes in rPSB neighboring residues dipoles[63].

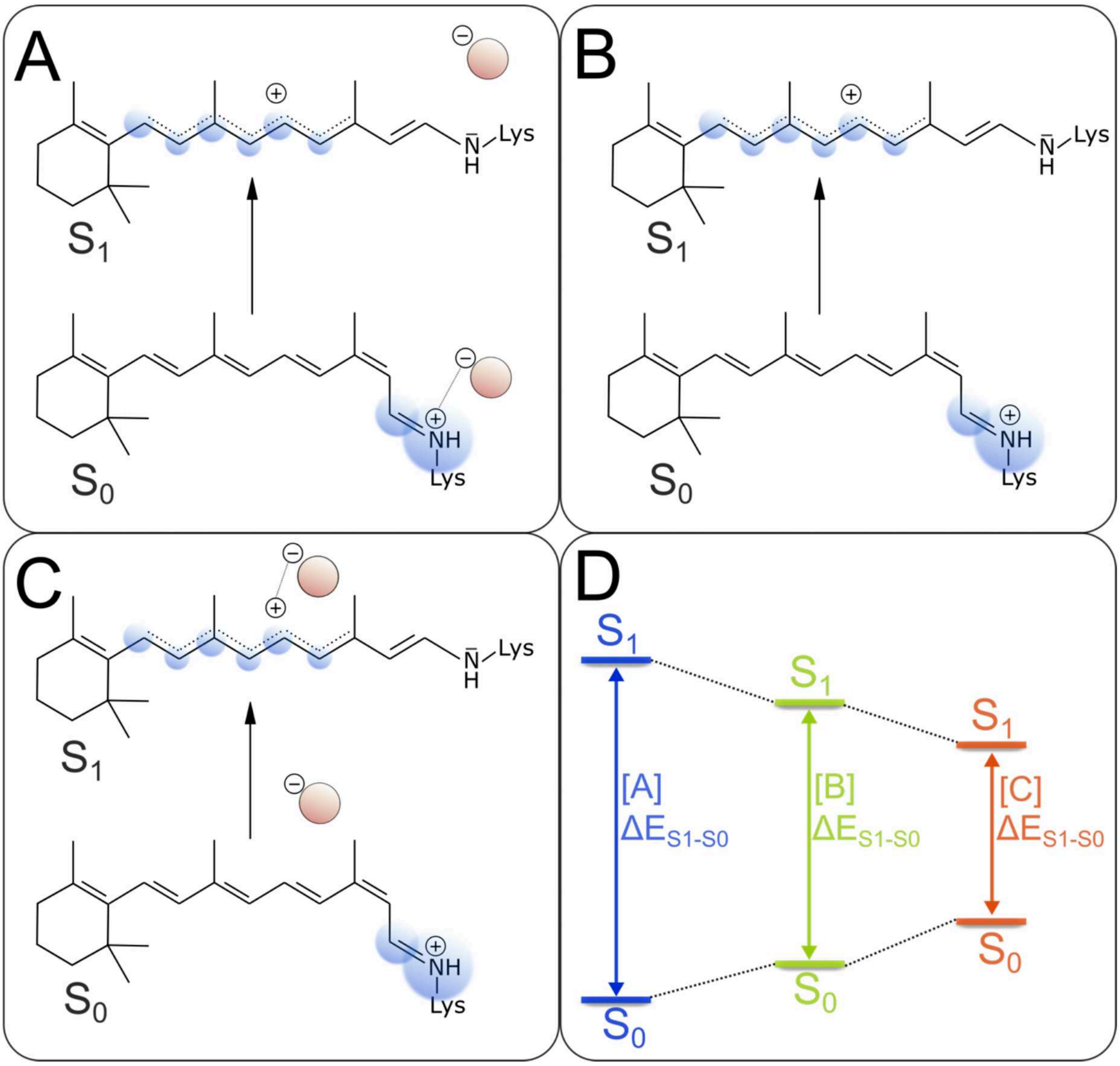


Figure 1.5.1 Spectral tuning exemplified in rhodopsins featuring rPSBAT. (A) Vertical excitation energy response to a negatively charged counterion located in proximity of RPSB positive charge. (B) Vertical excitation energy in the absence of a negatively charged counterion. (C) Vertical excitation energy when the chromophore counterion is re-located far from the PSB and closer to the β-ionone moiety. (D) Effect of electrostatic interactions on the vertical excitation energy may result in a spectral blue-shift [A], red-shift [C] with respect to the reference [B].

1.5.3 The Arch case: role of the main counterion

This model, although intuitive, is rather simplistic. In fact, in most rhodopsins, included the Arch set, which is the main topic of this dissertation, the electrostatics is extremely complex, with a varying and possibly large number of charged residues, multiple counterions, complex hydrogen bond networks (HBN), dipoles and polarization effects which most of the times are not accounted for both in theoretical and experimental studies.

An example of this complexity is outlined indeed by the Arch set itself. The Arch3 mutants generated throughout years of mutagenesis experiments aimed at both tuning the $\lambda^{\mathrm{a}}_{\mathrm{max}}$ to the red and increasing the FQY. Interestingly, all the Arch variants with red-shifted $\lambda^{\mathrm{a}}_{\mathrm{max}}$ and increased FQY selected by both site-directed and random mutagenesis experiments feature modifications to residues located in proximity of the RPSB moiety, and at least one of the point mutations is directed to a residue of the counterion complex.

In wild-type Arch3, the counterion complex consists of two aspartic acid residues located at the two sides of RPSB, namely D222 and D95. In QuasAr1, D95 is substituted by a histidine (H95), as opposed to QuasAr2 featuring the Q95 replacement. The D95H substitution is also included in the 13 point mutations that distinguish Archon2 from the wild type. Arch7 and Arch5 both share the D95 replacement by a glutamic acid (E95), however, Arch7 also features the D222 to S222 substitution[20,46].

It is clear, at this point, that mutating one or more charged residues in proximity of rPSB can have massive effects on the photophysical and photochemical properties of a rhodopsin. Interestingly, in the Arch family, experiments seem to outline a general correlation between $\lambda^{\mathrm{a}}_{\mathrm{max}}$ and FQY, whose investigation would be crucial to foresee the rational design of red-shifted Arch-based GEVIs. Since a point charges based theoretical model appears unsatisfactory to describe the spectral tuning of the retinal chromophore within the extremely complicated electrostatics of the cavity binding pocket, a large part (**Chapter 4**) of the present dissertation will be dedicated to proposing a computational model to overcome this limitation. As the focus of my research work was dedicated to investigating the photochemistry of Arch variants, which offer the perfect example of the importance of the counterion and its electrostatics interaction with the chromophore electron density, I developed a theoretical investigation tool that allows to model said interactions as hypothetic delocalization of the main counterion negative charge and document its effect on λ^{a}_{max}.

1.5.4 The photoisomerization reaction

Thus far, we have delivered the message that rhodopsins can exploit photons of light to initiate their biological function, and how optimizing the efficiency towards absorption of photons with specific wavelength, has deep implications both in nature and in technological applications. Now, we will proceed on discussing what happens upon photoexcitation of the rPSB, and how the photon energy can trigger different photochemical or photophysical processes. In natural rhodopsins, to initiate biological function, the very photochemical reaction is invariably an ultrafast isomerization of a double bond, generally the *cis* to *trans* isomerization of the C_{11}=C_{12} bond for animal rhodopsins and *trans* to *cis* C_{13}=C_{14} isomerization for microbial rhodopsins (**Figure 1.5.2A** and **Figure 1.5.2B**). The efficiency of photoisomerization is summarized by the ratio of photons leading up to a successful isomerization to photon absorbed, or photoisomerization quantum yield (QY, as opposed to the FQY described before). For instance, in Rh, where photoisomerization triggers the events required by vision, the QY is exceptionally high, (0.66 ± 0.03), and the excited state lifetime (ESL) is extremely low (<100 fs), meaning that the photochemical reaction is ultrafast. If a photon is exploited for ultrafast photoisomerization, the very short timescale will not allow for the photon to be emitted as fluorescence, meaning that generally $QY \propto 1/FQY$. For the QY to be very high, (or FQY to very be low), the S_1 PES should feature a barrierless or almost barrierless isomerization path leading with the S_1 electronic gradient driving the rPSB chromophore to a decay funnel (CoIn) and stereo-selectively to the S_0 PES of the isomer photoproduct (**Figure 1.5.2**). On the other hand, if S_1 features a sizeable barrier, the decay through CoIn is slowed down, and the molecule may remain in the excited state long enough to emit fluorescence (more details regarding the rPSBAT case are provided in **Section 3.7**). Microbial rhodopsins, including the main target of this dissertation, the Arch set, generally feature longer ESLs than animal rhodopsins, suggesting that their S_1 PES along the all-*trans* -> 13-*cis* isomerization path is generally less steep than the main isomerization channel in animal rhodopsins (11-*cis* -> all-*trans*).

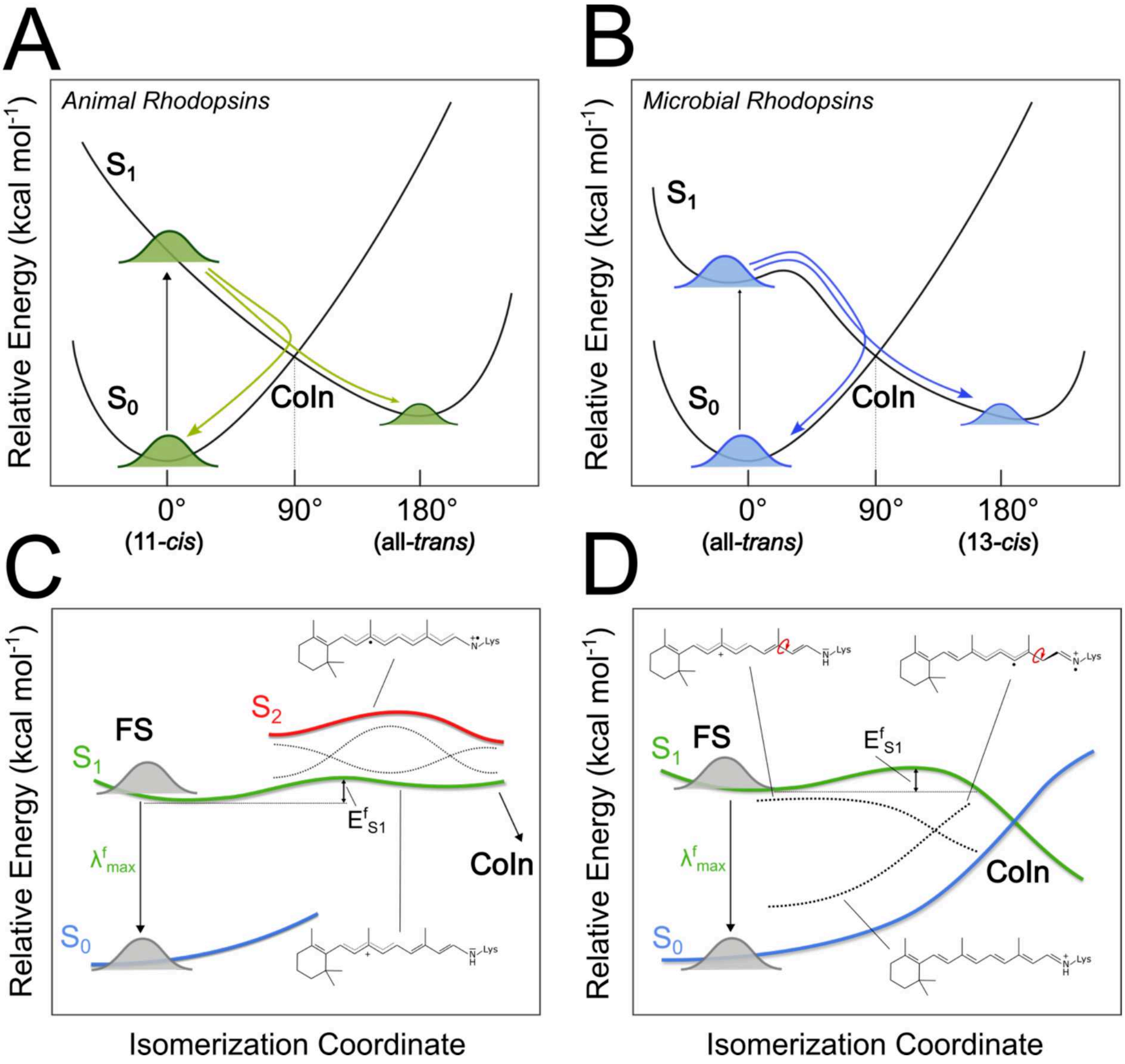


Figure 1.5.2 Schematic representation of the rPSB photoisomerization reaction in (A) animal and (B) microbial rhodopsins. In many natural animal rhodopsins (as Rh) the isomerization takes place in an ultrafast timescale through barrier less profiles, as opposite to microbial rhodopsins, where longer ESLs suggest the presence of an isomerization barrier on S_1. Vertical arrows represent the photon absorption initiating photoisomerization. Conical intersections (CoIn) are intersection points of S_0 and S_1 electronic PES. Here, we are interested in intersections along the torsional coordinate. (C-D) Scheme of barrier genesis mechanisms along the S_1 isomerization path of a fluorescent rhodopsin. (C) A shallow surface or a barrier may be generated via S_2/S_1 mixing (see ref. [66]). (D) Two-state mechanism proposed for the Arch set is generated via coupling of the diabatic S_1 and S_0 states. The dotted curves represent diabatic energies associated with the different electronic characters. Resonance formulas are also given.

1.5.5 Fluorescence mechanisms in microbial rhodopsins

The mechanism at the origin of the S_1 photoisomerization barrier (E^f_{S1}) genesis in rhodopsin is still largely unexplored. A large part of the present dissertation, summarized in **Chapter 3** and in **PAPER I**, will be dedicated to justifying the origin of increasing E^f_{S1} values in the Arch-set, and relate such values to the observed experimental trend in FQY. To investigate such mechanism at the atomistic level, multiconfigurational quantum chemistry (MCQC, see **Chapter 2**), is the only viable option. In the following, I will briefly introduce the most relevant theoretical findings regarding the fluorescence mechanism in these systems.

The first three electronic singlet states (S_0, S_1, and S_2) of an all-trans polyene as rPSBAT in its S_0 equilibrium can be mapped to $1A_g$, $1B_u$ and $2A_g$ symmetries (see ref. [67]). The $1A_g$ character is related to S_0 and has a closed-shell-like covalent electronic structure. The $1B_u$ state corresponds to S_1 and is dominated by a singly excited electronic configuration leading to a charge-transfer character, while the $2A_g$ state, corresponding to S_2, is dominated by double excitations and has a diradical character. According to the selection rules for electronic transitions, $1A_g$ -> $1B_u$ and $1B_u$ -> $2A_g$ transitions have non-zero oscillator strength[68].

To this date, the only theoretical study providing mechanistic insights into the genesis of an S_1 barrier at the electronic structure level was conducted on ASR and 2 of its mutants by Marin et al.[66], summarized in **Figure 1.5.2C**. According to this mechanism, a barrier on S_1 may be generated by the mixing of the reactive S_1 state with the partially bound, diradical, S_2 state. Such mixing creates small S_1 energy barriers or, possibly, “shallow” regions of the PES due to the increase of $2A_g$ diradical character associated to the S_2 wave function. In fact, the diradical character is associated to a partially locked $C_{13}=C_{14}$ bond (**Figure 1.5.2C**), as opposed to an elongated $C_{13}=C_{14}$ bond characterizing the S_1 state with $1B_u$ character. In synthesis, a shallow S_1 path can be attained via multiple crossing along the isomerization coordinate of the $2A_g$ and $1B_u$ diabatic states along a degenerate or nearly degenerate S_1/S_2 region of the adiabatic PES. Consequently, a shallow S_1 PES will slow down the evolution of the excited chromophore reactive torsion toward the CoIn, allowing the molecule to re-emit a photon of fluorescence.

This mechanism has also been proposed for BR, which, however, displays extremely low FQY in its DA state (**Section 1.4**).

One of the main findings in this thesis, discussed in **Chapter 3** and **PAPER I** and **II**, is a proposal of a new, previously unreported mechanism for the fluorescence in the Arch set.

I will show that in this family, the fluorescence can be explained based on a two diabatic-states model (**Figure 1.5.2D**). Here, the coupling between the reactive S_1 state with the covalent S_0 state leads to a destabilization of the CoIn structure with respect to the planar emitting S_1 minimum (FS). It is important to emphasize that, as opposed to the S_2/S_1 mixing, which is most likely in the FC/FS region, where S_1 and S_2 might happen to be degenerate, the S_0/S_1 coupling mechanism proposed for the Arch set depends on the evolution of the isomerization coordinate, or double bond twisting motion/deformation. Consequently, as discussed in **Section 3.8** the isomerization barrier is located along a distorted geometry of the chromophore, as the reactive $C_{13}=C_{14}$ progresses towards CoIn. In fact, as soon as the isomerization coordinate increases, the S_1 electronic character shifts from the mixed $1A_g/1B_u$ character of the FS region to the pure covalent/diradical character of the $1A_g$ diabatic state of **Figure 1.5.2D** in the CoIn region[69,70]. The $1A_g$ character has the $C_{13}=C_{14}$ bound, impairing the isomerization. Therefore, according to this mechanism, we hypothesize that the diabatic states crossing between $1A_g$ and $1B_u$ corresponds to the S_1 transition state (TS_{S1}), or energy maxima, where the $1A_g$/1Bu coupling is maximum[69].

Chapter 2

Methods

As discussed in Error! Reference source not found., notwithstanding the growing demand for rhodopsins with specific photochemical and photophysical properties, a fast and efficient rational approach for their design is still missing. Experimental approaches can be very expensive, complex and time consuming, but more importantly they cannot provide direct, atomic-level evidence of the mechanism of complex phenomena as spectral shift or fluorescence enhancement in rhodopsins. To this aim, computer-based simulations represent both a valuable alternative and a complement to experiments. In this Dissertation, I generated computer-based models of the proteins of the Arch set, which represent an invaluable opportunity to explore how limited modifications in the gene sequence can affect deeply the property of the first electronic excited state. In the following, I will provide a brief review of the theoretical models required to investigate the evolution of an electronically excited chromophore, such as the retinal chromophore, using modern computers. As it will be explained below, due to the large size of the photoreceptor proteins, such as rhodopsin, the hybrid quantum mechanics/molecular mechanics (QM/MM) embedding has been the main tool used to model them. In the chapter, we will revise the fundamental steps to build up from basic electronic structure methods to the construction of rhodopsin using standardized protocol such as the a-ARM protocol, used to generate the Arch set models. Such models represent the milestone to discuss the main findings of this dissertation, summarized in **Chapter 3** and **Chapter 4**.

2.1 The Hamiltonian Operator

According to the first fundamental postulate of quantum mechanics, any physical system can be specified by a wavefunction ($|\psi\rangle$). Acting with the proper operators on such a system (wavefunction), will yield a precise value for the corresponding observable (i.e., momentum, position or energy)[71]. For example:

$$\hat{p} = i\hbar \frac{d}{dx} \tag{2.1.1}$$

is the momentum operator, that acting on a wavefunction will return the momentum of the system for that particular state.

$$\hat{p}|\psi\rangle = i\hbar \frac{d|\psi\rangle}{dx} \tag{2.1.2}$$

Since most of the problems in computational chemistry revolves around the calculation of the energy associated to a specific chemical structure, it is straightforward to notice how the most important operator will be the energy operator, also known as the Hamiltonian operator:

$$\hat{H} = \hat{T} + \hat{V} \tag{2.1.3}$$

As shown in **Equation 2.1.3**, the energy of any chemical system will depend on kinetic and potential energy terms. These can as well be represented with their respective mathematical operators. Considering the system as constituted by nuclei and electrons, these latter being treated explicitly in quantum mechanics, it is useful to express the molecular Hamiltonian as:

$$\hat{\mathrm{H}}_{\mathrm{tot}} = \hat{\mathrm{T}}_{\mathrm{n}} + \hat{\mathrm{T}}_{\mathrm{e}} + \hat{\mathrm{V}}_{\mathrm{ne}} + \hat{\mathrm{V}}_{\mathrm{nn}} + \hat{\mathrm{V}}_{\mathrm{ee}} \tag{2.1.4}$$

2.2 The Schrödinger Equation

As discussed in **Section 2.1**, the state of any system in quantum mechanics can be specified by mean of a wavefunction. In computational chemistry, the most appropriate tool to describe a many electrons system is the Schrödinger equation (SE), a differential equation whose solution yields indeed the wavefunction as a function of time[71];

$$\left[\frac{-\hbar^2}{2m}\nabla^2 + V(\vec{r},t)\right]\psi(\vec{r},t) = i\hbar\frac{\partial}{\partial t}\psi(\vec{r},t) \tag{2.2.1}$$

Above, is shown the time-dependent non-relativistic Schrödinger equation (TDSE). Notice that the terms in square brackets of the left-hand side correspond to the Hamiltonian operator, introduced in **Section 2.1**. Thus **Equation 2.2.1** can be re-written, after separation of variables, as the well-known time independent Schrödinger equation (TISE)[72]:

$$\mathrm{H}(\vec{\mathrm{r}})\psi(\vec{\mathrm{r}}) = \mathrm{E}(\vec{\mathrm{r}})\psi(\vec{\mathrm{r}}) \tag{2.2.2}$$

This has the mathematical form of a secular equation (or eigenvalue equation), and its solutions are quantum states called stationary states. A huge effort in computational chemistry goes into solving the above equation and finding the energy of the stationary states. However, for systems of more than 2 particles it is not possible to find analytical solutions[71]. In fact, while it is relatively simple to write down an expression for the Hamiltonian operator, it is extremely challenging to do the same for $|\psi\rangle$. Nevertheless, approximated solutions can still be achieved by mean of several *ab initio* methods, some of which will be illustrated later in this chapter.

In this sense, a common simplification in quantum chemistry known as Born-Oppenheimer (BO) approximation consists in considering the electrons as particles moving in the electrostatic field of fixed nuclei, given the large ratio of nuclear particles masses to electron masses. Under this assumption, the TISE can be broke down in a nuclear and electronic part, where R and r denote the nuclear and electronic coordinates, respectively.

$$[V_{nn}(\mathrm{R}) + E_{el}(\mathrm{R})]\psi = E_{tot}\psi \tag{2.2.3}$$

Where $V_{nn}(\mathrm{R})$, the electrostatic repulsion of the nuclei is a constant which is easily derived from static Coulombic interactions. In other word, we have now that $\mathrm{E_{tot}}$ can be computed by solving the electronic part of the TISE for a specific geometrical arrangement of the nuclei.

2.3 The Hartree-Fock Method

Until know, we didn't discuss in detail how to construct the electronic Hamiltonian discussed at the beginning of this chapter. Since we are looking for solutions to the electronic TISE, we represent electrons for a molecule of a given geometry (assuming BO approximation), using molecular orbitals (MO) functions. As shown in **Equation 2.3.1**, we introduce the variable x to represent a single MO as product of a spin function $\omega(\mathrm{s})$ and a space function $\phi(\mathrm{r})$.

$$\phi(\mathrm{x}) = \phi(\mathrm{r})\omega(\mathrm{s}) \tag{2.3.1}$$

According to the Pauli exclusion principle, the wavefunction for a polyelectronic system should be antisymmetric, thus a convenient mathematical construct to represent such a wavefunction is a Slater determinant:

$$\psi(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N) = \frac{1}{\sqrt{N!}} \begin{bmatrix} \phi_1(\mathrm{x}_1) & \cdots & \phi_N(\mathrm{x}_1) \\ \vdots & \ddots & \vdots \\ \phi_1(\mathrm{x}_N) & \cdots & \phi_N(\mathrm{x}_N) \end{bmatrix} \tag{2.3.2}$$

Using this idea, one of the most effective and simple methods to solve the non-relativistic electronic TISE, is using the Hartree-Fock (HF) approximation to construct a single Slater determinantal wavefunction[73]. HF relies indeed on the variational principle to find the best set of molecular orbitals minimizing the energy of a polyelectronic wavefunction. To deal with electron-electron repulsion, HF takes advantage of the Fock operator ($\hat{f}$), a one electron operator where the simultaneous electrostatic interaction of one electron with the N-1 remaining electrons is formulated as a mean field (i.e., averaged over the N-1 remaining electrons positions). However, the formulation of $\hat{f}$ turns the electronic TISE in a pseudo-eigenvalue equation, depending on its own eigenfunctions (i.e., the molecular orbitals). For this reason, the HF equations must be solved iteratively, until the ground state (S_0) energy converges to a minimum. This procedure goes under the name of self-consistent field (SCF)[73].

2.4 The CASSCF Method

The mean-field approximation introduced by the Fock operator allows to provide the best approximate single-determinant wavefunctions to solve the non-relativistic TISE. However, the HF wavefunction has several limitations. Indeed, it cannot model correctly some quantum mechanical phenomenon, such as the Coulomb hole and the homolytic bond breaking, due to the lack of the so-called static and dynamic electron correlation. Furthermore, since HF is limited to closed-shell systems, it cannot deal with excited states in general, making impossible to investigate the photochemistry of molecules such as the rhodopsins discussed in this dissertation.

To overcome these limitations, it is possible to include in the construction of the electronic wavefunction all the possible Slater determinants for a given nuclear arrangement. This technique goes under the name of Full Configuration Interaction (FCI).

$$\psi_{\mathrm{FCI}} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{C_i}\psi_{\mathrm{i}} \qquad \psi_0 = \psi_{\mathrm{HF}} \tag{2.4.1}$$

The FCI method provides an exact solution to the electronic Schrödinger equation in the space spanned by the basis set by accounting for all the possible electronic excitations (i.e., all the possible electronic configurations). However, in a biological molecule such as rPSB this number the number of configurations to consider is impractical, and calculations are carried out using a truncated version of **Equation 2.3.1**. The approach employed in the presented research takes advantage of a specific truncated FCI variants called ‘complete active space self-consistent field (CASSCF)’ method[74,75]. The CASSCF method offers indeed the possibility to include in the FCI expansion only the MOs which are relevant for the chemical reactivity or photo-chemical reactivity of the molecule. Furthermore, in contrast with other truncated CI methods such as configuration interaction single double (CISD), in CASSCF, as the name implies, also the reference MO basis are optimized via SCF in the new space of configurations generated. This is important, because in CASSCF optimizing MOs for a single state or multiple states allows to study reaction paths and excited state dynamics, which is a crucial component of the calculation carried out in this thesis.

To describe the photon-absorption and more generally the rPSB photoisomerization process described in this dissertation, we included in the CAS space the 6 π and 6 π* orbitals of the retinal chromophore. Since we are including 12 active electrons in 12 orbitals, the corresponding wavefunction will be called CASSCF(12,12). **Figure 2.4.1** offers a schematization of these concepts.

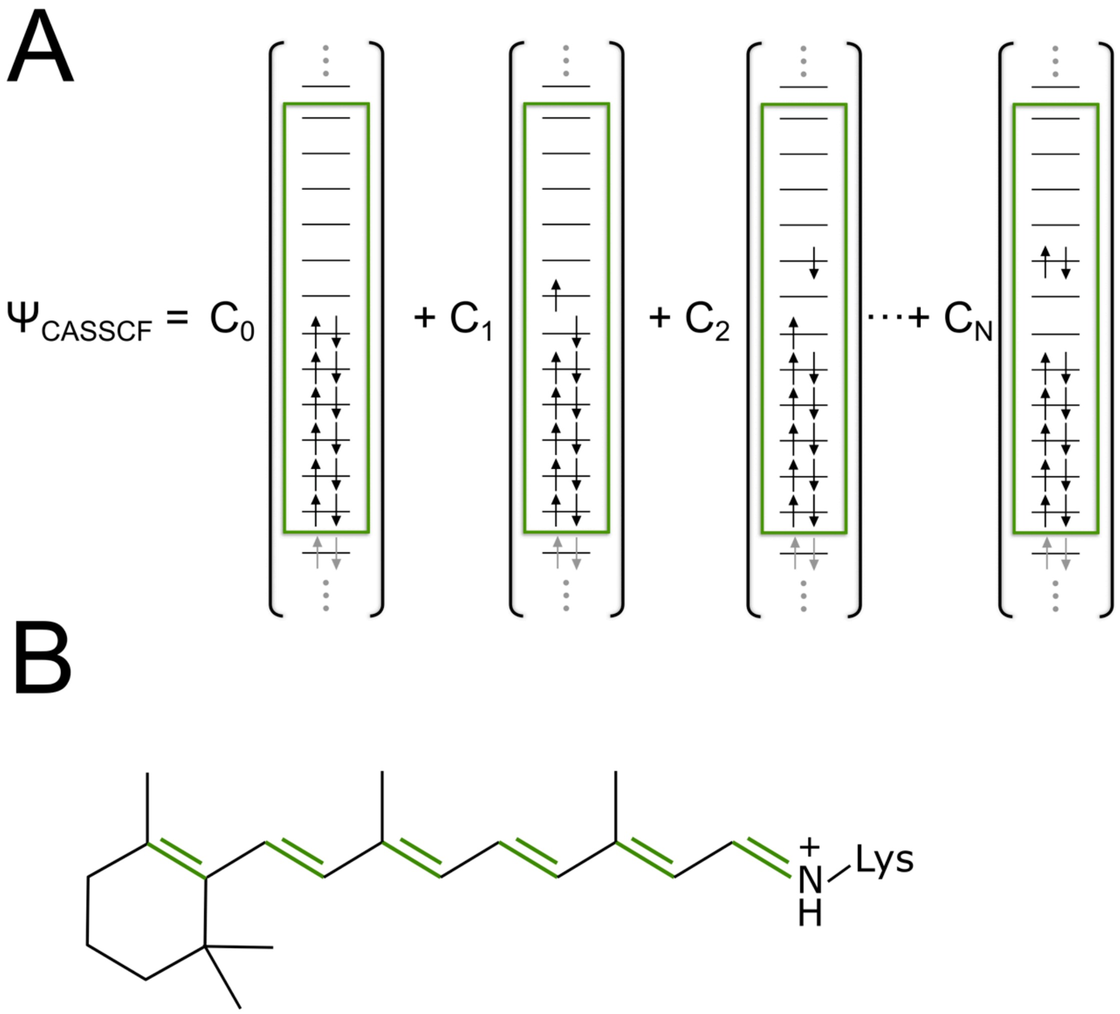


Figure 2.4.1 Schematic representation of a (12,12) CASSCF wavefunction. The active orbitals are framed in green boxes, active electrons are shown as black arrows and inactive orbitals/electrons are shaded in gray. (B)The active orbitals correspond to the 6 conjugated double bonds of the retinal chromophore.

2.5 The CASPT2 method

Throughout the discussion of the results of this Dissertation, the CASSCF method will be employed to for calculating geometries and reaction paths for the rPSBAT chromophore of the rhodopsins of the Arch set. However, the CASSCF method is still lacking, at least in part, a good description of dynamic electron correlation. Indeed, while dynamic correlation is somewhat accounted for in the space spanned by the active space orbitals, its global description is still unsatisfactory. Among the approaches available to recover the missing dynamic electron correlation, we briefly discuss the perturbative methods CASPT2 and its extended multi-state (XMS) variant, XMS-CASPT2, employed in this dissertation due to their relatively cheap computation time[76–79]. According to the CASPT2 theory, we express the energy as the zeroth order term plus some higher order correction:

$$(\hat{H}_0 + W)|\psi_i\rangle = \varepsilon_i|\psi_i\rangle \tag{2.5.1}$$

If the solution to the zeroth order term is known, it is possible to solve the Equation above by Taylor expanding the $\hat{H}_0$ eigenvalues and eigenvectors.

$$\varepsilon_i = E_i^0 + \lambda E_i^1 + \lambda^2 E_i^2 + \ldots + \lambda^n E_i^n \tag{2.5.2}$$

and

$$|\psi_i\rangle = |\psi_i^0\rangle + \lambda|\psi_i^1\rangle + \lambda^2|\psi_i^2\rangle + \ldots + \lambda^n|\psi_i^n\rangle \tag{2.5.3}$$

Substituting these expansion in the TISE it is possible to arrange n perturbation equations which can be solved for λ. In the CASPT2 variants, the correction to energy and wavefunction are respectively of the second order and first order. This is a reasonable compromise between accuracy and efficiency and it is particularly suitable when the perturbation is applied to a CASSCF zeroth order reference. In general, CASPT2 yields 0-0 excitation energies which are in better agreement with experimental data with respect to CASSCF. One of the shortcomings of the CASPT2 method is the inability to provide smooth PESs around near degeneracy points, such as conical intersections (CoIn), which are critical elements in the description of the photochemical processes characterizing the systems under investigation in this dissertation[77]. For this reason, on several occasions, we employed the XMS-CASPT2 variant of the CASPT2 method. Briefly, in the XMS formalism, the wave functions are invariant with respect to unitary rotations of the reference functions, which in turn provides improved potentials in the near crossing regions and so-called CIs[77].

2.6 QM/MM methods

The capabilities and accuracy of ab-initio QM calculations of the electronic structure of molecules presented thus far makes them particularly appealing, but the required CPU time makes their usage impractical to calculate energies, geometries and properties of systems of more than a few tens of atoms. Accordingly, to describe large systems such as the photoactive microbial rhodopsins discussed in this dissertation, we took advantage of a hybrid quantum mechanics/molecular mechanics (QM/MM) scheme[80,81]. To do this, we conceptually map the systems into two domains; (i) the chemically or photochemically relevant site, the rPSBAT chromophore, is treated at the QM level while (ii) the opsin, is treated at the much cheaper molecular mechanics (MM) level using classical forcefields.

In general, the QM/MM Hamiltonian operator can be expressed as:

$$\widehat{\mathrm{H}}_{\mathrm{tot}} = \widehat{\mathrm{H}}_{\mathrm{QM}} + \widehat{\mathrm{H}}_{\mathrm{MM}} + \widehat{\mathrm{H}}_{\mathrm{QM/MM}} \tag{2.6.1}$$

The distinction of the existing QM/MM methodologies implementations arise from the selected formalism to describe the QM/MM interaction term ($\widehat{\mathrm{H}}_{\mathrm{QM/MM}}$) in **Equation 2.6.1**, encapsulating the mutual electrostatic interactions of the QM electron density with the MM point charges as well as the criteria to describe the frontier between these two. This scheme should not break the uniqueness of the QM/MM PES. For all the models generated in this dissertation, we consistently employed the standard QM/MM treatment at the core of the a-ARM protocol (discussed in the following). The Link Atom (LA) scheme is used to describe the QM/MM. This consists in placing a hydrogen atom (HLA) along the covalent bond connecting the QM and MM fragments, in order to saturate the valence of the QM part[82]. The degrees of freedom of the HLA are restrained by using the Morokuma scheme and in our models, it will be located along the Cδ-Cε retinal-linked lysine bond, as reported in previous contributions[33,34,57]. Concerning the electrostatic interactions between the two subsystems, the wavefunction of our QM fragment experiences the electric field of fixed MM point charges, while polarization of the MM fragment by the QM subsystem is presently not accounted for.

2.7 The a-ARM protocol

The theoretical background provided in this chapter, intends to provide the key conceptual points needed to discuss the workflow of the Automatic Rhodopsin Modelling Protocol (ARM)[34]. Indeed, *all the results discussed in the present dissertation depend on rhodopsin QM/MM models generated with this technology*. This protocol allows for a substantially automatic, fast and standardized construction of basic monomeric 'gas phase' hybrid quantum mechanical/molecular mechanical (QM/MM) models of wild-type rhodopsin systems and their mutants. Recently, an updated version of ARM featuring an increased level of automation has been proposed as an high-throughput tool for computational photobiology. This version, from now on called simply a-ARM, has been benchmarked by calculating the absorption vertical excitation energy (ΔE^{a}_{S1-S0}) for a set of more than 30 phylogenetically distant rhodopsins, showing a mean absolute error (MAE) of c.a. 3.0 kcal mol^{-1} blue-shifted with respect to experimental data[33]. The a-ARM pipeline consists in a collection of bash scripts controlled by a python-based driver directing calls to different programs (see **Figure 2.7.1**) that finally generate a gas-phase uncharged monomer of rhodopsin system. The QM/MM treatment of rhodopsins in a-ARM (at the core of the protocol) requires to identify three fundamental fragments (see **Figure 2.7.1**) in the three-dimensional input structures, which may be crystal structures or homology models:

- QM atoms: comprise the retinal chromophore atoms, the linker lysine side chain atoms from the N-terminal to Cδ and the link atom. These atoms will be treated at the QM level.
- MD atoms: consists in those atoms that are free to relax during a-ARM molecular dynamics. These include the QM atoms and the side chains of the residues surrounding the chromophore (cavity residues).
- MM frozen atoms: represented by those atoms whose coordinates are always kept fixed. Therefore, they include all atoms except MD atoms. They are treated at the MM level.

a-ARM takes as input a plain PDB file, which can correspond to a crystal structure retrieved from the PDB[30] or to a comparative model such as MODELLER[83] or SCWLR4[84]. The initial setup is as follows:

1. The retinal protonated Schiff base (rPSB), chromophore bounded lysine , main and second counterion are identified automatically without calls to external programs.
2. PROPKA3[85] is invoked to predict the protonation state of the ionizable side chains. The global charge of the protein is neutralized, where necessary, by automatic placement of Cl^{-} and Na^{+} ions at the intracellular (IS) or extracellular (OS) side.

3. The residues forming the chromophore cavity are identified by Fpocket. In case the main and second counterion are not detected, they will be automatically included by a -ARM.

Once the initial setup is done, hydrogen atoms are added to the model by DOWSER[86] and GROMACS[87] in different steps. At this point, the system is equilibrated by mean of an MM energy minimization followed by multiple MM molecular dynamics (N=10 repetitions with different initial velocities) of 1 picosecond on the automatically included MD atoms. The 10 independent models obtained are then prepared for the QM/MM calculations (performed by MOLCAS[88] and TINKER[89]) . The hydrogen link atom (HLA) is added at the proper position (QM/MM frontier) and the QM atoms and MM atoms are readily identified. The 10 models are optimized in parallel through different steps featuring an increasing level of theory: from HF/3-21G/MM to single-root CASSCF(12,12)/6-31G*/MM to obtain the final geometries. These geometries are finally used to calculate 10 vertical excitation energies (ΔE^{a}_{S1-S0}) using a 3-roots CASPT2/6-31G*/MM calculation to correct a 3-roots state average CASSCF(12,12)/6-31G* reference wavefunction[34]. The MM force-field which is used both in the MDs and subsequent QM/MM calculation is the AMBER94 force field[90].

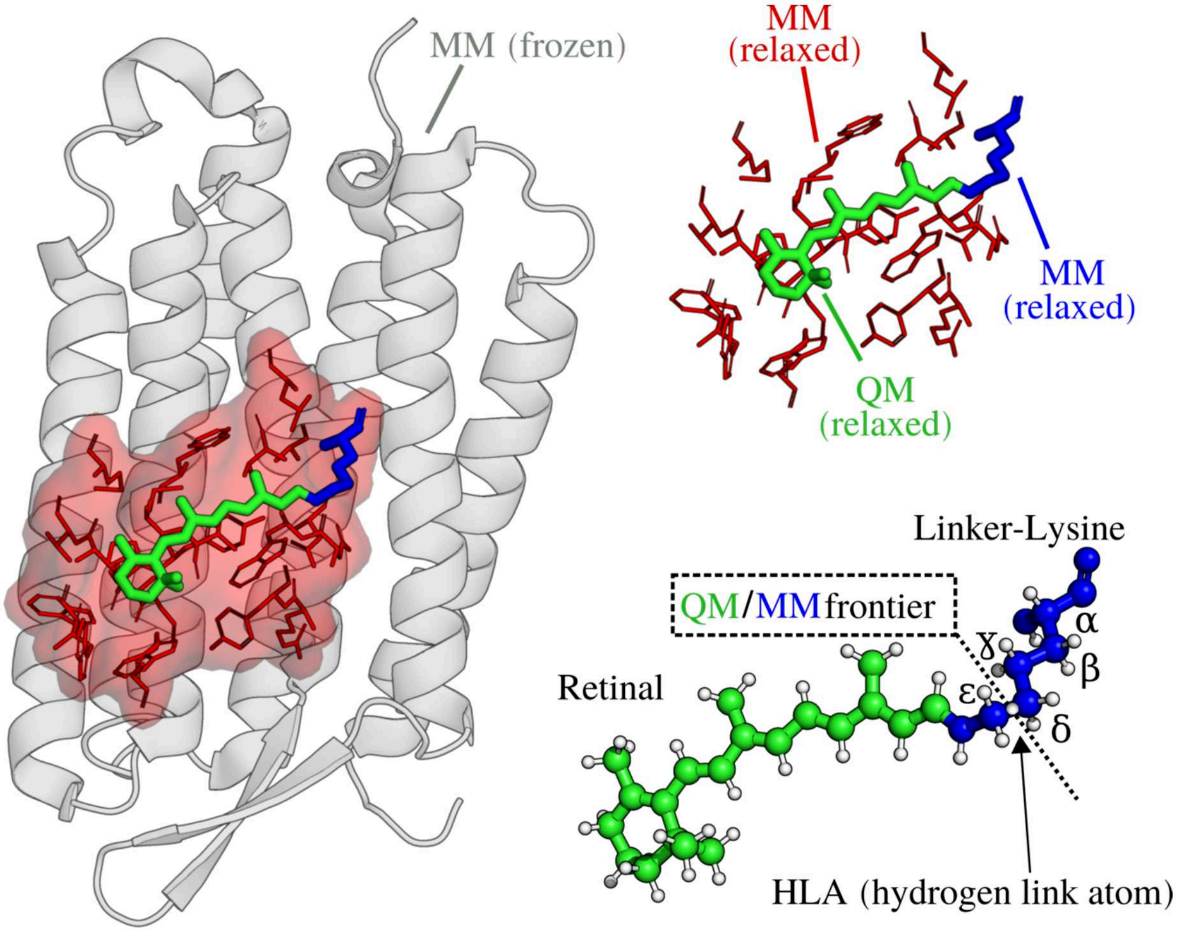


Figure 2.7.1 The a-ARM protocol QM/MM scheme. Schematic representation of an in silico designed QM/MM model generated with a-ARM. The a-ARM protocol partitions the rhodopsin rhodopsins structures in 3 subsystems which are treated with different approximation along the program workflow. These are the MM frozen atoms (in gray), MM relaxed atoms (in red) and QM atoms (in red). A frontier is defined along the $C\delta$ and $C\varepsilon$ carbons of the chromophore linker-Lysine. The QM valence is saturated by a dummy atom called hydrogen link atom (HLA).

Chapter 3

Fluorescence Enhancement in the Arch Family

In this chapter, based on the findings of **PAPER I**, addresses important issues regarding the visualization of neuronal activity using microbial rhodopsins. However, the significance of the presented results goes well beyond the development of novel tools for imaging. In fact, it shows how exactly a microbial rhodopsin, whose natural function is to use photon energy for catalyzing an isomerization reaction, can be re-engineered to become highly fluorescent. That has led to uncover a mechanistic theory connecting, at the molecular level, variations in protein sequence and changes in the quantum efficiency of light emission. I will show how the implementation of such theory reproduces the fluorescent intensities observed for a series of known Archaerhodopsin-based optogenetics reporters. The research, entirely carried out using multi-state and multi-configurational quantum mechanical computations, is connected to an experimental study recently published in Science by Romei et. al, which however is applied to Green Fluorescent Protein (GFP) variants. Our computations not only show that the empirically derived GFP fluorescence mechanism is also present in Archaerhodopsins (i.e., in a dramatically different class of proteins), but exposes the original intimate electronic and nuclear clockwork making such mechanism possible and thus paving the way to the engineering of highly fluorescent proteins in general.

3.1 Abstract

Investigating the dynamics of neuronal networks communication requires bright fluorescent probes which can localize in the neuron membrane. Although rhodopsins have already been exploited to control and visualize neural activity, their fluorescence is very dim, impairing the development of novel optogenetic tools. Here, we investigated the excited state of a set of mutants of the wild-type microbial rhodopsin Arch3, using quantum mechanics (QM), molecular mechanism (MM), hybrid QM/MM modelling to develop a theory relating amino acid sequences to fluorescence brightness. We document that the experimentally observed fluorescence quantum yield is directly proportional to an isomerization barrier which is regulated by the *relative energy difference of two minima on the first excited state potential energy surfaces. Furthermore, we propose that in this class of microbial rhodopsins the electrostatic interaction between the protein environment and a peculiar chromophore excited state electronic structure is key to enhance fluorescence. These findings allow to lay out a set of rules which will lead to the rational design and engineering of rhodopsin probes with brighter and brighter fluorescence emission.*

3.2 Introduction

As discussed in **Section 1.1**, Optogenetics relies on synthetic biology for the development of devices which can at the same time modulate and report the electrical activity of extended neural networks, both in vitro and in vivo. Depending on their characteristics, microbial rhodopsins may be used for triggering or reading out action potentials (APs). The progenitor of this second class of so-called genetically encoded voltage indicators (GEVIs), is Archaerhodopsin-3 (Arch3), a wild-type archaeal rhodopsin from Halorubrum Sodomense. Arch3 emits weak fluorescence in the near-infrared (NIR) after red-light absorption, which is sensitive to the membrane voltage. However, the fluorescence is dim and its on-set is slow because it does not come directly from the dark-adapted (DA) state but from a later photointermediate. Moving from single cell to population wide-field imaging, requires improved spatial and temporal resolution, which scientists have been trying to achieve via directed evolutionary approaches and random mutagenesis experiments. These attempts culminated in the discovery of a number of Arch3-variants with improved photophysical and biochemical properties, such as the Archers, the QuasArs, the Archons, Arch5 and Arch7, among others. All of these rhodopsins are spontaneously fluorescent and feature improved fluorescence quantum yield (FQY), kinetics and voltage sensitivity, which enabled their applications in imaging of acute brain slices, but also living mammals and invertebrates. Furthermore, their red-shifted absorption and emission makes them particularly suitable to be coexpressed with optogenetic actuators, as part of devices called optopatch, which can be used for all-optical electrophysiology[13,20].

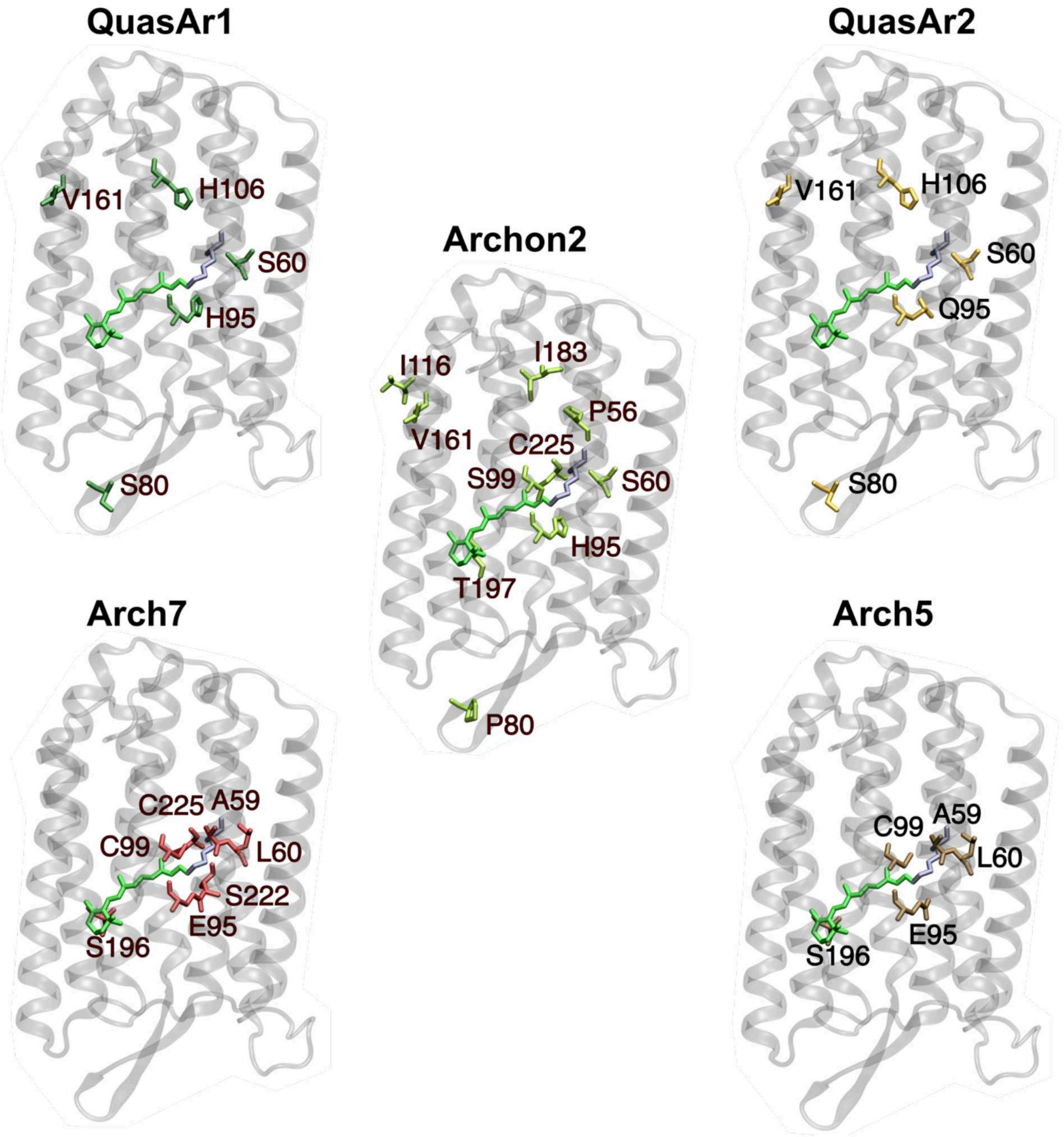
It is unquestionable how random and site-directed mutagenesis, as well as, in principle machine learning (ML) based techniques, are prosecutable brute force solutions to address the problem of fluorescence enhancement in GEVIs of this class. However, a deeper understanding of the main mechanistic ideas behind the fluorescence properties tuning, could speed up the future design of neuronal reporters, redirecting the experimental efforts with a more efficient rational design approach. Furthermore, the insights gained might be applicable to other families of fluorescent chromophores/dyes and even suggest how to modulate properties which are different from fluorescence.

In the present contribution, we provide a theory for fluorescence enhancement for the Arch set (see **Section 1.4**), a diversified set composed of seven microbial rhodopsins. These are the two wild types Arch2 (PDB ID: 3WQJ) and Arch3 (PDB ID: 6GUX), and five of the most natively fluorescent Arch3-mutants, namely QuasAr1, Archon2, QuasAr2, Arch5 and Arch7. Remarkably, the wild types have not reported to be fluorescent in their DA state[49]. On the contrary, the mutants have significant experimental FQY , ranging from ca. 0.004 (QuasAr2) to 0.01 (Arch7), which make them amongst the most fluorescent membrane integral proteins ever reported[46]. Interestingly, these variants have also extremely red-shifted maximum absorption and emission wavelengths ($\lambda^{a}_{max}$and λ^{f}_{max}), which appears to be correlated to their increased brightness. Below, by constructing QM/MM models of the target proteins and using a multiconfigurational quantum chemistry (MCQC) based approach, we suggest how the observed difference in FQY are modulated by the relative differences in shape of the S_1 potential energy surfaces of the rhodopsins in the set, particularly, by the height of the torsional barriers[69]. Differently from what previously postulated for another class of microbial rhodopsins (i.e. ASR and its mutants)[66], we propose a model for fluorescence in the Arch family which does not involve interactions between S_1 and higher excited states but is modulated by the evolution along the isomerization coordinate of the electronic character of the S_1 state itself. In addition, we show how the trend in $\lambda^{a}_{max}$is related to trending differences in topographical properties of the S_1 PES, which are encapsulated by the relative energy differences of two minima on the first excited state. We found, as foreseen by the Hammond postulate, that the gap between these two minima controls the magnitude of the barriers and therefore the FQY, thus spontaneously manifesting a potential target for fluorescence engineering in the Arch family.

3.3 From amino acid sequences to 3D structures

While Arch2 and Arch3 crystallographic structures are available (PDB IDS: 3WQJ and 6GUX), none of the other Arch set structures has yet been reported. Therefore, to be able to perform *in silico* QM/MM modelling of the Arch3 mutants, we first generated their three-dimensional structures via a comparative modelling approach designed to exploit the high sequence similarity of the mutants with Arch3. In fact, QuasAr1 and QuasAr2 differ from their progenitor for only 5 mutations. The P60S, T80S, D106H and F161V mutations are shared by QuasAr1 and QuasAr2, whereas at position 95 the aspartic acid of Arch3 is substituted by a histidine (D95H) in QuasAr1 and by a glutamine in QuasAr2 (D95Q). Arch5 also differs for 5 mutations from Arch3, namely D95E, T99C, V59A, P60L and P196S. Finally, Arch7 features all the Arch5 mutations plus the replacements D222S and A225C, yielding for a total of 7 substitutions with respect to Arch3. Finally, Archon2 differs from Arch3 for the substitutions T56P, P60S, T80P, D95H, T99S, T116I, F161V, T183I, L197I, A225C. All mutations are collected in **Figure 3.3.1**.

Comparative modeling was performed with the software MODELLER[83]. The structures of the targets (QuasAr1, QuasAr2, Archon2, Arch7 and Arch5) were generated by simply copying the coordinates of all the conserved amino acids from the template to the new structures and relaxing the atom coordinates of the mutated side chains. The crystallographic water molecules were also transferred from the template. The relaxed cartesian coordinates of the mutated side-chains are randomized and then refined by MODELLER via cycles of molecular dynamics (MD) and conjugate gradient (CG) optimizations to minimize an objective function derived from spatial restraints. These operations are repeated 200 times using the different seeds produced by the randomization step to yield 200 potential target structures differing in mutated side-chain conformation. The structures are then scored with the discrete optimized protein energy (DOPE) statistical potential. The final representative variant model is selected as the one featuring the lowest DOPE score.


QuasAr1

Archon2

QuasAr2

Arch7

Arch5

Figure 3.3.1 Pictorial representation of the mutations localization in the Arch3 mutants discussed in this chapter with respect to the parent wild-type rhodopsin Arch3. Mutated amino acids are shown in sticks and the corresponding label is also given. For each model, the retinal chromophore and linker-lysine are displayed using green and blue ball and sticks, respectively.

3.4 Generating the QM/MM models with the a-ARM protocol

As discussed in **Sections 2.6 and 2.7**, although QM/MM models are particularly suitable to study the photoreactivity of a system where one or more excited states are necessarily involved, their construction may require many parameters, ranging from the selection of the protonation states, the number of atoms included in the QM region, the QM theory, polarization of the MM embedding etc. Since we are not interested in elucidating the properties of a single protein but rather in capturing the different physics in the Arch set rhodopsins and evaluate how it reflects on their photochemistry, it was necessary to use a congruous and standardized approach to generate the corresponding QM/MM models. To this aim, we generated all the QM/MM Arch rhodopsins model using the a-ARM protocol discussed in **Section 2.7**.

3.5 Validation of the QM/MM models

Before the discovery of NeoR, the mutants of Arch where considered the most red-shifted and most fluorescent rhodopsins ever investigated[70,91–93]. As a first step toward the investigation of the excited state properties of this family, we focus indeed on evaluating whether the QM/MM models generated with a-ARM are indeed able to reproduce the experimental observed differences in λ^{a}_{max} across the set. Notice that by focusing on trends rather than absolute values, we avoid systematic errors. For instance, those related to the elementary structure of a-ARM models and the one relative to measured ΔE^{a}_{S1-S0} and ΔE^{f}_{S1-S0} obtained by converting the wavelength of spectral band maxima (λ^{a}_{max} and λ^{f}_{max} respectively) into photon energies. **Figure 3.5.1A** demonstrates that the computed and measured ΔE^{a}_{S1-S0} correlate linearly within the Arch set. Overall, the computed models feature a mean absolute error (MAE) of ca. 3.0 kcal mol^{-1} and a trend mean absolute error (TMAE) of 0.7 kcal mol^{-1}, which is within the documented error bars of the a-ARM protocol. **Figure 3.5.1B** displays the corresponding ΔE^{f}_{S1-S0} for the members of the set featuring a stable S$_1$ planar fluorescent state (FS). The presence of such an intermediate is assessed by locating a minimum on the S$_1$ PES. The computed ΔE^{f}_{S1-S0} MAE is 4.0 kcal mol $^{-1}$ and 0.6 kcal mol^{-1}, respectively, although, in contrast with ΔE^{a}_{S1-S0}, ΔE^{f}_{S1-S0} values are underestimated (i.e., red-shifted) with respect to the experimental quantities.

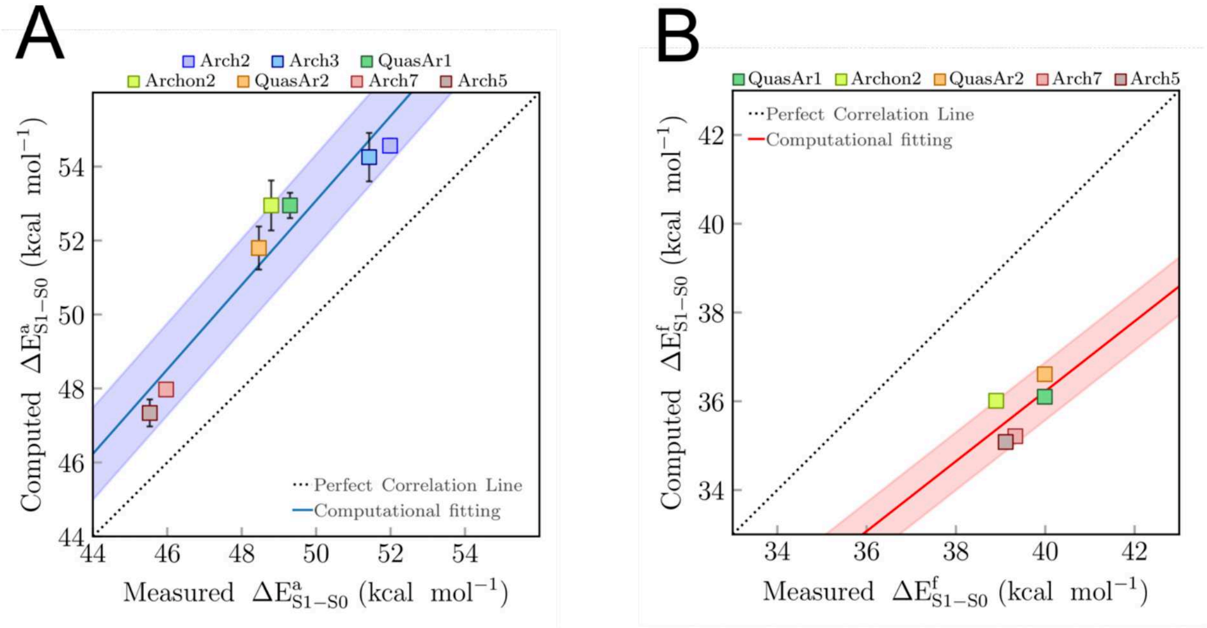


Figure 3.5.1 (A) Comparison between observed and computed values of vertical excitation energies ($\Delta E^{a}_{S1\text{-}S0}$) in kcal mol^{-1} for the wild type rhodopsins Arch2 and Arch3 and the Arch3-based mutants, QuasAr1, QuasAr2, Archon2, Arch5 and Arch7. Standard deviations are given as black vertical segments. Least square regression is also shown. (B) Comparison between experimental and computed fluorescence emission ($\Delta E^{f}_{S1\text{-}S0}$) kcal mol^{-1} for the proteins of the set with reported fluorescence in the DA state.

3.6 On the correlation between FQY and photoisomerization barriers

As anticipated in **Section 3.2**, we will now proceed and demonstrate a correlation between quantities *computed* using the QM/MM models and the *observed* fluorescence quantum yield (FQY) of the Arch set. The FQY expresses how efficiently a molecule converts an absorbed photon into an emitted photon:

$$\mathrm{FQY} = \frac{\text{No. of emitted photons}}{\text{No. of absorbed photons}} \tag{3.6.1}$$

and is a function of the relative rate of radiative and non-radiative de-activation processes:

$$\mathrm{FQY} = \frac{k_r^S}{k_r^S + k_{nr}^S} \tag{3.6.2}$$

where k_r^S is the rate constant for radiative deactivation ($\tau^S=1/k_r^S$ is called radiative lifetime) and k_{nr}^S accounts for all the possible non-radiative decays of excited state molecules from S_1 to S_0. In the present research, we assume that: (i) k_r^S is similar in all closely related Archaearhodopsin homologues and (ii) that k_{nr}^S is determined by the S_1 double-bond isomerization barrier (E^{f}_{S1}) of the chromophore. If i-ii hold, it is then possible to hypothesize the existence of a direct proportionality between E^f_{S1} and FQY. More specifically, FQY is modulated by controlling the access to the S_1 double-bond isomerization path leading to the conical intersection (CoIn) region responsible for S_1 decay. This assumption is supported by the τ^S value of virtually all rhodopsins being estimated in 1-10 ns range by the Strickler-Berg equation. This is several order of magnitude larger than measured τ^S even for highly fluorescent rhodopsins, indicating that $k^S_r << k^S_{nr}$ and, as a working hypothesis, a k^S_{nr} corresponding to an ultrafast chemical reaction. Under the assumptions i-ii, an Arrhenius model would predict E^f_{S1} and FQY to be linked by a logarithmic relationship. However, an Arrhenius model cannot be applied in a non-equilibrium situation such as that of barrierless (or nearly barrierless) isomerization reactions. Accordingly, we do not attempt to fit the E^f_{S1} and FQY proportionality relation based on an Arrhenius model but only demonstrate that the FQY and E^f_{S1} are directly proportional.

3.7 Photoisomerization mechanism

The results from **Section 3.5** above indicate that a-ARM models may be used for investigating the mechanism allowing for different levels of fluorescence in Archaerhodopsin variants. They also indicate that the members of the pairs Arch2/Arch3, QuasAr1/QuasAr2 and Arch5/Arch7 are similar. Accordingly, when appropriate, we have limited the analysis to the representative variants Arch3, QuasAr2 and Arch7. It is useful to recall that rhodopsin function is generally triggered by photon absorption and consequent photoisomerization (**Figure 1.2.1**). As discussed in **Section 3.6**, we assume that CoIn-mediated internal conversion and S_1 fluorescence emissions are the two possible competitive process. This assumption, backed up by experimental findings, allows to establish the S_1 photoisomerization barrier (E^f_{S1}) as the critical quantity determining FQY. In general, the relevant changes occurring in the S_1 PES along the all-*trans* -> 13-*cis* isomerization coordinate of microbial rhodopsins, can be described by two critical geometrical normal modes; upon photoexcitation to S_1, the BLA mode drives the initial chromophore relaxation from the FC point (i), while the reactive torsion associated to the *trans* C_{13}=C_{14} bond (from here on called α), drives the chromophore to the 13-*cis* photoproduct (ii).

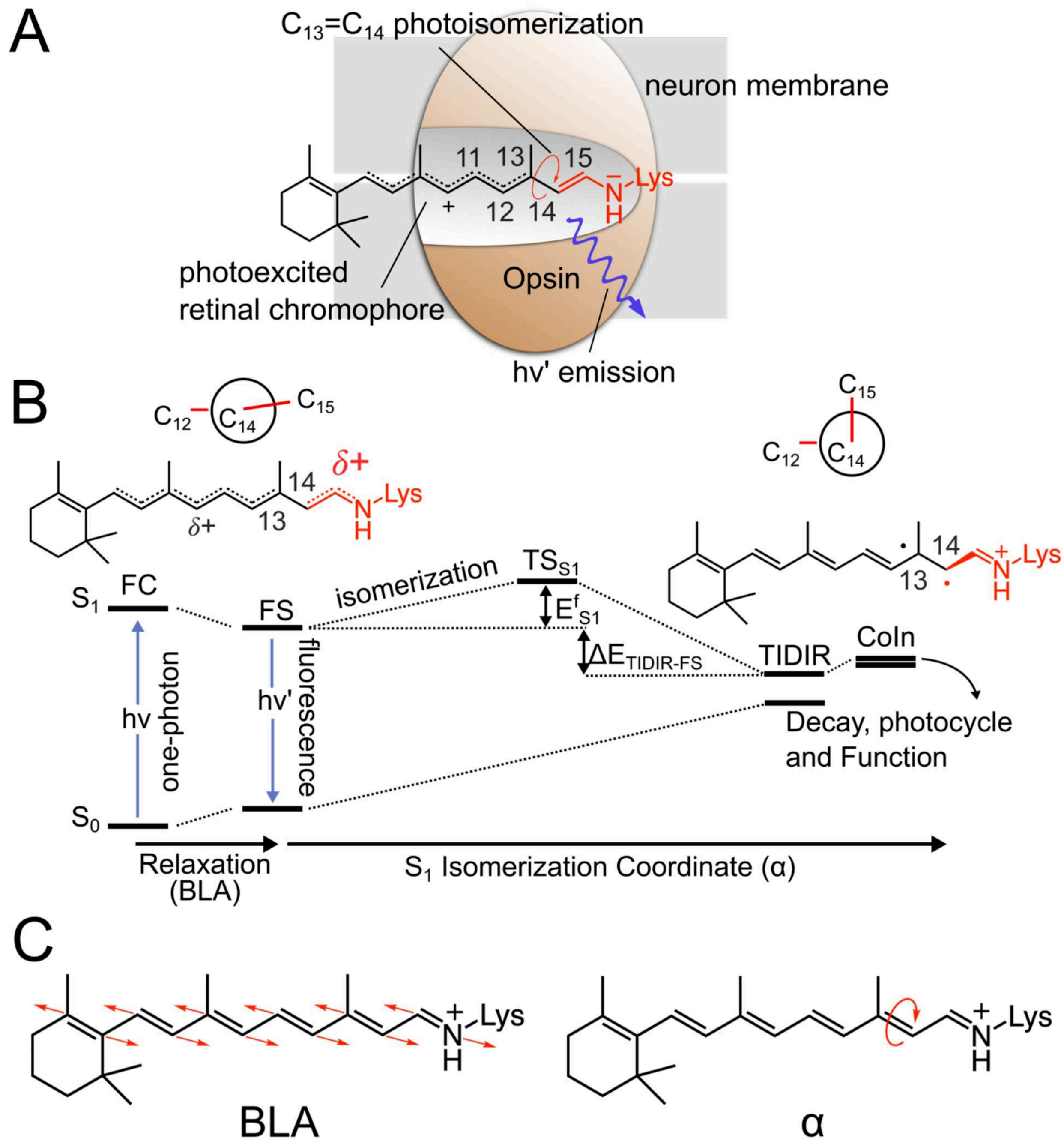


Figure 3.7.1 Photoisomerization mechanism of Archaerhodopsins. (A). Lewis formula representing the initial S_1 chromophore structure. (B). Representation of the chromophore isomerization path. FS corresponds to the fluorescent state. TIDIR represents the photoisomerization channel located near CoIn. FS and TIDIR are represented by Lewis formulas displaying distinct degrees of double bond twisting and charge transfer. (C). Main components of the reaction coordinate. BLA is numerically defined as the difference between the average single-bond length minus the average double-bond length along the C5 to N conjugated chain (for convenience, below we consider the BLA of the framed moiety exclusively). α is defined by the dihedral angle C12-C13-C14-C15.

As shown in (**Figure 3.7.1**), we found that in the Arch set, the BLA initially intercepts a planar, potentially emissive FS minima, while the progressive twisting of α, drives the evolution from FS to a second, twisted, S_1 minima, with diradical electronic character (TIDIR), located in proximity of the decay channel. As detailed below, this exotic intermediate, not documented before in computational studies, offers the possibility to describe the fluorescence enhancement mechanism in the Arch family in terms of the Hammond postulate from organic chemistry[94].

Looking at the S_1 PES reported in **Figure 3.7.2**, it is evident that our MCQC models can reproduce the correct progression of experimental FQY in the Arch set. By comparing Arch7, one of the most fluorescent models, with the wild-type Arch3, we observe an increased slope along the S_1 isomerization coordinate connecting FS to TIDIR. Such increase is evident both at the zeroth-order level ((**Figure 3.7.2A** and **B**, left panels**)**, left panels**)** and after the geometrical perturbative correction at the XMS-CASPT2/CASSCF(12,12)/MM level (**Figure 3.7.2A** and **B**, central panels), named, in short, XMS-CASPT2/MM. In **Figure 3.7.2A** and **B** we also display the evolution of the positive charge on the C14-C15-N moiety for the two models (right panels).

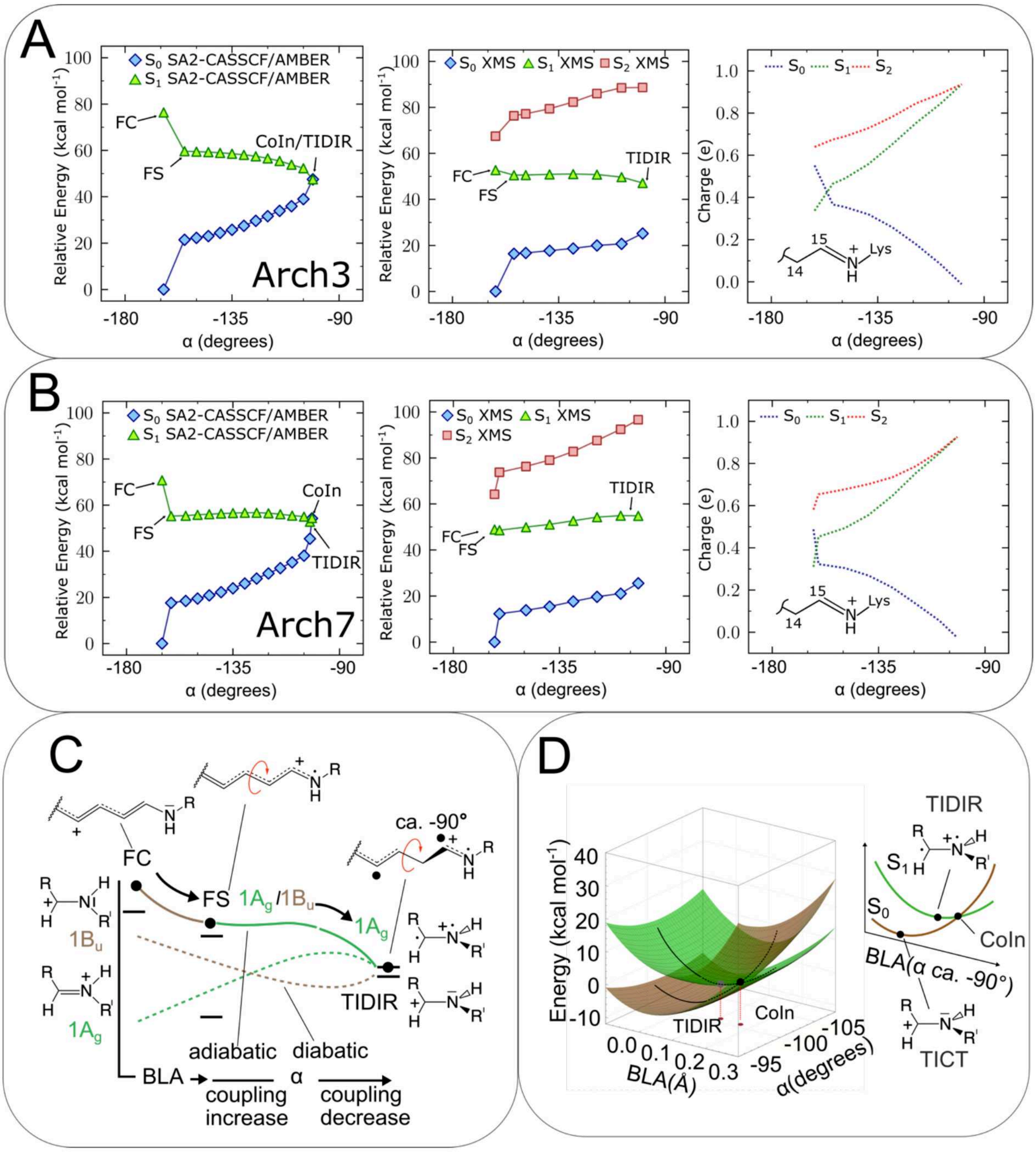


Figure 3.7.2 (A-B) Energy data and charge evolution on the C14-C15-N moiety from the S_1 QM/MM relaxed scans along α (C12-C13-C14-C15 dihedral) for Arch3 (A) and Arch7 (B). Left columns show the reaction paths calculated at the SA2-CASSCF/MM level, central column shows the same calculation at the XMS-CASPT2/MM level and the right column displays the XMS-CASPT2/MM charges evolution. (C) Schematic decomposition of adiabatic energies in diabatic contribution associated to the Lewis Formula of the CH2=NH2(+) iminium cation. (D) Arch7 CoIn representation in the branching space showing the position of TIDIR.

This is important because tracking the evolution of the iminium positive charge at different α values allows to relate changes in the chromophore S_1 electronic structure to a diabatic states representation recently proposed for a set of variants of the GFP protein family named DRONPA2[95]. The Arch set calculated S_1 PES in our models can indeed be described in simplistic valence bond terms as mixing of two pure diabatic states, namely a covalent configuration ($1A_g$) and a charge-transfer configuration ($1B_u$). As discussed in other contributions regarding different rhodopsin models, the $1A_g$ configuration features a C13=C14 double-bond and the positive charge mostly localized on the C14-C15-N fragment, while the $1B_u$ configuration is characterized by a C13-C14 single-bond with the positive charge spread on the π-system resulting in a neutral C14-C15-N moiety (**Figure 3.7.2C**).

When we apply this representation to the S_1 reaction paths of the Arch set QM/MM models, we consistently observe for all the paths (see ref. [69]) that at low values of α (e.g., in the FC/FS region), S_1 is dominated by $1B_u$ character as generally found in other rhodopsins models. However, in opposition to previous theoretical findings on different rhodopsins, as soon as α increases, we found that the $1A_g$ configuration becomes progressively more dominant, leading to a completely pure $1A_g$ character at the TIDIR geometry. In synthesis, as summarized in **Figure 3.7.2C** at α=0°, we have a mixture of $1A_g$ and $1B_u$, with a predominance of $1B_u$, while at α=90° or TIDIR geometry, $1A_g$ is dominant, meaning that due to orthogonality of the π-orbitals, the C13=C14 double bond features two distinct non-interacting radical centers and the positive charge confined on the C14-C15-N moiety. Furthermore, the oscillator strength ($f_{S1\text{-}S0}$), at the TIDIR, is close to 0 (forbidden transition), as expected from a stationary point located close to a CoIn (**Figure 3.7.2D**). Such characterization of the electronic character offers the possibility to rationalize the genesis of E^f_{S1} in terms of diabatic states picture. We hypothesize that along the S_1 isomerization path, as soon as the chromophore relaxes from FC, the $1A_g/1B_u$ coupling increases because of the diminishing S_1-S_0 gap. However, as α becomes progressively more twisted, the coupling diminishes and vanishes as soon as α approaches 90°. As it will be discussed later using a more quantitative approach, we propose that the trending differences in E^f_{S1} across the Arch set, can be mapped to a simple diabatic scheme where E^f_{S1} is a function of α and two parameters: (i) the energy difference of $1A_g$ and $1B_u$ and (ii) the associated coupling.

We should now emphasize that the TIDIR is a stationary point which may exist on the S_1 PES solely because of the topography of the CoIn. As schematized in **Figure 3.7.2D**, lifting the CoIn degeneracy along the BLA-dominated branching vector can intercept a stationary point on S_1 only if the CoIn is sloped. On the contrary, both a peaked and a sloped CoIn topography give rise to a twisted intramolecular charge transfer (TICT) transition state on the S_0 PES, as it has been shown to characterize the thermal isomerization process of bovine rhodopsin.

3.8 Fluorescence mechanism

After having validated the accuracy of our QM/MM models in **Section 3.5** we show that, as anticipated, a direct proportionality between our calculated E^f_{S1} and FQY indeed exists in **Figure 3.8.1**. More importantly we notice that the same proportionality holds when E^f_{S1} is replaced by the relative energy difference of FS and TIDIR ($\Delta E_{TIDIR-FS}$). This finding can be rationalized assuming the Hammond-Leffler postulate from organic chemistry can be applied to the S_1 photoisomerization reaction in our set. In other words, if FS is the reactant and TIDIR is the product, E^f_{S1} magnitude and position (X^f_{S1}) of the transition state (TS_{S1}) depends on the endothermicity of the FS to TIDIR transformation. If the product (TIDIR) is higher in energy than FS, the transition state (TS_{S1}), will resemble the product rather than the reagent (FS). The more TIDIR is destabilized, the more X^f_{S1} will be closer to TIDIR and E^f_{S1} will be higher in energy.

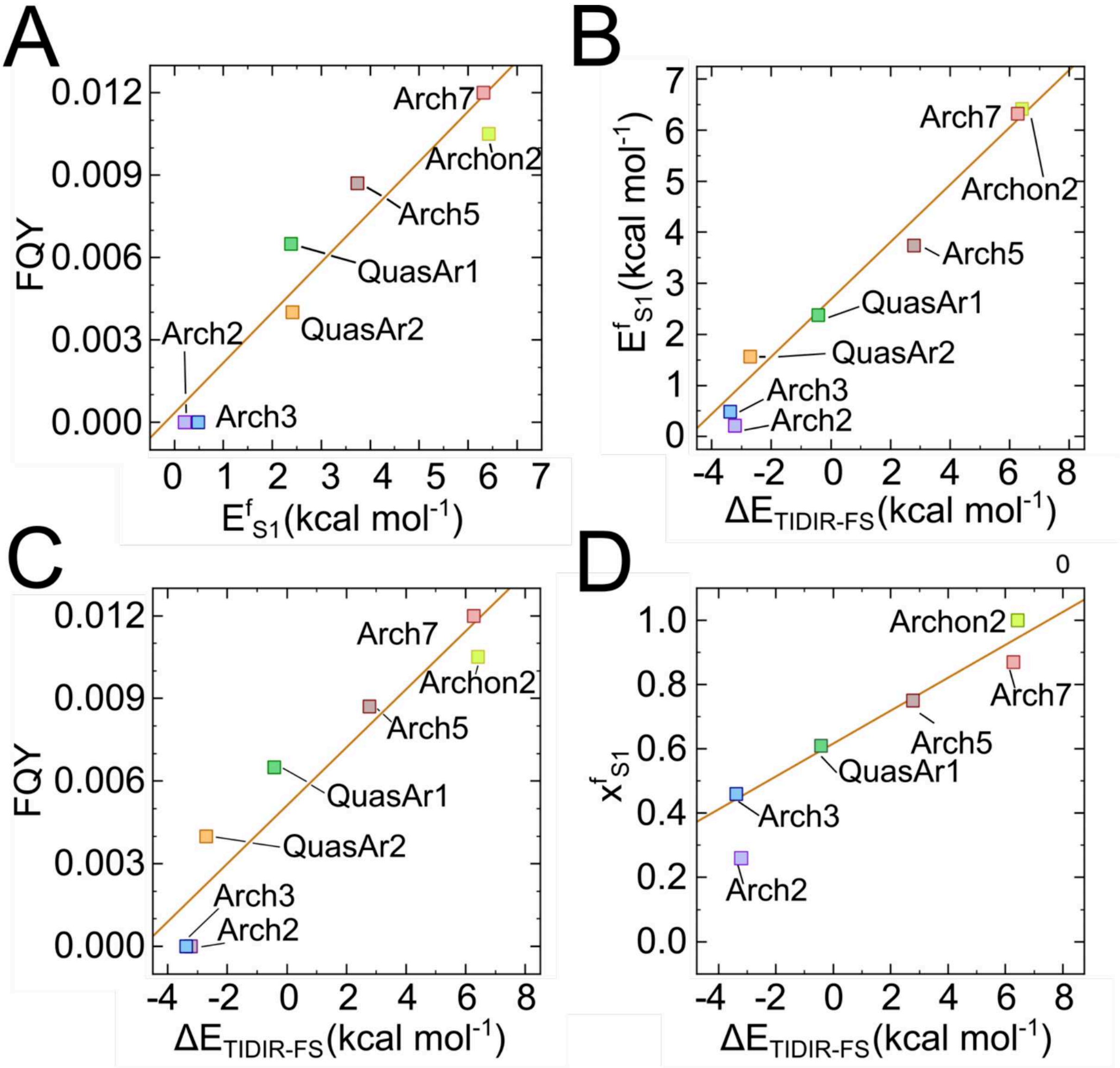


Figure 3.8.1 (A) Linear relationship between E^f_{S1} and FQY and between S_1 isomerization energy ($\Delta E_{TIDIR\text{-}FS}$) and FQY (B). Linear fits are given as orange lines. (C) Linear relationship between E^f_{S1} and $\Delta E_{TIDIR\text{-}FS}$ and between S_1 photoisomerization transition state position (X^f_{S1}) and isomerization energy (D).

To support this finding, recalling the diabatic representation from **Section 3.7**, we build a simple 2 by 2 Hamiltonian to fit the S_0 and S_1 energy profiles of the Arch models as function of α. The diagonal terms of the Hamiltonian are the pure diabatic states ($1A_g$, $1B_u$) and the off-diagonal terms represent their coupling, which is expressed as a simple harmonic function of α that maxes out at 45°. The fundamental assumption of our diabatic scheme is a simple valence-bond fashion description of the adiabatic PES based on two pure resonance formulas, where $1A_g$ represent the covalent formula and $1B_u$ the charge-transfer formula, and the respective characters are tracked by following the evolution of the positive charge at different α values. As shown in **Figure**

3.8.2, the analytical PESs generated using our elementary Hamiltonian model can correctly describe the main features of the QM/MM adiabatic PESs, namely the linear relationship between E^f_{S1} and FQY and the position X^f_{S1} of $TS_{S1.}$

A deeper look into the features of the adiabatic S_1 PES of our models offers a further rationalization of the reciprocal differences between Arch mutants. In **Figure 3.8.1** we observe that the modulation of $\Delta E_{TIDIR\text{-}FS}$ obviously depends on the relative stabilization/destabilization of both FS and TIDIR with respect to the reference S_0 energy at the FC point. Our analysis suggests that the modulation of the FS energy plays the dominant role in Arch2, Arch3, QuasAr1 and Quasar2. As the 0-0 excitation encounters its minimum at Arch5, the TIDIR destabilization becomes the prevalent factor in regulating the $\Delta E_{TIDIR\text{-}FS}$ gap, and it is maximum for Arch7 and Archon2. These data indicates that in the most fluorescent mutants, where FS is already close in energy to S_1 FC, the destabilization of TIDIR and of the corresponding CoIn is the main contributor to the FQY regulation.

To ensure that the observed topological differences across the Arch set, mirrored in the $\Delta E_{TIDIR\text{-}FS}$ value, are providing valuable information on the excited state dynamics of these systems, we propagated 450 fs semiclassical QM/MM S_1 dynamics with no kinetic energy for a representative subsample of the set, namely Arch3, QuasAr1 and Arch7. In agreement with the calculated QM/MM PESs and by extension with the experiments, only the wild-type non-fluorescent Arch3 reacts within the simulation time (**Figure 3.8.3**).

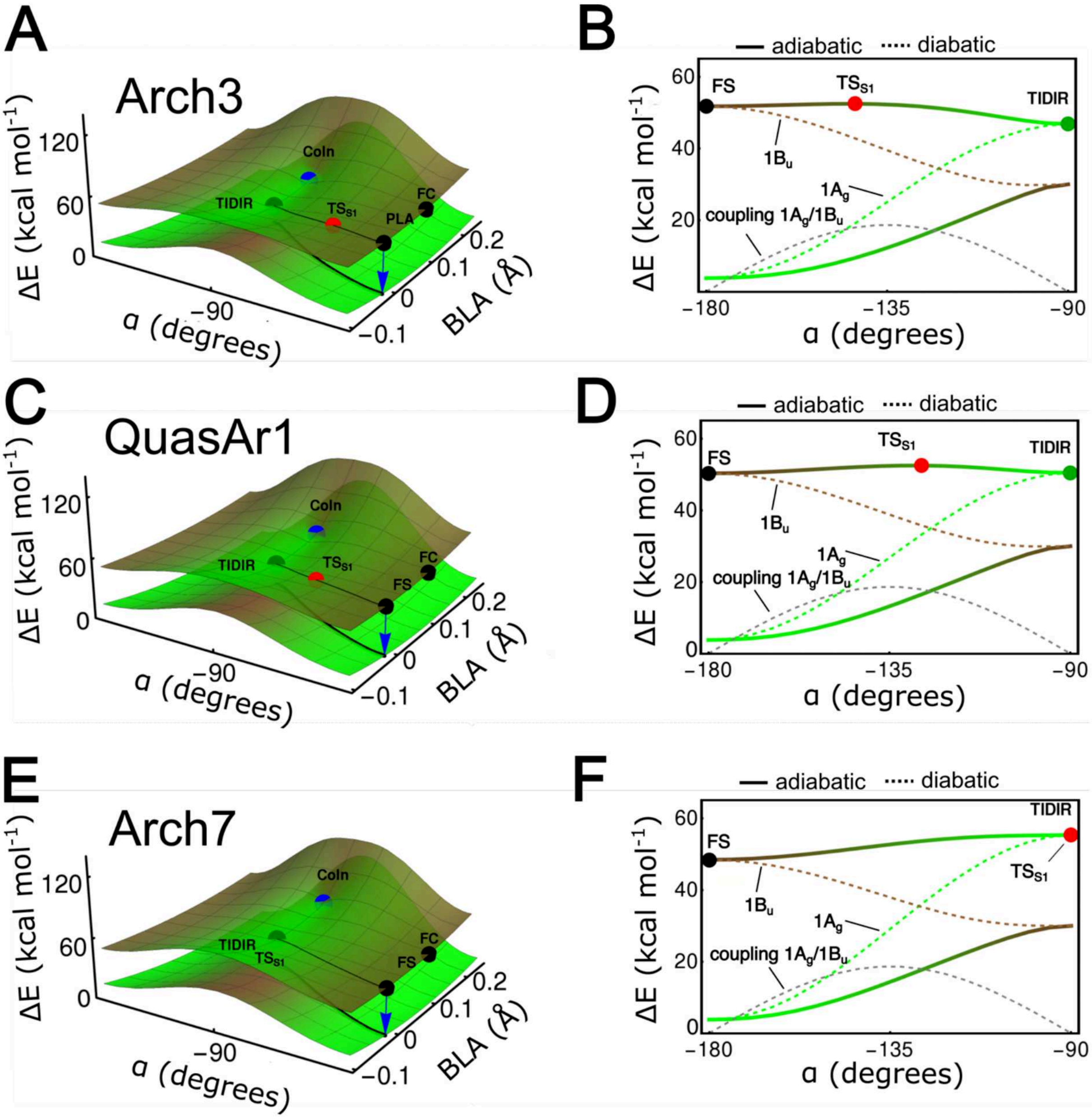


Figure 3.8.2 Diabatic representation of the Arch set PESs.Change in the S_1 and S_0 PESs topography, computed using a 2-state 2-mode model Hamiltonian In the first column we plot the PESs along the α and BLA modes. The color depicts the percentage of diabatic character: brown and green encode the $1B_u$ and $1A_g$ character respectively. Left column: position of the five critical S_1 PES points FC, FS in black, TIDIR in green, CoIn in blue and TS_{S1} in red plotted as a function of the parameter n. Right column: diabatic and adiabatic energy profiles along the S_1 MEP connecting FS to TIDIR and comprising TS_{S1}. The energies of adiabatic states correspond to full lines. The energies of diabatic states are represented with dashed lines. The off-diagonal element, H_{cp}, representing the electronic coupling is shown with a dashed grey line. Here, the diabatic states corresponds to either $1B_u$ (brown) or $1A_g$ (green) electronic characters. The E^f_{S1} and $\Delta E_{TIDIR\text{-}FS}$ values for each term of the series are given and, due to the model parametrization, reproduce the trend computed for the Arch series.

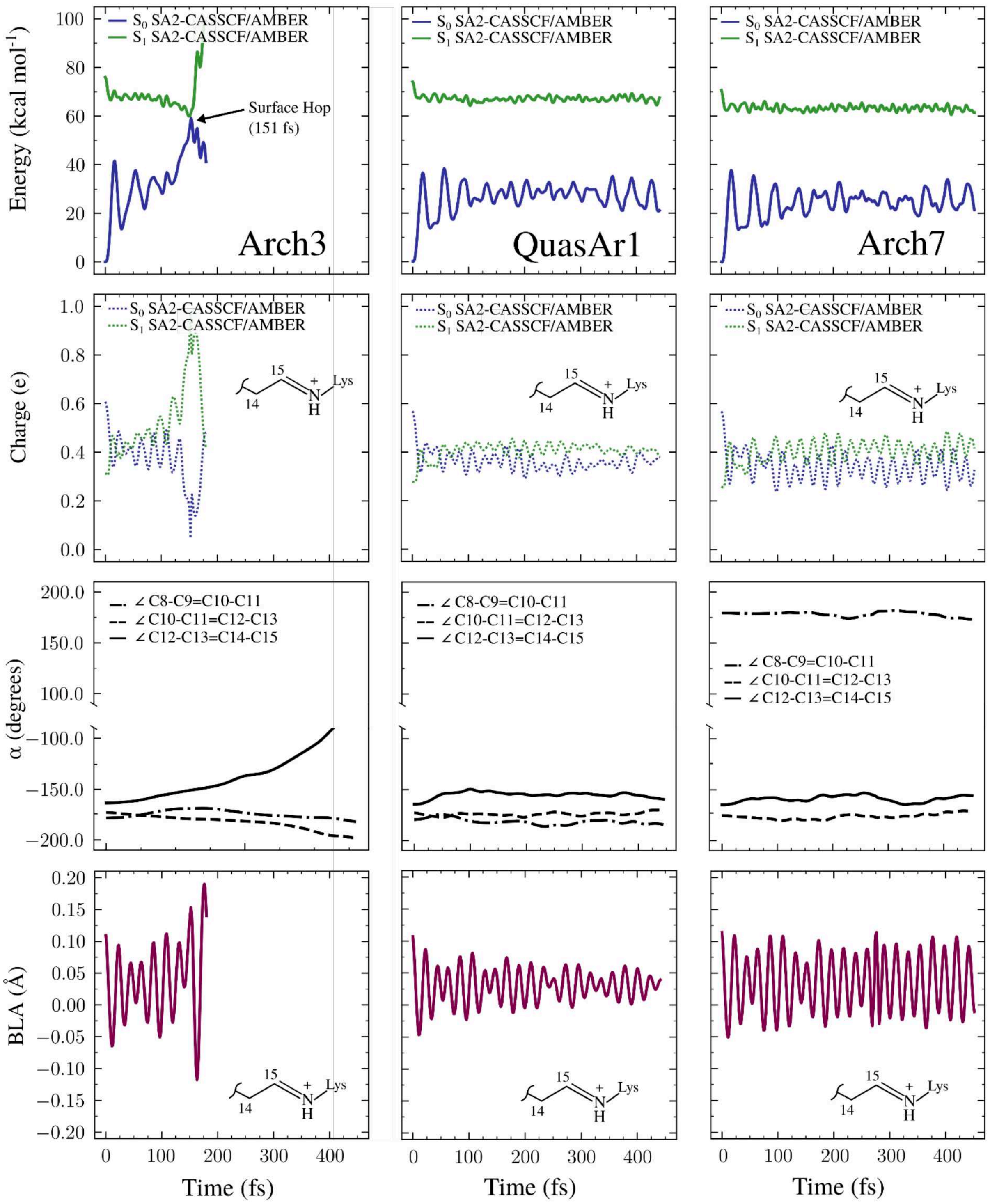


Figure 3.8.3 FC trajectories of the Arch models. Overview of the results extracted from the FC trajectories propagated on the S_1 PES for Arch3 (left column), QuasAr1 (center column) and Arch7 (right column). The energy profiles (top row) are given at the SA2-CASSCF/MM level. The second row displays the evolution of the corresponding fractional Mulliken charge associated to the depicted C14-C15-N moiety for the corresponding wavefunction. The third row shows the time evolution of the main torsional coordinates of the chromophore π-system and include α (the C12-C13-C14-C15 dihedral). A significant deviation from planarity is observed only in correspondence of an isomerization event, as in the case of Arch3 which undergoes all-trans -> 13-cis isomerization. The bottom row shows the evolution of the BLA associated to the same C14-C15-N moiety.

3.9 Molecular determinants of the isomerization barrier

Thus far it was demonstrated that in our MCQM the isomerization energy, or $\Delta E_{TIDIR\text{-}FS}$, correlates with the experimental FQYs for the entire Arch-set. In order to explore how variations in the protein sequence are affect the relation between $\Delta E_{TIDIR\text{-}FS}$ and FQY, we take Arch3 and Arch7 has the limit case of lowest and highest FQY[69].

Since, as discussed in **Section 3.8** and documented in **Figure 3.7.2A-B**, FS and TIDIR are distinct stationary points on the S_1 PES featuring distinct electronic character ($1A_g/1B_u$ at FS and $1A_g$ at TIDIR) mirrored in completely different positive charge distributions, we must conclude that $\Delta E_{TIDIR\text{-}FS}$ can be modified by the opsin electrostatics. Furthermore, from our calculations, it appears that the S_1 electronic character and its evolution along the isomerization coordinate are similar across the Arch set, emphasizing the role of the opsin electrostatic potential (ESP_{opsin}) in the regulation of the $\Delta E_{TIDIR\text{-}FS}$ quantity. To increase $\Delta E_{TIDIR\text{-}FS}$ and consequently FQY, ESP_{opsin} should ideally stabilize FS and destabilize TIDIR.

By looking at the S_1 PES represented **Figure 3.9.1A**, it is evident that in the absence of the protein electrostatics, both Arch3 and Arch7 display a sizable isomerization barrier, imposed by the opsins geometries on the geometrical distortion associated to all-*trans* to 13-*cis* isomerization[69]. However, in the same Figure, it is evident that Arch3 opsin charges acts on this "steric barrier" by stabilizing TIDIR with respect to FS and decreasing $\Delta E_{TIDIR\text{-}FS}$, which, as discussed in **Section 3.7**, is related to low E^f_{S1}. On the contrary, in Arch7, the opsin charges destabilize TIDIR. It is important to notice that the so-called "steric barrier", is the results of the combined effect of steric and electrostatic effects on the chromophore photoisomerization coordinate, both in Arch3 and Arch7 as well as in the other members of the Arch-set (see ref. [69]).

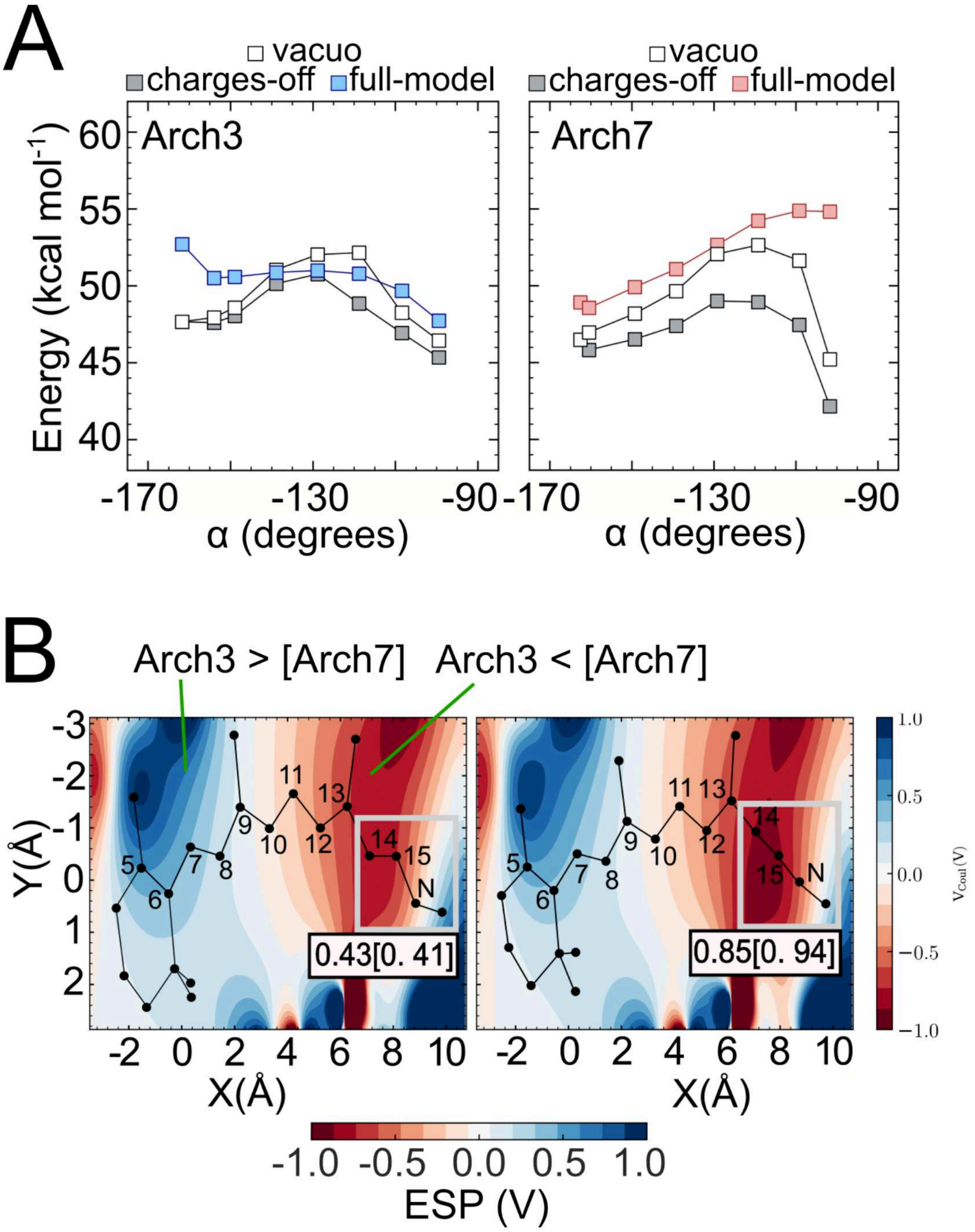


Figure 3.9.1 (A) Effect of the opsin charges on the isomerization energy profile for Arch3 (left) and Arch 7 (right). (B) Two-dimensional plots showing the Arch3-Arch7 electrostatic potential difference at FS (left) and TIDIR (right). The total charge of the framed Schiff-base moiety is also given.

The opposed electrostatics effects on $\Delta E_{TIDIR\text{-}FS}$ documented above, must result from the different side-chain mutations in the two proteins Arch3 and Arch7. After having observed the effect of these mutations from an energetic decomposition standpoint, we propose a further rationalization based on mapping ESP_{opsin} on the chromophore solvent accessible surface. This map provides limited information on the Arch3-Arch7 differences, but rather suggest a complex ESP_{opsin} with a prevalent negative value in the region closer to the PSB. However, by plotting the difference (ΔESP_{opsin}), between the Arch3 and Arch7 electrostatics and mapping it along a cross-section roughly parallel to the chromophore π-skeleton (see **Figure 3.9.1B**), we notice that Arch3 preferentially stabilize the positive charge fully confined on the SB moiety of TIDIR while the opsin of Arch7 better stabilizes the delocalized charge spread towards the β-ionone of FS[69]. These effects can be also interpreted using the Hamiltonian model discussed in **Section 3.8**, focusing in particular on the electronic character of the TIDIR structure. As the positive charge fully confined on the C14-C15=N moiety is typically associated to the $1A_g$ diabatic character, Arch3 more negative electrostatics projected in this region must stabilize TIDIR, as opposed to the Arch7 electrostatics, which is less negative in this region and destabilizes it consistently with the diabatic picture of **Figure 3.8.2**.

3.10 Conclusions

Generating more and more efficient rhodopsin-based GEVIs, is a challenge that nowadays is mostly addressed via experimental brute force approaches, where tuning the protein fluorescence is performed via round of random and/or directed mutagenesis. In this work, we have investigated the increase in fluorescence quantum yield in the Arch set by constructing MCQC. The most significant findings can be summarized as follows: (i) in Arch3 and its variants, the FQY is primarily determined by the competition between S_1 emission and isomerization extending the mechanism reported for DRONPA2[95] to retinal proteins[69], (ii) the rate of isomerization depends by $\Delta E_{TIDIR\text{-}FS}$ quantity, and in particular by the destabilization with respect to FS of the newly reported TIDIR intermediate and (iii) the TIDIR stabilization/destabilization is modulated by the protein electrostatics, rather than geometry.

The atomistic mechanism for the TIDIR stabilization appears to be a consequence of the cavity electrostatics on the distinct charge distributions of the FS and TIDIR chromophores. As schematically (top schemes) and computationally (bottom plots) displayed in Figure 4E the chromophore positive charge is largely delocalized in FS but confined on the C_{14}-C_{15}-N fragment in TIDIR. In this situation the described change in electrostatics from Arch3 to Arch7 leads to a reduced stabilization of the Schiff base confined with respect to the delocalized charge. In simple terms, the effect of the amino acid (and atomic point charges) replacement resulting in the Arch3 to Arch7 progressive TIDIR destabilizing along the Arch-set, can be interpreted as the gradual relocation of a negative charge from the Schiff base region to a region closer to the β-ionone ring.

The presented mechanism has both methodological and biological implications. The first is related to the fact that TIDIR and FS, being energy minima on the S_1 potential energy surface, are computationally fast to locate. It is thus possible to envision the development of a tool for the *in silico* selection of highly fluorescent Arch3 variants. Such tool would be based on the automated[33] construction of MCQC models and the calculation of the corresponding $\Delta E_{TIDIR-FS}$ value to be maximized via *in silico* mutational experiments. The biological implication is instead related to the hypothesis that the changes in sequence, and thus electrostatics, leading to a negligible Arch3 fluorescence could have occurred also in nature through a natural selection process aimed at increasing the protein photoisomerization rate and improve the protein function. We cannot exclude that the same mechanism plays a role in other microbial rhodopsin evolution. On the other hand, an inverse natural selection process leading to the suppression of the photoisomerization, and maximization of the fluorescence output may have generated the recently discovered Neorhodopsin from *Rhizoclosmatium globosum*[96] that displays an intense fluorescence with a ca. 0.2 FQY and therefore much closer to that of DRONPA2 and other optimized green fluorescent protein variants.

Chapter 4

Rhodopsin color-tuning via counterion migration-delocalization

In the previous chapter, I discussed the computational tools allowing to investigate the excite state property of a rhodopsin and provided a theoretical justification to the increased FQY exhibited by Arch3 variants with respect to the wild type. Although enhancing fluorescence of microbial rhodopsins is a highly desirable technological feature in optogenetics, scientists are looking for variants which are at the same fluorescent and feature absorb light in the red or near-infrared (NIR). As discussed in **Section 1.4**, GEVIs absorbing light in the red of the visible spectra are well-suited to work in parallel with actuators and pose less challenges to fluorescence microscopy. Since most of the protein of the Arch family already exhibits increased FQY and red-shifted λ^{a}_{max}, I extended the analysis discussed in **Chapter 3** to understanding the rules of spectral tuning in the same sub-family of rhodopsins. Interestingly, in the following, I demonstrate that instead the modulation of spectral tuning and fluorescence appear to be correlated, as it is speculated in recent experimental literature[43].

4.1 Abstract

The understanding of the electrostatic rules controlling the spectral tuning of the retinal chromophore of rhodopsins is still unsatisfactory. This issue originates from the overwhelming structural complexity of the opsin cavity hosting the chromophore. The fact that the cavity may incorporate the chromophore counterion in different positions and that the remaining polar residues may have different orientations in response to both electrostatic and steric factors leads to a large number of patterns whose electrostatic effect is not obvious. In this work we use a computational model of a selected rhodopsin to study the effect of the hypothetic "delocalization" of the chromophore counterion on both the chromophore geometry and excitation energies. By applying a variational optimization protocol to the model, we demonstrate that both the counterion position and its "effective delocalization" are primary factors in controlling the chromophore vertical excitation energy and that an extended

delocalization is necessary to achieve a low excitation energy We conclude our analysis by showing that the same model can be used to derive a first-order relationship between λ^{a}_{max} and chromophore isomerization barrier that correctly reflect the observed trends relative to a set of Arch3-based optogenetic tools.

4.2 Introduction

After the discovery of Archaerhodopsin-3 (Arch3) from *Halorubrum Sodomense*[32,49], microbial rhodopsins have gained popularity as potential optogenetic fluorescent reporters for the visualization of neuronal action potentials[41]. However, Arch3 maximally absorbs 556 nm light and therefore relatively far from the desired red-light absorption. In fact, the use of red-light to promote fluorescence is a desired technological feature due to the low phototoxicity and higher tissue penetration of this wavelengths. For this reason, scientists have been looking for natural and artificial rhodopsins with redder and redder maximum absorption wavelength ($\lambda^{\mathrm{a}}_{\mathrm{max}}$)[46,50–52]. On the other hand, the attempts to engineer rhodopsins with a red-shifted $\lambda^{\mathrm{a}}_{\mathrm{max}}$, it has only been partially successful as the understanding of the, so called, color tuning mechanism is still incomplete.

The color tuning mechanism of rhodopsins has been studied for a long time and several coupled factors have been found to play a role in its regulation. Here we focus on the complex interaction of the charges carried by opsin amino acidic matrix with the retinal-protonated Schiff base (rPSB) chromophore. Such electrostatic tuning has been extensively investigated experimentally by using retinal analogs[97–99], via site-directed mutagenesis[100–106] and more generally via structural analysis[107,108]. Nakanishi and coworkers were the first to propose a general external point charge model to explain the experimental observations[109]. According to this model, a negative charge in proximity of the β-ionone ring of the retinal[109,110] is key to achieve red $\lambda^{\mathrm{a}}_{\mathrm{max}}$ (See also **Section 1.5.2**). More recently, in order to get a deeper understanding of electrostatic tuning, Borhan, Geiger and coworkers used, as a lab model, a set of synthetic rhodopsin mimics based on human cellular retinol-binding protein II (hCRBPII)[110,111]. The preparation and analysis of a set of mimics displaying $\lambda^{\mathrm{a}}_{\mathrm{max}}$ values from 425 to 644 nm indicates that the reddest $\lambda^{\mathrm{a}}_{\mathrm{max}}$ can be obtained by projecting an evenly distributed negative electrostatic potential (ESP) on the rPSB[63,110,111]. The authors stress the fact that a delocalized and negative cavity ESP would stabilize an iminium positive charge maximally delocalized along the chromophore backbone (i.e. one would constitute a negative replica of the other).

A computational approach should, in principle, provide a deep understanding of the molecular level factors determining the specific λ^{a}_{max} of rhodopsins. However, even if the range of such studies is wide when compared to their experimental counterpart, these are focused on determining specific steric and electrostatics contributions to the λ^{a}_{max} in specific rhodopsins or rhodopsin variants. These analyses, although well suited to capture specific λ^{a}_{max} effects, do not capture the general relationship between the ESPs structure and large bathochromic shifts as the one found in the recently reported wild-type Neorhodopsin, featuring a near infrared absorption of 690 nm[96]. Furthermore, no reported computational study has addressed the qualitative conclusions that a negative charge delocalization, not only a change in the charge position, is an important color-tuning mechanism. In this work, we attempt to fill this void.

The present investigation is based on an unusual quantum mechanics / molecular mechanics (QM/MM) model of Arch3 where (i) the charges on all chromophore cavity residues has been zeroed and (ii) a unit negative charge is not only free to migrate from one residue to the next, but it can fragment and distribute on different cavity residues. We will show that, when used together with a specific conjugated-gradient optimization protocol, the model is able to simultaneously describe the λ^{a}_{max} effect of charge migration and delocalization defining the Nakanishi and Borhan-Geiger models. More specifically, the model is used to carry out a set of computational investigations based on the assumption that the total counterion charge is always represented by a unit negative charge free to transit from localized to largely delocalized configurations. It is found that the preferred counterion configuration for red-light absorbing models as well as for models maximizing the excited state lifetime is always highly delocalized.

The contribution is organized as follows:

(i) Details of the Arch3 QM/MM model construction.

(ii) The aim of optimizing the distribution of the negative main counterion charge in Arch3 cavity binding pocket is discussed, together with the technical details of the optimization algorithm.

(iii) Theoretical justification of the methodology employed. It is shown, with a simple model system of a rhodopsins cavity, how the main counterion electrostatic potential diffusion operated by dipoles can be qualitatively mimicked by a system of fractional negative charges.

(iv) The proposed theoretical framework outlined by our computational model is compared to the existing electrostatics tuning theories by Borhan-Geiger and Nakanishi.

(v) A correlation between $\lambda^{\mathrm{a}}_{\mathrm{max}}$ and photoisomerization barrier is pointed out, and a mechanistic explanation for the recent experimental finding of red-shifted absorbing fluorescent microbial rhodopsins is provided.

4.3 Arch3 model structure

Due to its applications in optogenetics, Arch3 has been extensively studied experimentally. In the last decade scientists have put considerable effort in the optimization of Arch3's spectral properties, particularly in identifying variants displaying both red-shifted $\lambda^{\mathrm{a}}_{\mathrm{max}}$ and increased fluorescence quantum yield (FQY), in order to develop improved genetically encoded voltage indicators (GEVIs), or molecules which are able, via fluorescence emission, to mark the passage of action potential (AP) across the neuronal membrane[32,49,112,113]. Tuning $\lambda^{\mathrm{a}}_{\mathrm{max}}$ represents a technologically appealing feature, as it allows to use trigger fluorescence with light which is less likely to interact with cell tissues. Since these efforts culminated in the discovery of several Arch3's variants with red-shifted $\lambda^{\mathrm{a}}_{\mathrm{max}}$, and in most cases increased FQY, we deem that Arch3 represents the best platform to explore electrostatic tuning of $\lambda^{\mathrm{a}}_{\mathrm{max}}$.

In general, to study rhodopsins spectral properties *in-silico*, a quantum-mechanical multiconfigurational approach is mandatory (see **Chapter 2**). However, multiconfigurational studies of systems of a few dozens of atoms are already challenging, such that hybrid quantum-mechanics (QM) molecular-mechanics (MM) schemes are generally preferred due to the reduced computational load. For this reason, we started our electrostatic tuning study of Arch3, by generating its corresponding QM/MM model. For our simulations to be as realistic as possible, we constructed Arch3's QM/MM model using the well resolved dark-adapted (DA) crystallographic structure reported by Bada-Juarez et al. (PDB ID: 6GUX)[32]. Starting from 6GUX, a QM/MM model was automatically constructed using the previously reported advanced automatic rhodopsin modelling (*a*-ARM) protocol, a software which has been indicated to build fast and congruous and reproducible QM/MM models of rhodopsins. The *a*-ARM QM/MM model of Arch3, or reference for all the successive calculations, is a gas-phase globally uncharged monomer optimized on the ground state (S_0) at the single-state CASSCF(12,12)/6-31G*/MM (in short SA2-CASSCF/MM) level of theory.

4.4 General aims and optimization algorithm

Starting from this QM/MM model of Arch3, we propose to study the electrostatic modulation of $\lambda^{a}_{\mathrm{max}}$ from a totally new perspective from what has been previously reported in the literature. Instead of looking at differential molecular contributions to variation in $\lambda^{a}_{\mathrm{max}}$ among different models, we reverse engineered the charge distribution (and consequently the ESP) of the Arch3 retinal binding pocket to reproduce specific $\lambda^{a}_{\mathrm{max}}$ values. To preserve a realistic electrostatics environment, only the MC charge (in our QM/MM model the aspartic acid residue D222) was allowed to fraction or displace to other cavity residues to modulate the opsin ESP in order to fit different $\lambda^{a}_{\mathrm{max}}$ values. We show that, in agreement with the experiments on hCRBPII, and in contrast to the external point charge models, the negative charge features a homogeneously increased spreading and fractioning going to redder and redder $\lambda^{a}_{\mathrm{max}}$.

Briefly, the algorithm works as follows; we start from the S_0 optimized structure of Arch3, select a target vertical excitation energy value ($\Delta E^{*,a}_{S1-S0}$) and optimize the electrostatics of the cavity such that the difference between the absolute values of ΔE^{a}_{S1-S0} and $\Delta E^{*,a}_{S1-S0}$ is minimized. In other terms, we define ΔE^{a}_{S1-S0} as a function of the cavity charge distribution, or $\Delta E^{a}_{S1-S0}(\mathbf{q})$ (**Figure 4.4.1A**). To achieve this goal, the cavity residues charges of the Arch3 QM/MM model cavity are represented by the sum of their MM force field partial charges (q_i), which is 0 for neutral residues and +1 or -1 for charged residues. Since the cavity binding pocket defined by the *a*-ARM protocol (**Figure 4.4.1B**) does not include charged residues except for MC, we can conveniently represent the model virtual MC charge distribution by a vector ($\mathbf{q}$ = q_0, q_1, …, q_N, where N is the number of residues in the chromophore binding pocket), whose elements are bound to host a negative charge comprised between 0 and -1. Thus, at the initial step of the optimization process, $\mathbf{q}$ will always be represented by a vector which has N-1 elements equal to 0 and one entry corresponding to the original main counterion, equal to -1. Furthermore, we impose that the global charge of the cavity $(\mathbf{q})$ must always equal -1 (as in the starting QM/MM model), to make sure that the final optimized model is as realistic as possible.

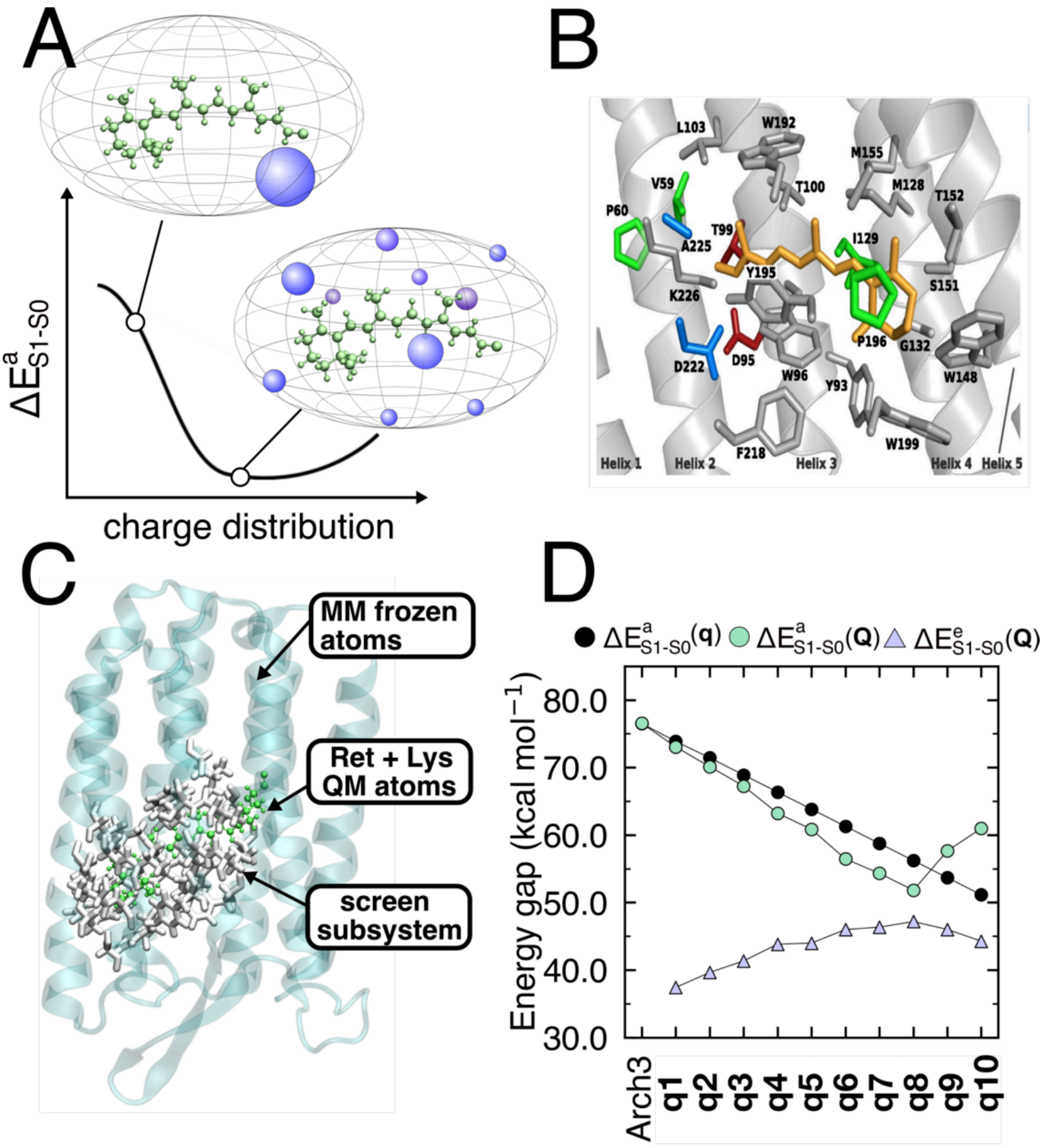


Figure 4.4.1 (A) Illustration of the methodology discussed in this work. The theoretical minimum in λ^a_{max} can be explored by optimizing the environment charge/electrostatics to reproduce a target $\lambda^a_{max.}$ (B) Cavity of Arch3. The corresponding residues constitute the screen subsystem of following panel C. Adapted from ref. 46 (C) Depiction of the systems employed in the optimization procedure. (D) Diagram showing vertical excitation energies obtained via optimization of different distributions of negative

charge in black circles, and after geometrical relaxation on S_0 at the same charge distribution in green circles.

As summarized in **Figure 4.4.1C**, the QM/MM model is therefore partitioned in 3 subsystems; the atoms treated at the multiconfigurational QM level, or the retinal chromophore and linker lysine sidechain, a '*screen subsystem*', represented by the cavity binding pocket residues, or the entries of vector $\mathbf{q}$, and the MM opsin atoms which are frozen at their crystallographic position as electrostatic and geometrical constraints to enforce the conservative design of the simulation. Given this set of rules, the problem of finding the $\Delta E^{*,a}_{S1-S0}$ value, can be formulated as a simple constrained optimization problem as follows:

$$\begin{gathered}\min\left(\left(\Delta E^{a}_{S1-S0}(\mathbf{q}) - \Delta E^{a,*}_{S1-S0}(\mathbf{q})\right)^2\right)\\ \text{subject to}\sum_{i}^{N} q_i = -1\,;\; q_1, q_2, \dots, q_N \le 0\end{gathered} \tag{4.4.1}$$

When a residue hosts a fraction of negative charge different from 0 such fraction is equally distributed amongst the atom of the residues. Since ΔE^{a}_{S1-S0} is calculated at the SA2-CASSCF/MM level and no gradient of the CASSCF wavefunction is available with respect to the charges, the minimization is performed by computing the gradient numerically, with the following two-point formula:

$$\frac{\left(\Delta E^{a}_{S1-S0}(\mathbf{q}) - \Delta E^{a,*}_{S1-S0}(\mathbf{q})\right)^2 - \left(\Delta E^{a}_{S1-S0}(\mathbf{q}+\Delta\mathbf{q}) - \Delta E^{a,*}_{S1-S0}(\mathbf{q})\right)^2}{\Delta\mathbf{q}} \tag{4.4.2}$$

where $\Delta\mathbf{q}$ is set to 0.001. The optimization of $\mathbf{q}$ regards only the 'screen subsystem' and is carried out at fixed geometry with $\Delta E^{a}_{S1-S0}(\mathbf{q})$ evaluation carried out via single point energy calculations. In practice, this approach relies on an interface built by the author and exploiting the SciPy built-in trust region algorithm driver[114] to minimize $\Delta E^{a}_{S1-S0}(\mathbf{q})$ and find optimal $(\mathbf{q})$ provided with the gradients calculated as in **Equation 4.4.2** by [Open]Molcas[115] via QM/MM single point calculations at the SA2-CASSCF/MM level.

4.5 Mathematical framework

Although the computational setup is designed to preserve the structure of Arch3's cavity and generate configurations with realistic electrostatics, it could be objected that residues hosting fractional charge may not represent correctly a system of dipoles (neutral cavity residues) interacting with a negatively charged main counterion. For this reason, we propose a simple mathematical justification of our methodology based on a minimal example. As shown in **Figure 4.5.1A**, the cavity of Arch3 may be schematized as follow; (i) the MC is represented by a negative point charge at the origin of the system, (ii) the cavity residues are represented by two identical dipoles symmetrically positioned to left and right to the central charge (iii) and finally the retinal is exemplified by a horizontal line 5Å far from the dipoles plus charge system. This setup, where the dipoles mimic the solvation effects of amino acidic residues side chain on the main counterion, is called the *biological model*, as it is representative of a generic realistic biological rhodopsin cavity. By elongating and shortening the bond of the dipoles, it is evident from **Figure 4.5.1A** how the dipoles have the effect of diffusing the electrostatic potential generated by the unperturbed negative charge. In **Figure 4.5.1B**, using the same minimal setup, is it shown that it is indeed possible to find a fractional distribution of the negative charge that allows to replicate, with qualitative accuracy, the diffusion effect operated by the cavity dipoles on the electrostatic potential projected on the retinal chromophore. Since the scheme is symmetric, also the fractioning necessary to reproduce the electrostatic potential is symmetric. As explained above, this model, which we call the *chemical model*, has the desirable feature of being easily manipulated through an optimization approach allowing to simply diffuse or relocate a negative charge to optimize a target property, which would imply a more complex and involved setup for the "*biological model*".

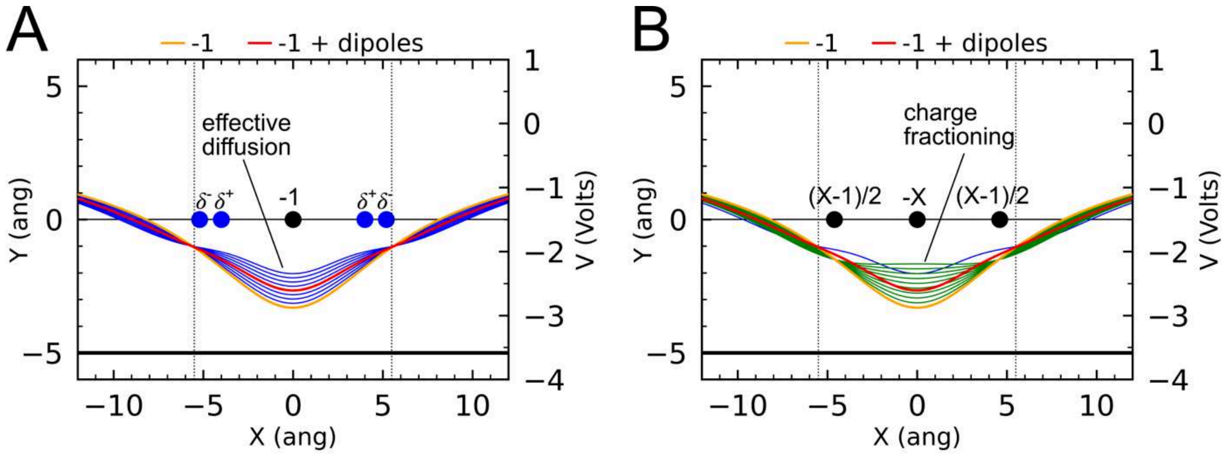


Figure 4.5.1 (A) Effect of dipoles on the ESP projected by a negative point charge (orange line). The two dipoles are horizontally aligned to the left and right of the central negative charge and their parameters (charge and bond length) are taken from the carbonyl group of acetone after an HF/6-31G* calculation. The ESP projected by the negative charge plus dipole system is represented by the red line. By shortening and elongating the dipole bond, the negative charge ESP is diffused (blue lines). (B). The negative charge plus dipoles system is substituted by three negative fractional charges summing up to -1. Different fractioning patterns can qualitatively reproduce the diffusion effect on the negative charge ESP by the dipoles (green lines).

4.6 Algorithm application and results

To test our algorithm, we selected a range of 10 $\Delta\mathrm{E}^{\mathrm{a},*}_{\mathrm{S1-S0}}$ values, spanning from 76.5 kcal mol^{-1} to 51.2 kcal mol^{-1} (e.g., a range of 25 kcal mol^{-1} with spacing of ca. 2.5 kcal mol^{-1}) and for each one of the $\Delta\mathrm{E}^{\mathrm{a},*}_{\mathrm{S1-S0}}$, optimized a charge distribution $\mathbf{q}$. As shown in **Figure 4.4.1D**, for all the 10 cases, the optimization converged successfully with 10 different solutions $\mathbf{q}$ which will from now on be labeled as $\mathbf{q_1}$ to $\mathbf{q_{10}}$. However, identifying a distribution of the MC negative charge yielding the target $\Delta\mathrm{E}^{\mathrm{a},*}_{\mathrm{S1-S0}}$ values at fixed geometry is not sufficient to demonstrate that the new electrostatics corresponds to an hypothetical DA structure with that particular $\Delta\mathrm{E}^{\mathrm{a},*}_{\mathrm{S1-S0}}$, as the CASSCF wavefunction has been optimized only with respect to the charges and not with respect to the nuclear coordinates. To ensure that our distributions ($\mathbf{q_1}$ to $\mathbf{q_{10}}$) generate ground state (S_0) minima spanning the originally designed 25 kcal mol^{-1} $\Delta\mathrm{E}^{\mathrm{a},*}_{\mathrm{S1-S0}}$ range, for each of the 10 QM/MM models with $\mathbf{q}$ distribution, we geometrically relaxed the atoms of the screen and QM subsystems on S_0 to obtain 10 new structures which now differ in $\mathbf{q}$ and geometry ($\mathbf{r}$). We introduce for these new models a new notation which accounts both for the virtual MC charge distribution and geometry as $\mathbf{Q_1}$ to $\mathbf{Q_{10}}$.

Figure 4.4.1D shows that, for a specific $\mathbf{q}$, $\Delta E^{a}_{S1-S0}(\mathbf{Q})$ shows a generally red-shifted behavior with respect to $\Delta E^{a}_{S1-S0}(\mathbf{q})$ until $\Delta E^{a}_{S1-S0}(\mathbf{q_8})$ = 56.2 kcal mol^{-1}, whereas the difference is the opposite for $\mathbf{q_9}$ and $\mathbf{q_{10}}$ with $\Delta E^{a}_{S1-S0}(\mathbf{Q_9})$ and $\Delta E^{a}_{S1-S0}(\mathbf{Q_{10}})$ which are blue-shifted with respect to the non-geometrically relaxed counterpart. As shown in **Figure 4.6.1** a close inspection of models $\mathbf{Q_9}$ and $\mathbf{Q_{10}}$ suggests that at extremely red-shifted ΔE^{a}_{S1-S0} values, the electrostatics imposed is so strong that the retinal protonated Schiff base nitrogen atoms undergoes a complete pyramidalization, justifying the inversion of the trend observed in **Figure 4.4.1D**. For this reason, models corresponding to $\mathbf{Q_9}$ and $\mathbf{Q_{10}}$ will not be discussed in subsequent analysis. We also notice that the final $\Delta E^{a}_{S1-S0}(\mathbf{Q})$ range, going from 74.0 kcal mol^{-1} to 51.8 kcal mol^{-1}, has a similar span (23.2 kcal mol^{-1}) to the originally proposed $\Delta E^{a,*}_{S1-S0}$ range, and we shall from now on discuss, without loss of generality, the properties of the $\mathbf{Q}$ models, which are more realistic in terms of mimicking potentially existing DA states of rhodopsin systems.

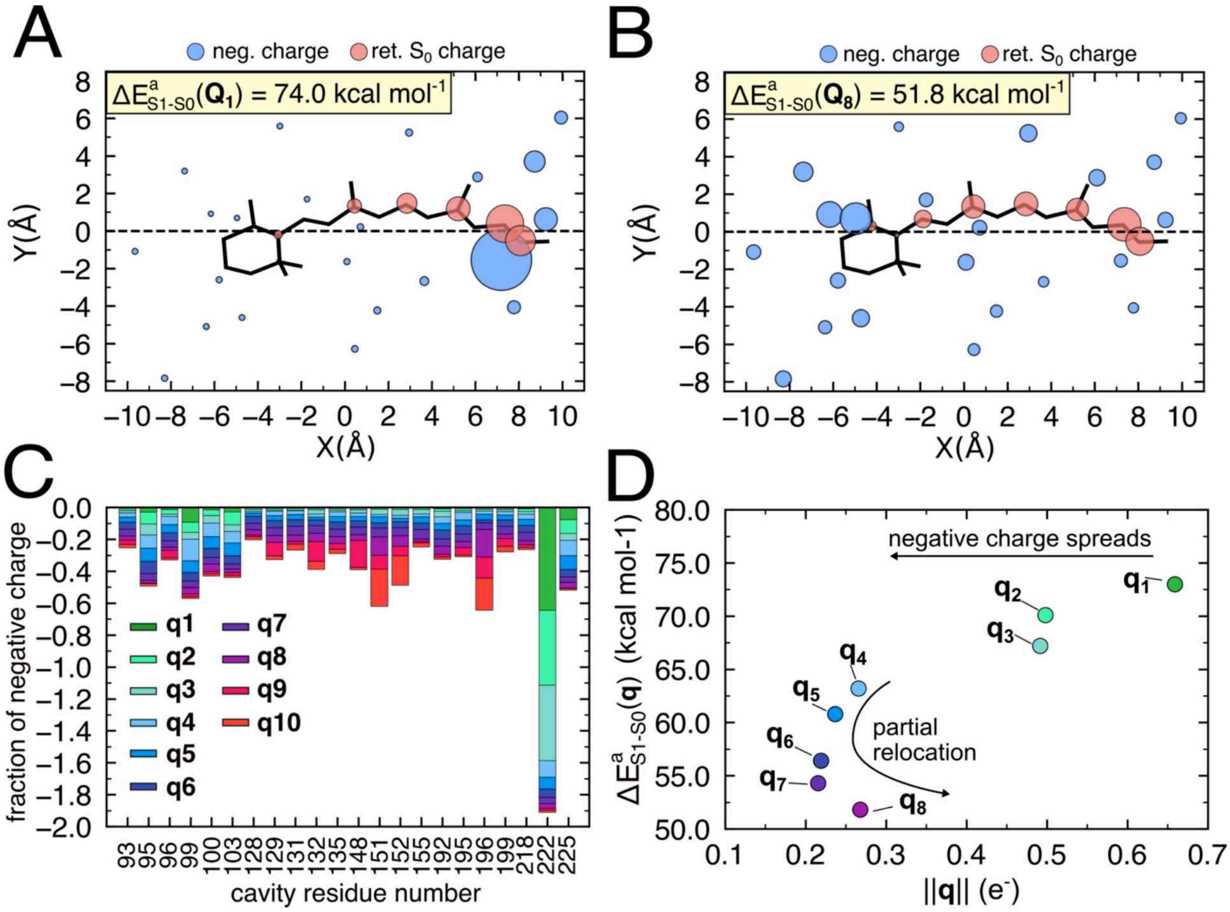


Figure 4.6.1 (A) and (B) show how the cavity negative charge is fractioned among the cavity residues (blue circles) and the localization of chromophore S_0 positive charge (red circles) for the structures corresponding to $\Delta E^a_{S1\text{-}S0}(Q_1)$ = 74.0 kcal mol^{-1} and $\Delta E^a_{S1\text{-}S0}(Q_8)$ = 51.8 kcal mol^{-1}. (C) Histogram plot showing the charge distributions (q_1 to q_{10}) at the end of the optimization procedure targeting the 10 $\Delta E^{a,*}_{S1\text{-}S0}$ values from 76.5 (q_1) to 51.2 kcal mol^{-1} (q_{10}) (D) Norm of the same charge distributions as indicator of the diffusion of the MC charge.

In **Figure 4.6.1A** and **Figure 4.6.1B** we compare the virtual MC charge distribution and chromophore SA2-CASSCF/MM S_0 charge associated to the larger and smaller $\Delta E^a_{S1-S0}(\mathbf{Q})$, or the models with the most blue and red-shifted absorption among the optimized $\mathbf{Q}$, which are, namely, $\Delta E^a_{S1-S0}(\mathbf{Q_1})$ = 74.0 kcal mol^{-1} and $\Delta E^a_{S1-S0}(\mathbf{Q_8})$ = 51.8 kcal mol^{-1}.

The virtual MC charge for the blue-shifted model is largely localized on the PSB side, and, although partially fractioned, largely resemblant of the parent model Arch3, where the cavity negative charge is hosted exclusively by D222. On the other hand, the red-shifted model features a complete spread of the negative charge which has fractioned and relocated to all the cavity residues, with the barycenter is slightly closer to the β-ionone half of the chromophore. Associated to the differences in cavity electrostatics, the S_0 chromophore positive charge is more localized on the PSB side of in the blue-shifted model while it is largely more delocalized on the π-system in the red-shifted model.

Figure 4.6.1C and **Figure *4.6.1*A D** shows that the fragmentation is somehow limited for vertical excitation energies up to $\Delta E^{a}_{S1-S0}(\mathbf{Q_3})$ = 69.2 kcal mol^{-1}, where the dominant fraction of virtual MC charge is still located at D222, while it is global for all the other distributions ($\mathbf{Q_4}$ to $\mathbf{Q_8}$). As shown in **Figure 4.6.1D,** the spread of the negative charge can be conveniently quantified by taking the norm of the $\mathbf{Q}$ vectors representing the different virtual MC charge distribution. As already pointed out, the fragmentation (rather than relocation) of negative charge, occurring spontaneously from our optimization procedure, becomes more and more dominant as target $\Delta E^{a,*}_{S1-S0}$ becomes smaller, except for $\mathbf{Q_8}$, where we assist to a partial re-localization of the negative charge on residues located to the β-ionone side of the retinal.

4.7 A novel theoretical framework for color tuning

The congruous increase of spread or fractioning of negative charge following almost perfectly the trend in blue to red-shifted λ^{a}_{max} absorptions (i.e., from large to small ΔE^{a}_{S1-S0} values) is reminiscent of the mechanisms of wavelength regulation discussed by Borhan and coworkers[110] for mutants of human cellular retinol binding protein II (hCRBPII). It was suggested that to achieve a maximally red-shifted λ^{a}_{max} in this protein, the solution is three-fold: (i) removing the MC, (ii) creating an isolated environment encapsulating the retinal and (iii) project a uniformly neutral electrostatics potential (ESP) across the polyene chain to allow the maximal delocalization of retinal-PSB positive charge. These findings are in contrast with the previously reported external point charge model, which suggests that a red-shifted absorption may be achieved by introducing a negative charge in the vicinity of the retinal β-ionone side or a positive charge closer the PSB side[64].

Our simulations do not perfectly mimic the conditions of hCRBPII experiment; however, contrarily to hCRBPII, Arch3 is a membrane protein condition (ii) is automatically enforced. Furthermore, our results hint at similar conclusions as those suggested at points (i) and (iii). Looking at the progression from $\Delta E^{a}_{S1-S0}(\mathbf{Q_1})$ to

$\Delta E^{a}_{S1-S0}(\mathbf{Q_8})$ and the corresponding solutions $\mathbf{Q_1}$ to $\mathbf{Q_8}$, we assist, in general, to an increasing pattern of MC charge fractioning and homogeneous distribution and relocation to all the Arch3 cavity residues.

Although in our model we do not replace the MC with another non charged residue, the fraction of negative charge on D222 in models from $\mathbf{Q_4}$ to $\mathbf{Q_8}$ is so small **(Figure 4.6.1C)** that in practice, point (i) above is enforced. Furthermore, if we compare the ESP projected by the opsin on a retinal cross-section, the virtual charge distributions associated to $\mathbf{Q_1}$(a blue-shifted model, **Figure 4.7.1A**) and $\mathbf{Q_8}$ (a red-shifted model, **Figure 4.7.1B**), we observe that the latter features a much more uniform ESP across the retinal conjugated π-system. Although the ESP in model $\mathbf{Q_8}$ is not neutral as required by point iii above, but rather uniformly negative, we believe, as outlined in **Figure 4.7.1A** and **Figure 4.7.1B** already discussed, that the delocalization effect exerted on the rPSB positive charge is the same.

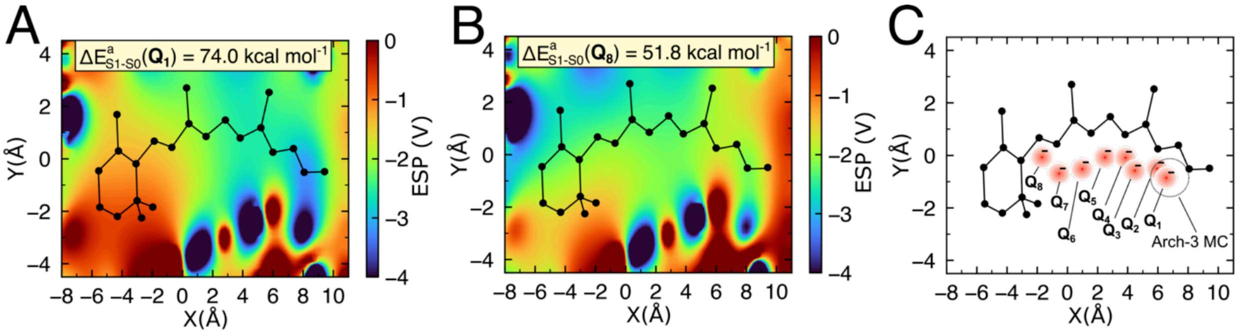


Figure 4.7.1 Electrostatic potential projected on a retinal cross section by the protein environment for the QM/MM models corresponding to $\Delta E^{a}_{S1\text{-}S0}(Q_1)$ = 74.0 kcal mol^{-1} (A) and $\Delta E^{a}_{S1\text{-}S0}(Q_8)$ = 51.8 kcal mol^{-1} (B). respectively. (C) centroid of virtual MC negative charge for the models Q_1 to Q_8.

Careful examination of the centroid of negative virtual MC charge for models $\mathbf{Q_1}$ to $\mathbf{Q_8}$ (**Figure 4.7.1C**), however, suggests that the mechanism of wavelength modulation in rhodopsins toward red-shifted value may be regulated by both the principles unveiled in hCRBPII and from the simple external point-charges scheme. We found indeed that progressing towards red-shifted λ^{a}_{max}, the centroid of the complex diffused patterns of negative charge moves progressively towards the β-ionone moiety, reminiscent of the external two point-charge model of Nakanishi and coworkers.

In **Figure 4.7.2A** and **Figure 4.7.2B** we show that the changes in ESP from $\mathbf{Q_1}$ to $\mathbf{Q_8}$ not only determine a difference in S_0 positive charge delocalization, but also a completely different change in electron density ($\delta\rho_{abs}$) associated to the $S_0 \rightarrow S_1$ transition at the Franck-Condon (FC) point. At $\Delta E^{a}_{S1-S0}(\mathbf{Q_1})$ = 74.0 kcal mol^{-1} , the transition involves the canonical redistribution of positive charge from the iminium position, where it is mostly localized in the S_0 state, to the β-ionone moiety after excitation In contrast, at $\Delta E^{a}_{S1-S0}(\mathbf{Q_8})$ = 51.8 kcal mol^{-1}, the positive charge of the chromophore is already highly delocalized in S_0 and the excitation to S_1 is associated to a $\delta\rho_{abs}$ which is not directed to either ends of the chromophore.

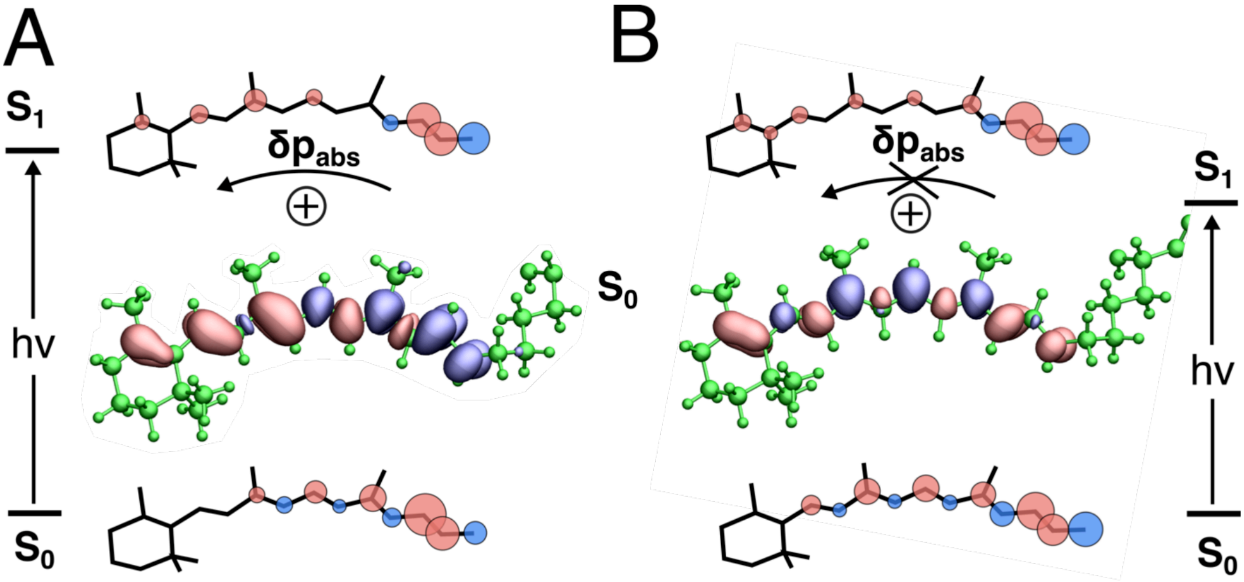


Figure 4.7.2 Electron density difference ($\delta\rho_{abs}$) associated to the vertical $S_0 \rightarrow$ S1 transition at the FC point for structure $\Delta E^{a}_{S1\text{-}S0}(Q_1)$ = 74.0 kcal mol^{-1} (A) and $\Delta E^{a}_{S1\text{-}S0}(Q_8)$ = 51.8 kcal mol^{-1}, (B). The blue and red areas correspond to the zones of electron density increases and decreases respectively. Isovalues set to 0.002 a.u. The retinal charges > 0.05 of the states involved in the transition are also shown as blue circles (negative charge) and red circles (positive charge).

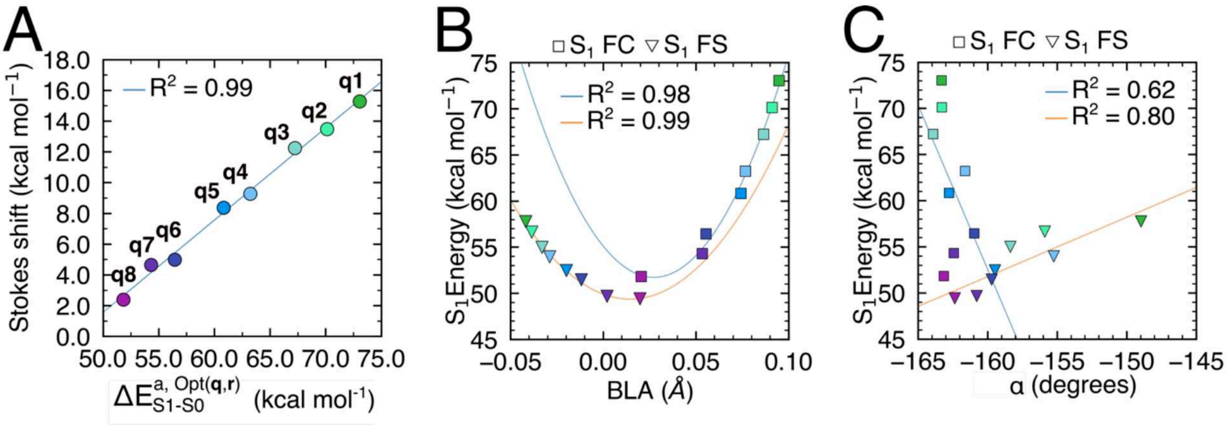


Figure 4.7.3. Properties of Q_1 to Q_8 models excited state. (A) Stokes Shift (B) BLA and (C) reactive torsion.

This is consistent with the analysis discussed in **Figure 4.7.3**, where we show the results of the S_1 geometry optimization starting from the S_0 optimized structures corresponding to $\mathbf{Q_1}$ to $\mathbf{Q_8}$. For all the structures a planar minimum, or fluorescent state (FS) on the excited state potential energy surfaces was identified. Shifting the vertical excitation energy from blue to red has the effect of progressively diminishing the magnitude of $\delta\rho_{abs}$ associated to the S_0-S_1 transition, which in turn results in a progressively smaller Stokes Shift (**Figure 4.7.3A**). This is perfectly reflected by the evolution of the bond length alternation (BLA) along the series. **Figure 4.7.3B** shows indeed that the differences between the BLA of the S_0 optimized structure and S_1 optimized structures progressively diminishes going from blue to red shifted models, in a perfectly regular fashion modellable by a parabola. Interestingly, the BLA difference between the FC and FS geometries of the most red-shifted model, was the closest one, and both the BLAs of the FC and the BLA of FS for this model are close to zero and lie on the minimum of the fitted curve. Such property appears to demonstrate that at $\Delta E^{a}_{S1-S0}(\mathbf{Q_8})$ = 51.8 kcal mol^{-1} represent the lowest vertical excitation energy possibly achievable by tuning of the electrostatics environment.

4.8 The relationship between red shift and excited state lifetime and observed FQY

In microbial rhodopsins, all-*trans* retinal is typically isomerized into the 13-cis form, even if the reverse reaction is possible in several rhodopsins. However, since Arch3 is phylogenetically closely related to bacteriorhodopsin (BR), a widely studied microbial rhodopsins showing 100% selective all-*trans* to 13-*cis* isomerization, we investigate the S_1 PES along the C_{13}=C_{14} torsional coordinate for our set of configurations, via a relaxed scan. We define the S_1 isomerization barrier ($\mathrm{E}_{\mathrm{S1}}^{\mathrm{f}}$), as the highest point in S_1 energy along the approximated minimum energy path (MEP). As shown in **Figure 4.8.1**, we observe a that a decrease in $\Delta\mathrm{E}_{\mathrm{S1-S0}}^{\mathrm{a}}$ is associated to a linearly increasing $\mathrm{E}_{\mathrm{S1}}^{\mathrm{f}}$ along the photoisomerization coordinate. This conclusion seems to agree with recent experimental findings, discussed in **Chapter 3**, where engineered mutants of Arch3 (such as the Archers, the Archons, Arch7, Arch5 and so on) with red-shifted $\lambda_{\mathrm{max}}^{\mathrm{a}}$ were found to display increased fluorescence quantum yield (FQY). Similarly, one could explain the extreme photo physics and unusually high FQY observed in NeoR, which is also the most red-shifted rhodopsin and in general transmembrane protein reported to date[70,91–93]. Interestingly, our electrostatics model suggests that to minimize $\Delta\mathrm{E}_{\mathrm{S1-S0}}^{\mathrm{a}}$ we select a configuration featuring a highly diffused main counterion negative charge, which may explain why microbial rhodopsins with extremely red-shifted absorption, generally display mutations at the level of the main counterion complex (see **Section 1.5.3**).

Furthermore, we observe that such increase in $\mathrm{E}_{\mathrm{S1}}^{\mathrm{f}}$ magnitude is accompanied by a similar increase in other two quantities shaping the S_1 PES; these are (i) the height, with respect to FS, of an S_1/S_0 conical intersection (CoIn) associated to the isomerization coordinate, or $\Delta E_{\mathrm{CoIn\text{-}FS}}$, and (ii) the energy difference between a twisted S_1 minima located in the vicinity of CoIn with diradicaloid character, a twisted intramolecular diradical intermediate (TIDIR), and FS, or $\Delta E_{\mathrm{TIDIR\text{-}FS}}$. Moving to red-shifted absorption, the distance in energy between CoIn and TIDIR increases progressively, suggesting a progressively more sloped topology of CoIn.

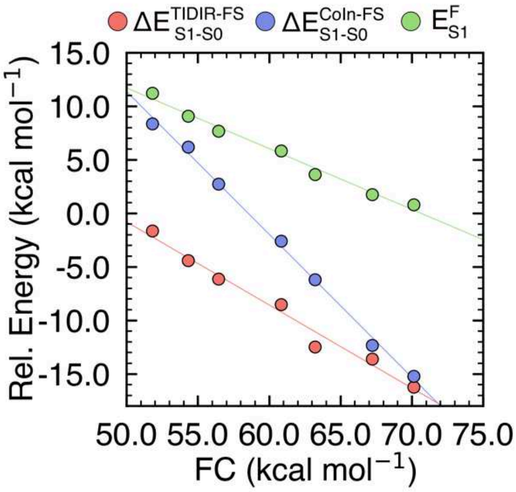


Figure 4.8.1 Linear relationship between S_1 FC energy and (i) isomerization energy ($\Delta E_{TIDIR-FS}$), (ii) CoIn energy with respect to the planar fluorescent minimum (FS) and (iii) isomerization barrier E^{f}_{S1} in red, blue and green dots, respectively.

4.9 Conclusions

In the past, the investigation of the key factors determining how to push the λ^{a}_{max} of different rhodopsins closer and closer to the near-infrared (NIR) range, has aimed at explaining the behavior observed in a number of specific cases and generated conclusions not extensible to other rhodopsins. The only attempt to provide a general theoretical framework for the "a priori" description of the color-tuning effects of the ESP is the external two-point charges model by Nakanishi and coworkers. Here, we developed a simple computational tool to evaluate the effect of counterion charge fractioning and diffusion on the chromophore vertical excitation energy. We found that the "effective delocalization" of the negative counterion charge is the main determinant of the vertical excitation energy in Arch3. Within the theoretical framework of our model, we observe that an extended delocalization of the counterion is necessary to achieve a low excitation energy (and thus red-shifted absorption). This conclusion provides a reasonable explanation for the outcome of several mutagenesis experiments performed on Arch3 and aimed at shifting its λ^{a}_{max} to the red. In fact, all the engineered Arch3 variants with a longer λ^{a}_{max} with respect to the progenitor invariantly feature one or multiple mutations to counterions located in the vicinity of RPSB. Since our delocalized models indicate how the electrostatics projected by the opsin should be in order to achieve pre-determined values of λ^{a}_{max}, in future studies we will use this knowledge to propose novel Arch3 mutants absorbing in the red or NIR. Furthermore, we have used the same artificial models to show the existence of a linear relationship between $\lambda_{\max}^{a}$ and chromophore isomerization barrier. This relationship agrees with experimental observation documenting that Arch3 variants with red-shifted $\lambda_{\max}^{a}$ consistently display an increased FQY with respect to the wild type.

Chapter 5

Conclusions and outlook

The present dissertation is focused on the investigation of the photochemistry and photophysics of the microbial rhodopsin Arch3 and a set of its variants. In particular, we focused on disentangling the mechanism regulating the fluorescence emission and the spectral shift of this class of rhodopsins.

Mechanism of Fluorescence Enhancement

In Chapter 3, we have investigated the excited state properties of seven proteins: the wild types Arch2 and Arch3, and the Arch3-based mutants QuasAr1, QuasAr2, Archon2, Arch7 and Arch5. These microbial proteins have large potential applications in Optogenetics, and in some cases they have already been used as genetically encodable voltage indicators (GEVIs), to study networks of interacting neurons *in vitro* and *in vivo*. The *in-silico* modelling of these systems represents an invaluable resource for the understanding of fluorescence enhancement mechanisms, which, in turn, is needed to help and drive the design of highly efficient GEVIs. For this reason, we focused our research on emphasizing the differences in the excited states properties of the with respect to the wild types Arch2 and Arch3. To study excited states, however, a multiconfigurational approach is required, and we used to this scope a set of techniques based on hybrid QM/MM methods.

The findings discussed in Chapter 3 can be divided in three parts.

- In the **first par**t we demonstrated the *existence of a proportionality between observed FQY values* and *computed* $\mathrm{E}^{\mathrm{f}}_{\mathrm{S1}}$ by constructing and validating a set of multi-configurational quantum chemical (MCQC) models corresponding to seven different Arch3 variants (the Arch set). The models were then employed to map the corresponding S_1 isomerization paths. The calculated S_1 potential energy profiles shows that all paths feature the same general topography, connecting the planar emitting minimum (FS) to a diradical twisted minimum (TIDIR) located in the vicinity of a conical intersection (CoIn), where decay to the ground state (S_0) occurs. The computed $\mathrm{E}^{\mathrm{f}}_{\mathrm{S1}}$ values are found to be in the order Arch2 = Arch3 < QuasAr1 < QuasAr2 < Arch5 < Archon2 = Arch7, which is the experimentally observed order of FQY.

- In the **second part** of the research we showed that *the FQY variation is modulated by a modification of the protein electrostatics*. We started by showing that the S_1 energy profiles follow the Hammond postulate which relates position and magnitude of $\mathrm{E}_{\mathrm{S1}}^{\mathrm{f}}$ with the energy difference between the FS "reactant" and the TIDIR "product" ($\Delta E_{FS\text{-}TIDIR}$). Indeed, the energy profiles are exergonic and thus a decrease in $\Delta E_{FS\text{-}TIDIR}$ values causes an increase in E^f_{S1} that, in turn, enhances FQY. As a result, a linear relationship between the easily computed $\Delta E_{FS\text{-}TIDIR}$ and the observed FQY values also exists. This indicates that higher fluorescent variants can be obtained by looking for mutants *simultaneously stabilizing and destabilizing the FS and TIDIR respectively*.

- In the **third part** we looked for the molecular-level factors modulating the critical $\Delta E_{FS\text{-}TIDIR}$ quantity. *We found that the electronic structure of the retinal chromophore at FS and TIDIR are very different and that this is reflected by their different charge densities*. This is important because a variation in the protein electrostatic potential (ESP_{opsin}) that simultaneously stabilize and destabilize the FS and TIDIR charge densities, respectively would decrease $\Delta E_{FS\text{-}TIDIR}$ and therefore increase $\mathrm{E}_{\mathrm{S1}}^{\mathrm{f}}$. This is particularly evident after comparing the two limit cases of Arch3 (FQY ~0.0001) and Arch7 (FQY ~0.012), where similar S_1 electronic charges distribution along the isomerization coordinate benefit from different stabilization by ESP_{opsin}.

In conclusion, the research discussed in **Chapter 3** of my dissertation unveils a previously unknown rational route to fluorescence enhancement in a set of Arch3 variants whose FQY values increase of two order of magnitude from 10^{-4} to 10^{-2}. The enhancement is due to specific ESP_{opsin} changes induced by the change in the protein amino acid sequence and acting all along the excited state isomerization path. It is important to emphasize that this link has been found through a study of the $\Delta E_{FS\text{-}TIDIR}$ variation indicating that and exquisite theoretical mechanistic entities such as conical intersections reveal their importance for the design and comprehension of optogenetics tools. Overall, our results show how our MCQC approach is providing an accurate frame for both spectral (trends in $\lambda^{\mathrm{a}}_{\mathrm{max}}$ and $\lambda^{\mathrm{f}}_{\mathrm{max}}$) and photochemical/photophysical properties (i.e., trends in FQY) in an extremely important sub-family of microbial rhodopsins. Our conclusions are solely based on electronic structure theory and basic biological-chemical arguments related to simple electrostatics and structural factor which we mapped to atomic differences in the gene sequences.

These findings open-up the possibility to envision a computational-driven, rational approach to the high-throughput design of neuronal reporters with enhanced fluorescence, based on the analytical tuning of factors which we have demonstrated to be fundamental to this aim.

To this aim, in my dissertation we identified a new, previously not reported exotic diradical intermediate (TIDIR), located in proximity of a decay funnel, controlling the isomerization energy and finally the FQY, which represents a perfect target for fluorescence engineering. In fact, the investigation of S_1 PES electronic structure allowed us to propose a general scheme for the ideal opsin electrostatic potential to be projected on the rPSBAT chromophore in order to optimize FQY. In fact, a promising application of the hereby proposed fluorescence design rule for the Arch family would be the large-scale *in-silico* rational screening of new Arch variants displaying enhanced FQY. The screening process could be designed to propose amino acid substitutions under the constraint that the newly generated structure $\mathrm{ESP}_{\mathrm{opsin}}$ is optimized to stabilize FS and destabilize TIDIR.

Mechanism of Spectral Tuning

While in **Chapter 3** I have mostly elucidated the mechanism of fluorescence enhancement in Arch proteins, Chapter 4 is almost exclusively dedicated to proposing an alternative to the external two-point charges model as theoretical framework to explain color tuning in rhodopsins. Since there is a high demand in neurobiology applications for rhodopsins exhibiting both bright fluorescence and red-shifted $\lambda^{\mathrm{a}}_{\mathrm{max}}$, I selected Arch3 as the reference study system also for this scenario. In fact, most of Arch3 variants featuring increased FQY are also absorbing in the red side of the visible spectrum.

The findings of this part of my dissertation can be summarized as follows:

- We proposed a new computational method based on a variational optimization protocol to study the effect of complex electrostatic interactions in terms of hypothetic delocalization of the chromophore main counterion on the geometry and vertical excitation energy of the chromophore. We demonstrate that by allowing the counterion negative charge to both fraction and diffuse to cavity amino acids, we can qualitatively mimic the diffusion effect typically exerted by interacting systems of dipoles on the counterion charge.
- Using this optimization protocol, we demonstrated that the counterion position and its *effective delocalization* are the most important electrostatic determinants to the chromophore vertical excitation energy in microbial rhodopsins.
- As previously documented by Borhan and coworkers on a completely different system, the human cellular retinol binding protein II (hCRBPII)[63], we conclude that also in the Arch family, the key to obtain a maximally red-shifted $\lambda^{a}_{\max}$ is projecting a uniformly negative ESP on the chromophore atoms. Such ESP would provoke a complete delocalization of the retinal iminium positive charge on the π-backbone. Consequently, the BLA and Stokes Shift of the chromophore vanish, as observed recently in NeoR[70,91–93].
- We observe a linear relationship between $\lambda^{a}_{\max}$ and the S_1 barrier along the all-*trans* to 13-*cis* isomerization path, justifying the observed properties of recently discovered Arch3 variants (such as those investigated in **Chapter 3** and **PAPER I**) or NeoR and its variants (see also **PAPER II**).

The development of this optimization tool and the rationalization discussed above, offer several exciting future perspectives. In fact, we envision the application of the optimization protocol as a tool to propose mutants with the desired $\lambda^{a}_{\max}$. Since we can now identify the general cavity electrostatics (ESP^{*}_{cavity}) which determines the optimal interaction with the retinal chromophore to produce a specific value of $\lambda^{a}_{\max}$, we are working on a high-throughput mutation strategy which is able to select those mutations which project on the chromophore an electrostatic potential (ESP_{cavity}) which minimizes the difference ESP_{cavity} - ESP^{*}_{cavity}. Furthermore, the optimization protocol architecture allows to easily modify the optimization target, such as λ^{f}_{max}, E^{f}_{S1}, S_0 isomerization barriers and so on, eventually imposing multiple constraints. To this aim, we could take advantage of our expertise on the FQY enhancement mechanism in the Arch family, to construct a new protocol which optimize the protein electrostatics to modulate the $\Delta E_{FS\text{-}TIDIR}$ quantity discussed above and extensively in **Chapter 3** and **PAPER I**.

Chapter 6

Articles

The larger part of my research work as a PhD student was dedicated to investigating the molecular determinants of fluorescence enhancement in Arch3 and its mutants, as summarized in **Chapter 3**. These efforts culminated in the publication of a scientific paper on Nature Communications (Barneschi, L. et al. *Nat. Comm*. 2022, 13, 1-9), listed below as **PAPER I**, in extended format. In 2020, Broser et. al., reported on the discovery of a new natural rhodopsin, Neorhodopsin (NeoR) expressed in *Rhizoclosmatium globosum,* a fungus from the phylum Chytridiomycota. The discovery of NeoR represents an unprecedented breakthrough in both the theoretical and experimental rhodopsin community. In fact, NeoR exhibits a FQY of ca. 0.2, more than two order of magnitudes larger than Arch7, which was previously thought to be the most fluorescent integral membrane protein ever reported. By relying on our findings on the mechanism of fluorescent enhancement in the Arch family, we constructed an MCQC model of NeoR and extensively investigated the property of its first electronically excited state (S_1). We found that NeoR fits almost perfectly in the general theory proposed to justify the increased FQY in the most fluorescent Arch3 variants. More specifically, we propose that NeoR represent an extreme case of such theory. In fact, the extreme photophysics of NeoR can be described in terms of the same $1A_g/1B_u$ diabatization scheme proposed for the Arch set. However, in NeoR the larger isomerization barrier is determined by an emphasized delocalization-confinement mechanism of the chromophore positive charge with respect to the Arch set. These findings, which are not discussed in the present dissertation, are available below (**PAPER II**).

In **Chapter 4**, we discussed a novel computational strategy to study the effect of the counterion charge effective delocalization (see **Section 4.5**) on the maximum absorption wavelength (λ^{a}_{max}) of the reference rhodopsin Arch3. A paper reporting on these results is currently in preparation. However, in a previous contribution focusing on a set of mutants of *Krokinobackter Eikastus* (KR2) a bacterial light-driven sodium pump, we have developed computation tools to study color tuning. In this case, since amino acid replacements were targeting a neutral amino acid position, we focused on energy decomposition schemes to disentangle steric and electrostatic contributions to the vertical excitation energy (ΔE^{a}_{S1-S0}) and consequently λ^{a}_{max}. Our conclusions, not discussed in this thesis, were reported in a scientific publication on Communications Biology (Nakajima et. al, *Comm. Bio.*, 2021, 4, 1-15), available below as **PAPER III**. The mechanistic findings and the code developed inspired the research on the novel technology discussed in **Chapter 4**.

PAPER I On the fluorescence enhancement of Arch neuronal optogenetic reporters

Publication: Barneschi, L., Marsili, E., Pedraza-González, L., Padula, D., De Vico, L., Kaliakin, D., Blanco-González, A., Ferré, N., Huix-Rotllant, M., Filatov, M. and Olivucci, M., 2022. On the fluorescence enhancement of arch neuronal optogenetic reporters. Nature communications, 13(1), pp.1-9.

Contribution: Conceptualization of the research, theoretical calculations, writing of the manuscript and revision, data analysis, graphics generation.

Current Status: Published, October 28, 2022.

nature communications

Article

https://doi.org/10.1038/s41467-022-33993-4

On the fluorescence enhancement of arch neuronal optogenetic reporters

Received: 27 May 2022

Accepted: 7 October 2022

Leonardo Barneschi[1], Emanuele Marsili[1,2,7], Laura Pedraza-González[1,8], Daniele Padula[1], Luca De Vico[1], Danil Kaliakin[3], Alejandro Blanco-González[3], Nicolas Ferré[4], Miquel Huix-Rotllant[4], Michael Filatov[5] & Massimo Olivucci[1,3,6]

The lack of a theory capable of connecting the amino acid sequence of a light-absorbing protein with its fluorescence brightness is hampering the development of tools for understanding neuronal communications. Here we demonstrate that a theory can be established by constructing quantum chemical models of a set of Archaerhodopsin reporters in their electronically excited state. We found that the experimentally observed increase in fluorescence quantum yield is proportional to the computed decrease in energy difference between the fluorescent state and a nearby photoisomerization channel leading to an exotic diradical of the protein chromophore. This finding will ultimately support the development of technologies for searching novel fluorescent rhodopsin variants and unveil electrostatic changes that make light emission brighter and brighter.

The imaging of neural activity requires bright fluorescent probes localized in the neuron membrane. Rhodopsins are membrane proteins that can be expressed in neurons and used for triggering, silencing and reporting on neuronal action potentials[1,2]. The prototypical fluorescent reporter is Archaerhodopsin-3 (Arch3), an archaeal rhodopsin from *Halorubrum sodomense*[3–5]. However, the fluorescence of Arch3 is extremely dim; its quantum yield (FQY) of ca. $1.1 \cdot 10^{-4}$ is almost four orders of magnitude lower than that of the green fluorescent protein (GFP)[6–11]. Furthermore, the fluorescence does not come from the dark-adapted state but rather from a photochemically produced photocycle intermediate that cannot deliver a prompt emission signal[11]. These facts not only make single neuron studies impossible with common microscopy techniques, but impair the investigation of neural activity at the population level as wide-field imaging techniques require improved spatial and temporal resolution[12]. In order to achieve better reporters, Arch3 has been engineered via directed evolutionary approaches and random mutagenesis, ultimately discovering variants such as the Archers[13], the QuasArs[12], the Archons[14], Arch5 and Arch7[13]. Recently, Hegemann and coworkers have investigated new Archon1 variants to elucidate the mechanism of fluorescence voltage sensitivity[15]. These variants feature a higher 10^{-3}–10^{-2} FQYs enabling, among other applications, the imaging of neuronal activity in living mammals and invertebrates[16–18]. It has also been suggested that such enhanced fluorescence originates, in contrast to Arch3, from a one-photon electronic excitation and, therefore, must come from the protein dark-adapted state[12,14,15]. Here we report on a (mechanistic) theory that explains how one-photon FQY is enhanced in Arch3 variants. We demonstrate that, to be predictive, such theory requires the mapping of an isomerization path producing a twisted intramolecular diradical intermediate (TIDIR) located in proximity of a conical intersection (CoIn). The energy difference between TIDIR and the emissive planar fluorescent state (FS), or $\Delta E_{\mathrm{TIDIR\text{-}FS}}$, controls the magnitude and positions of the S_1 photoisomerization barrier (E^{f}_{S1}) that, in turn, determines the FQY.

---

[1] Dipartimento di Biotecnologie, Chimica e Farmacia, Università di Siena, via A. Moro 2, I-53100 Siena, Italy. [2] University of Durham, Department of Chemistry, South Road, Durham DH1 3LE, United Kingdom. [3] Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403, USA. [4] Institut de Chimie Radicalaire (UMR-7273), Aix-Marseille Université, CNRS, 13397 Marseille, Cedex 20, France. [5] Department of Chemistry, Kyungpook National University, Daegu 702-701, South Korea. [6] University of Strasbourg Institute for Advanced Studies, 5, alleé duGeń eŕ al Rouvillois, F-67083 Strasbourg, France. [7] Present address: Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol, BS8 1TS, United Kingdom. [8] Present address: Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Giuseppe Moruzzi, 13, I-56124 Pisa, Italy. e-mail: olivucci@unisi.it

Based on the above result, we show that the increased FQY displayed by certain Arch3 variants is determined by inhibiting an electrostatic barrier suppression mechanism operating in the naturally occurring progenitor.

Results and discussion

Arch photoisomerization mechanism and S_1 PES topography

The function of rhodopsins is triggered by the photoisomerization of their retinal chromophores (Fig. 1A) that also operate as fluorophores[19]. Consistently with previous studies on fluorescent proteins[20], we hypothesize that the competition between isomerization and light emission determines the FQY value (see Supplementary Methods 3). This can be understood by examining a schematic representation of the chromophore excited state isomerization path (Fig. 1B) leading to a CoIn between the excited (S_1) and the ground (S_0) states where excited state decay occurs. From the scheme it is evident that the isomerization barrier (E^f_{S1}) located along the potential energy surface of the spectroscopically allowed S_1 state is a critical quantity. The higher E^f_{S1} is, the slower the isomerization is. Consequently, paths with high E^f_{S1} will feature an approximately planar fluorescent state FS with long S_1 lifetimes and, thus, high FQY. Below, we demonstrate a proportionality between *computed* E^f_{S1} and *observed* FQY by constructing multi-configurational quantum-chemical based models (MCQC) of the dark-adapted state of a set of Arch3 variants called the Arch-set (Supplementary Methods 1–3 and Supplementary Figs. 1–3). We will also show that the proportionality between *computed* E^f_{S1} and *observed* FQY remains valid when E^f_{S1} is replaced by the isomerization energy $\Delta E_{TIDIR\text{-}FS}$. Our MCQC models employ the well-established CASSCF[21] zeroth-order wavefunction defined by the selection of a (12,12) active space including all the π-electrons and orbitals of the retinal chromophore. Although the trends in spectral properties discussed throughout the text are well reproduced at this level, we discuss the results obtained after multi-state (XMS-CASPT2)[22,23] energy and geometrical corrections to the CASSCF geometries.

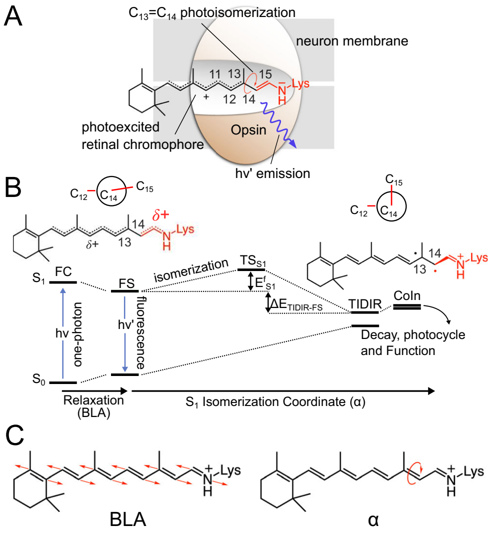


Fig. 1 | Photoisomerization mechanism of Archaerhodopsins. **A** Lewis formula representing the initial S_1 chromophore structure. **B** Representation of the chromophore isomerization path. FS corresponds to the fluorescent state. TIDIR represents the photoisomerization channel located near CoIn. FS and TIDIR are represented by Lewis formulas displaying distinct degrees of double bond twisting and charge transfer. **C** Main components of the reaction coordinate. BLA is numerically defined as the difference between the average single-bond length minus the average double-bond length along the C5 to N conjugated chain (for convenience, below we consider the BLA of the framed moiety exclusively). α is defined by the dihedral angle C12-C13-C14-C15.

Before focusing on the relationship between E^f_{S1} and FQY, it is useful to describe the changes of the all-*trans* retinal chromophore of Archaearhodopsins[24] along the isomerization path computed using the constructed Arch3 model. It is established that at least two geometrical coordinates are implicated in rhodopsin S_1 isomerization coordinates (Fig. 1C)[25–27]. The first describes the chromophore initial relaxation from the Franck-Condon (FC) point and corresponds to a bond length alternation (BLA) stretch. The second is the twisting (α) of the reactive *trans* $C_{13} = C_{14}$ bond ultimately leading to the 13-*cis* configuration. As shown in Fig. 2A and consistently with the scheme of Fig. 1B, such relaxation leads to the potentially emissive FS intermediate. This process is followed by progression along a flat potential energy region characterized by a monotonic decrease in α and connecting FS to a TIDIR intermediate located close to a sloped[28] CoIn. As it will be explained below, this previously unreported intermediate differs, in terms of electronic structure and topography, from the locally excited (LE) intermediate identified in a ring-locked derivative of bovine rhodopsin by Laricheva et al.[29].

Electronic character of the S_1 PES

In Fig. 2B we show that such a progression is replicated by the MCQC model of the top fluorescent variant Arch7. However, the comparison between the Arch3 and Arch7 S_1 energy profiles shows a significantly increased E^f_{S1} in the latter. Such an increase is seen in both the zeroth-order wavefunction calculation and in the more quantitative profiles obtained after applying geometrical and multi-state second-order perturbative correction (Supplementary Methods 4 and 5, Supplementary Figs. 4–7). Due to the high computational cost of QM/MM analytical Hessians, the TSs discussed throughout the text are approximated by the energy maxima along the relaxed scan connecting FS and TIDIR. These TSs must be considered approximate as it has not been possible to carry out a geometry optimization starting from a computed Hessian matrix as well as to compute a Hessian matrix at the end of the TS search. Furthermore, despite the increased FQY of the investigated Arch variants, recent measurements of the excited state lifetime (ESL) of some of the variants were found to be in the time range of picoseconds[30,31]. For this reason, we don't account in our calculation for T_1/S_1 intersystem crossing (ISC) as a viable competitive process to S_1 emission also considering that T_1 is a π-π* state with orbitals parallel (non-orthogonal) to those characterizing the S_1 state. Therefore, the singlet to triplet transition would be "forbidden" by the El-Sayed rule.

The progression along α is documented in Fig. 2C, D where we report the values of the relevant geometrical parameters of FS and TIDIR as well as the charge residing on the Schiff base moiety C_{14}-C_{15}-N. In fact, the geometrical changes are accompanied by variations of the chromophore electronic character. These can be conveniently followed by computing the fraction of positive charge and free valence (NUE)[32] residing on C_{14}-C_{15}-N (Supplementary Methods 6 and Supplementary Figs. 4, 7, 9 and 10). Following a recent report on the variants of DRONPA2[20], a soluble GFP-like protein, we interpret the results in Fig. 2A, B in terms of mixing of two diabatic states describing the covalent ($1A_g$) and charge-transfer ($1B_u$) characters of polyenes[33]. The computed C_{14}-C_{15}-N charge and NUE progression point to an S_1 state initially dominated by the $1B_u$ charge transfer character as already reported for other rhodopsins. However, such character decreases when relaxing to an FS that features a larger $1A_g$ weight, then constantly increasing all along the isomerization path. In summary, the electronic character evolves from mixed charge-transfer/covalent $1A_g/1B_u$ characters at FC and FS, to a substantially pure $1A_g$ character at

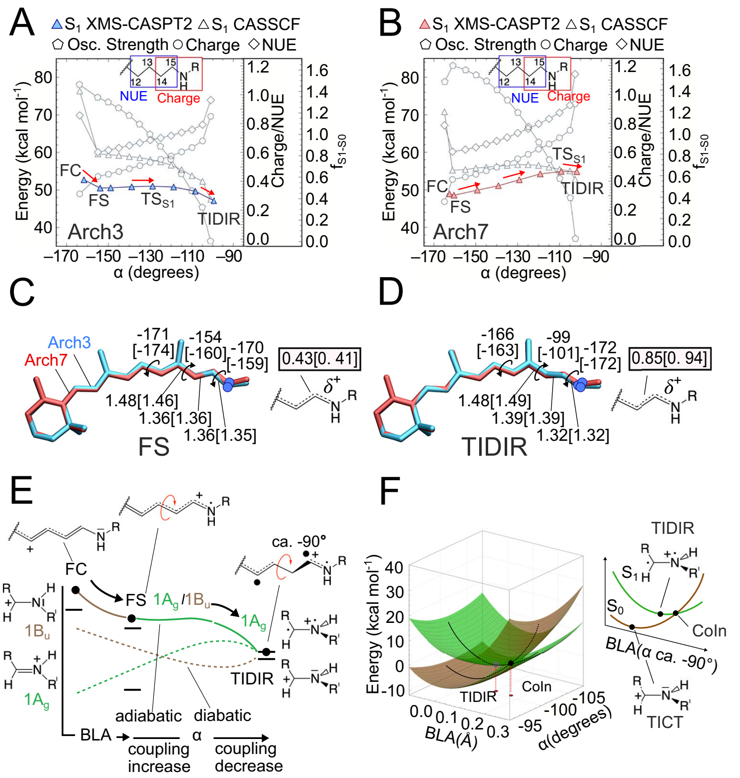


Fig. 2 | Comparison between computed Arch3 and Arch7 S_1 isomerization paths. **A** Variations in potential energy (relative to the S_0 equilibrium structure), charge distribution, free valence and oscillator strength along the Arch3 path. The energy profile in color is given after MS correction. Red arrows indicate progression along the isomerization coordinate. **B** Same data for Arch7. **C** Main geometrical chromophore parameters for the FS fluorescent intermediates of Arch3 and Arch7 (values in square brackets). The S_1 positive charge fraction on the C-C-N moiety are also given. **D** Same data for the photoisomerization channel TIDIR. **E** Schematic "decomposition" of the Arch3 adiabatic energy profile of panel **B** in terms of diabatic states associated to Lewis formulas of the $CH_2 = NH_2(+)$ minimal model. The progression of the electronic and geometrical structure of the relevant moiety of the full chromophore is given at the top. **F** Schematic representation of the CoIn region of Arch3 including the twisted diradical TIDIR along the relevant components of the reaction coordinate. The same coordinate also spans the CoIn branching plane. The potential energy scale is relative to the CoIn.

TIDIR that features an α value of ca. −90°, two spin-paired, but non-interacting, radical centers located on two orthogonal π-systems and a positive charge fully confined on C_{14}-C_{15}-N. This process is accompanied by a change in the S_1-S_0 oscillator strength ($f_{S1\text{-}S0}$) along α, from values typical of allowed electronic transitions to a forbidden transition at TIDIR.

The evolution going from a mixed $1A_g/1B_u$ character at FS to a $1A_g$ covalent/diradical character at TIDIR, provides information on the origin of the E^f_{S1} barrier. As illustrated in Fig. 2E the electronic coupling between the $1A_g$ and $1B_u$ diabatic states (represented, for simplicity, by the Lewis formula of a methylimine cation model) would initially increase due to the decrease in their energy gap. However, when α approaches orthogonality, the coupling decreases rapidly and becomes negligible at TIDIR. The negligible coupling at TIDIR is related to the vicinity of a CoIn (see Fig. 2F) where the $1A_g/1B_u$ coupling is zero. A CoIn deformation along the negative BLA direction lifts the degeneracy and intercepts TIDIR (see Supplementary Methods 7 and Supplementary Figs. 11 and 12). Notice that the same deformation along the S_0 potential energy surface achieves the twisted intramolecular charge transfer TICT structure[34] corresponding to the transition state driving the chromophore thermal isomerization also documented for bovine rhodopsin[33]. In conclusion, we associate the variation in the magnitude and position of E^f_{S1} (i.e., of TS_{S1} in Fig. 1B) with the variation along α of the $1A_g$ and $1B_u$ diabatic energies and their $1A_g/1B_u$ coupling.

S_1 isomerization barrier determines an increased FQY in the Arch variants

We now show that the isomerization mechanism above provides the basis for understanding the FQY variations along the Arch-set. To do so, we first demonstrate the existence of a correlation between computed E^f_{S1} and observed FQY values and then rationalize it based on FS and TIDIR charge distributions (or $1A_g$ and $1B_u$ character) calculated by constructing the MCQM models of all variants and using them to map the progression along α. The model accuracy is documented in Fig. 3A where we compare, after multi-state perturbative correction, the computed and observed trends of absorption ($\Delta E^a_{S1\text{-}S0}$, top) and emission ($\Delta E^f_{S1\text{-}S0}$, bottom) vertical excitation

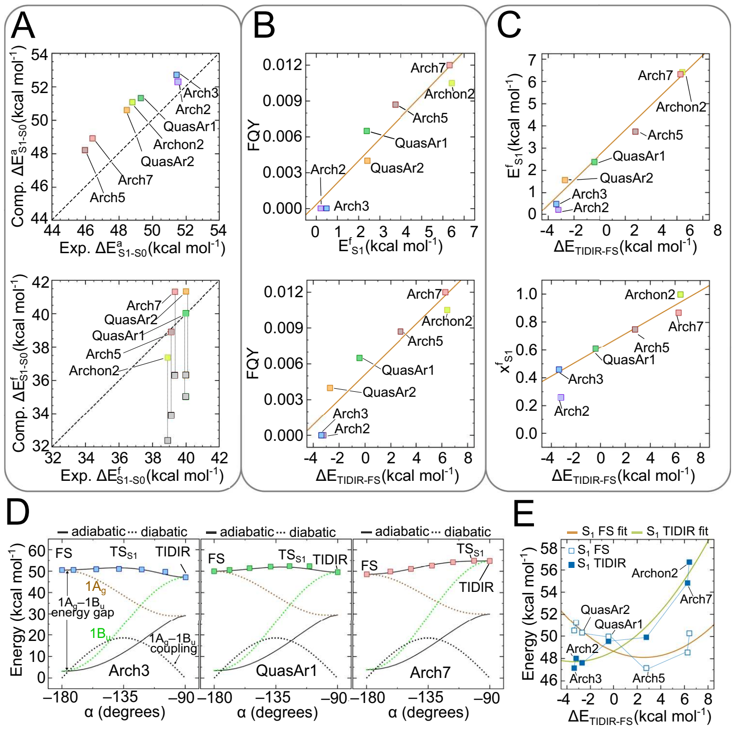


Fig. 3 | Relationship between computed S_1 isomerization properties and observed FQY along the Arch-set. **A** Comparison between computed and observed excitation energies associated with light absorption at the ground state equilibrium structure (top) and light emission from the fluorescent state (bottom). Vertical excitation energy associated to emission features a correction of 4.5 kcal mol^{-1} to account for kinetic energy derived from ref. 24. Gray and colored squares represent the emissions before and after correction, respectively. **B** Relationship between S_1 isomerization barrier and FQY (top) and between S_1 isomerization energy and FQY (bottom). Linear fits are given as orange lines. **C** Relationship between S_1 isomerization barrier and isomerization energy (top) and between S_1 isomerization transition state position and isomerization energy for three Arch3 variants. Linear fits are given as orange lines. **D** One-dimensional model for the relationship between barrier and isomerization energy. **E** Non-linear relationship between the FS and TIDIR S_1 excitation energies with respect to S_0 FC and the isomerization energy. Source data are provided in the Source Data file (d).

energies. The analysis of the results demonstrates that, in all variants, FS and TIDIR feature the same distinct charge distributions (Supplementary Figs. 4 and 7).

The bottom panel of Fig. 3B shows that computed E^{f}_{S1} and observed FQY variations are directly proportional. If one assumes that the Hammond-Leffler postulate[35,36] is valid for an S_1 isomerization, E^{f}_{S1} and $\Delta E_{TIDIR\text{-}FS}$ (i.e., the reaction endothermicity or, simply, isomerization energy in Fig. 1B) must also correlate. This leads to the conjecture that $\Delta E_{TIDIR\text{-}FS}$ is proportional to FQY. The bottom panel of Fig. 3B demonstrates that such proportionality exists. In other words, a progressive stabilization of the "reactant" FS and/or destabilization of the "product" TIDIR, must lead to higher FQYs. The relationship between the computed E^{f}_{S1} and $\Delta E_{TIDIR\text{-}FS}$ for the entire Arch-set can be modeled by using a basic two-state one-mode Hamiltonian that fits the computed S_1 and S_0 energy profiles along mode α in terms of $1A_g$ and $1B_u$ diabatic energies and a harmonic $1A_g/1B_u$ coupling function (Supplementary Methods 8 and Supplementary Fig. 12). The maximum along the resulting S_1 energy profile represents TS_{S1} and, as demonstrated in Fig. 3D for Arch3, QuasAr1 and Arch7, replicates the computed MCQC S_1 energy profiles. As shown in Fig. 3C the model Hamiltonian supports the existence of a proportionality between E^{f}_{S1} (top panel) and the TS_{S1} position X^{f}_{S1} (bottom panel) and $\Delta E_{TIDIR\text{-}FS}$. We stress that this mechanistic model assumes no "a priori" relationship between diabatics energy difference and diabatic coupling, but is a simple valence-bond type description of our adiabatic S_1 PES, assuming two pure resonance formulas ($1A_g$ and $1B_u$), whose weights are associated to the documented variation in positive charge distribution along the reaction path.

The $\Delta E_{TIDIR\text{-}FS}$ increases along the Arch-set may originate from either FS stabilization or a TIDIR destabilization effects (or from both effects). In Fig. 3E we show that the first effect (i.e., substantially the 0-0 excitation energy) shows a decrease until Arch5 that has the lowest excitation energy and then increase up to Archon2. In contrast, the TIDIR destabilization with respect to the same reference (i.e., the destabilization of the photoisomerization channel) increases almost monotonically indicating a higher sensitivity of FQY to mutations changing $\Delta E_{TIDIR\text{-}FS}$ rather than the difference in energy between FC and FS. The described $\Delta E_{TIDIR\text{-}FS}$ changes are confirmed by probing the light-induced dynamics of the Arch-set. To do so we employed the variants MCQC models to propagate a quantum-classical trajectory released from the FC point with zero initial nuclear velocities (Supplementary Methods 9 and Supplementary Figs. 14 and 15). As discussed in the literature, these trajectories mimic the evolution of the center of the S_1 population and are useful to detect barrierless or

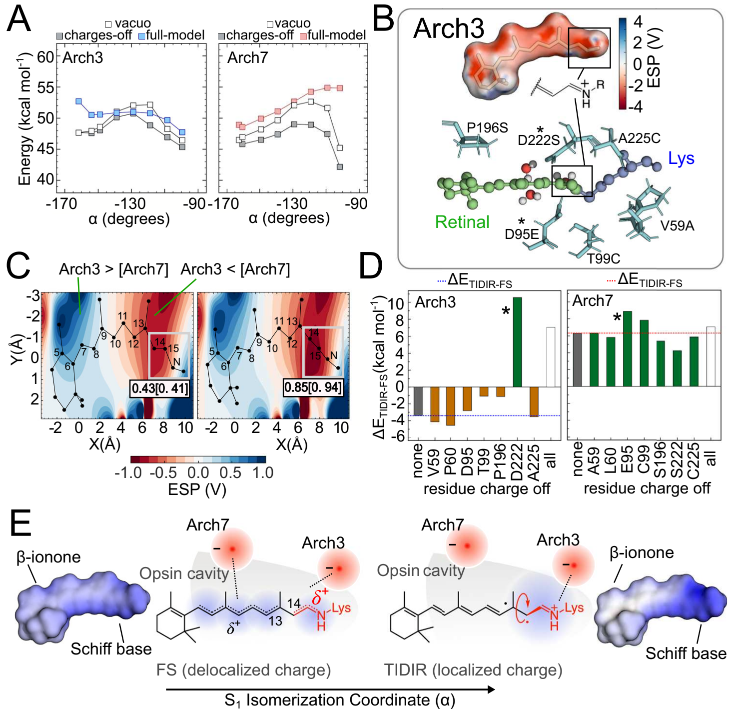


Fig. 4 | Origin of the S_1 isomerization barrier in Arch3 and Arch7. **A** Effect of the opsin charges on the isomerization energy profile for Arch3 (left) and Arch 7 (right). **B** Top. Arch3 opsin electrostatic potential projected on the chromophore solvent accessible surface. Bottom. The cavity amino acids distinguishing Arch3 from Arch7. The * symbol indicates the Arch3 (D222) and Arch7 (E95) chromophore counterions. **C** Two-dimensional plots showing the Arch3-Arch7 electrostatic potential difference at FS (left) and TIDIR (right). The total charge of the framed Schiff-base moiety is also given. **D** Variation in the Arch3 (left) and Arch7 (right) isomerization energy value after zeroing the residue charge of the cavity residues in panel **B** bottom. Reference values are shown as gray bars. Negative and positive isomerization energies are shown in orange and green, respectively. Finally, the effect of zeroing the charges of all residues simultaneously is also given as an open bar. **E** Point charge model for the delocalization-confinement mechanism. The Arch3-Arch7 ESP difference of **C** can be modeled in terms of a re-locating effective negative charge from the Schiff base region to a cavity region displaced towards the β-ionone ring of the chromophore. The increase in distance between the effective charge and the confined positive charge of TIDIR explains its electrostatic destabilization in Arch7 with respect to Arch3. The computed chromophore electrostatic potential projected on the chromophore solvent accessible surface is given at the left and right of the figure to support the schematic representation of a delocalized and confined charge given at the center. Source data are provided in the Source Data file (a, c, d).

nearly barrierless isomerization paths and validate the accuracy of computed torsional scans[27,37].

Consistently with the trend in E^f_{S1} values, the Arch3 trajectory reaches the CoIn region and decays to S_0 on a sub-500 fs timescale via a $C_{13}=C_{14}$ isomerization pointing to species that do not display significant fluorescence. In contrast, the QuasAr1 and Arch7 trajectories representing the remaining variants, orbit in the FS region for the entire 450 fs simulation time.

Molecular determinants of the isomerization barrier

Above we have demonstrated that our MCQM models produce a $\Delta E_{TIDIR\text{-}FS}$ trend proportional to the observed FQYs for the entire Arch-set. We now focus on the limiting cases of Arch3 and Arch7 to show that the proportionality between $\Delta E_{TIDIR\text{-}FS}$ and FQY is linked to the variations in the protein (opsin) sequence. This is possible because, as seen in Fig. 2C, D, the FS and TIDIR retinal chromophores have distinct charge distributions and, therefore, $\Delta E_{TIDIR\text{-}FS}$ must be sensitive to the protein electrostatics. In other words, an opsin electrostatic potential (ESP_{opsin}) stabilizing FS or/and destabilizing TIDIR would produce a larger $\Delta E_{TIDIR\text{-}FS}$ value and, in turn, enhance FQY. Such effect has been investigated by recomputing the S_1 energy profiles along α after setting to zero the opsin charges of the MCQC models (Supplementary Methods 10 and Supplementary Figs. 15–17). It is apparent from inspection of the energy profiles in Fig. 4A, B that, in the absence of the protein electrostatics, both Arch3 and Arch7 display a sizable energy barrier along α. However, the left panel of Fig. 4A demonstrates (compare the full model with the charges-off energy profiles) that, in Arch3, the opsin charges stabilize TIDIR with respect to FS yielding a low $\Delta E_{TIDIR\text{-}FS}$ value and, consequently, a negligible E^f_{S1}. In contrast, the right panel of Fig. 4A shows that, in Arch7, the opsin charges have the opposite effect. We can conclude that in the absence of the protein electrostatics, the geometrical deformation imposed by the opsin cavity on the chromophore backbone leads to a sizable E^f_{S1} value. Notice that the geometrical deformation is the result of both cavity steric and electrostatic effects on the chromophore isomerization coordinate and that these effects are common to Arch3 and Arch7, as well as to all members of the Arch-set (Supplementary Methods 10). This has been confirmed by recomputing the same energy profile in

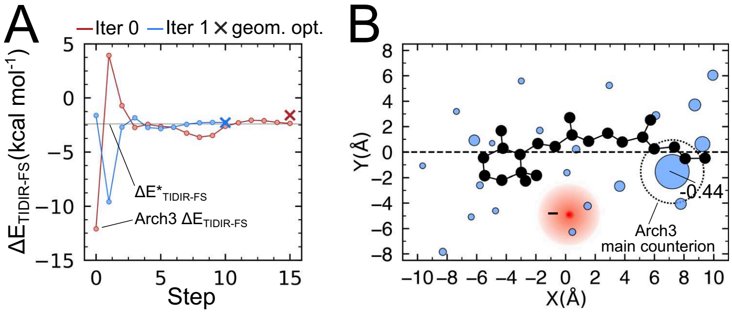


Fig. 5 | Predicted change of counterion charge distribution of Arch3 QM/MM model to yield the Arch7 $\Delta E_{TIDIR\text{-}FS}$ value. **A** $\Delta E_{TIDIR\text{-}FS}$ (SA2-CASSCF/AMBER level) change along the optimization steps leading from the Arch3 to the Arch7 value. The optimization is driven by the target $\Delta E_{TIDIR\text{-}FS}$ values ($\Delta E^*_{TIDIR\text{-}FS}$, indicated by a thin horizontal line) and corresponds to the minimization of the square of the scalar function $\Delta E_{TIDIR\text{-}FS}(q) - \Delta E^*_{TIDIR\text{-}FS}$ (i.e., $\Delta\Delta E_{TIDIR\text{-}FS}(q)$) as a function of the cavity residue charge vector q. Each iteration of the algorithm (two full iterations are reported in the panel) is divided in two parts; (i) first q is optimized at fixed FS and TIDIR geometries via conjugated-gradient optimization such that $\Delta E_{TIDIR\text{-}FS}(q) = \Delta E^*_{TIDIR\text{-}FS}$ and then (ii) the FS and TIDIR QM/MM model geometries are relaxed, and $\Delta E_{TIDIR\text{-}FS}$ is recomputed. If the difference between the $\Delta E_{TIDIR\text{-}FS}$ calculated after part ii and $\Delta E^*_{TIDIR\text{-}FS}$ is above a selected threshold, i-ii are repeated. The circles indicate the steps of part i, and the crosses indicate the results of geometrical relaxation of part ii. **B** Final charge distribution obtained after convergence of the optimization above. Two-dimensional representation of the fraction of negative charges residing in the cavity residues is proportional to the radius of the blue circles (the main counterion D222 final charge is indicated). The barycenter of the corresponding negative charge is shown as a red circle. The original localized Arch3 negative charge is indicated by the large open circle centered at the original counterion position Source data are provided in the Source Data file (a, b).

vacuo (i.e., in absence of the VdW interaction between protein cavity and chromophore) that still show a barrier in all cases.

The described electrostatic stabilization of TIDIR relative to FS in Arch3, and the consequent disappearance of the barrier, has been rationalized by mapping the ESP_{opsin} on the chromophore solvent accessible surface (Fig. 4B, top). The map shows a prevalent negative ESP_{opsin} value in the area surrounding the Schiff base moiety whose effect along the isomerization coordinate is described by plotting the difference (ΔESP_{opsin}) between the Arch3 and Arch7 electrostatics. In fact, the ΔESP_{opsin} mapped along a cross-section roughly parallel to the FS (Fig. 4C, left) and TIDIR (Fig. 4C, right) chromophore backbones, demonstrates that the opsin of Arch3 preferentially stabilizes the positive charge fully confined on the Schiff base moiety of TIDIR while the opsin of Arch7 better stabilizes the delocalized charge spread towards the β-ionone of FS. These effects can be also discussed using the model Hamiltonian presented above where the $1B_u$ diabatic dominates the FS while the $1A_g$ diabatic dominates the TIDIR. Since a Schiff base confined charge is a characteristic of $1A_g$, the Arch3 electrostatics must stabilize the covalent/diradical $1A_g$ character at TIDIR while the Arch7 electrostatics destabilizes it consistently with the diabatic model of Fig. 3D. It is possible to demonstrate that this effect remains valid when one uses a more realistic two-state two-mode Hamiltonian model generating two-dimensional adiabatic energy surfaces function along both BLA and α (Supplementary Methods 8).

Role of the main counterion in the barrier generation or suppression

At this point we propose an atomistic mechanism for the Arch3 barrier suppression. The differences in the ESP_{opsin} of Arch3 and Arch7 is a product of the protein sequence variation and, more specifically, from the seven residue replacements displayed at the bottom of Fig. 4B. In our MCQC models, the ESP_{opsin} is produced by the point charges centered on the atoms of each protein residue (Supplementary Methods 2). Therefore, the ΔESP_{opsin} of Fig. 4C must reflect the difference in point charges before and after the residue replacements. The effect of such difference has been investigated by recomputing $\Delta E_{TIDIR\text{-}FS}$ after setting the point charges of residues 59, 60, 95, 99, 196, 222, 225 as a group or individually. The results in Fig. 4D show that when setting to zero the charges of all residues, the $\Delta E_{TIDIR\text{-}FS}$ of Arch3 and Arch7 become similar. More specifically and consistently with the results in Fig. 4A, B, the Arch3 point charges must cause a ca. 9 kcal mol^{-1} decrease in $\Delta E_{TIDIR\text{-}FS}$ while the corresponding Arch7 residues do not seem to affect the original ca. 6 kcal mol^{-1} $\Delta E_{TIDIR\text{-}FS}$ value. From the analysis of the individual residues, it is apparent that the counterion at position 222 dominates the Arch3 electrostatics. Indeed, during the Arch3 to Arch7 transition the counterion is moved from position 222 to position 95 while the position 222 is taken by an uncharged cysteine. In conclusion, the counterion relocation appears to be the main mechanism for the $\Delta E_{TIDIR\text{-}FS}$ modulation. This is not a general mechanism. In fact, other members of the Arch-set all have the counterion in position 222 (Supplementary Methods 1) but display a regularly increasing $\Delta E_{TIDIR\text{-}FS}$ value. In these cases, the effect of the residue replacements must cause a ESP_{opsin} change that partially screen the counterion effect found in Arch3, which may be interpreted in terms of a "virtual relocation" of the counterion as explained below.

We propose that this relocation and "diffusion" of the negatively charged counterion (a virtual counterion), coupled with a delocalization-then-confinement mechanism of the positive charge of the chromophore, explains the regular change in isomerization barriers. In brighter Arch-3 variants (Arch-5 and Arch-7) the virtual counterion is increasingly distant and more diffuse from the Schiff base moiety. In these variants, at α = 0° the chromophore positive charge is delocalized, and its centroid is close to the counterion, leading to a stabilization of FS. As soon as α progresses and approaches a 90° twist, the confinement of the chromophore charge on the Schiff base moiety gradually increases, causing its centroid to drift away from the virtual counterion, inevitably determining a de-stabilization of TIDIR.

To support this conclusion, we developed a basic model that allows to compute (i.e., optimize) the protein charge distribution inducing a specific $\Delta E_{TIDIR\text{-}FS}$ value. This is done by allowing the relocation and fragmentation (i.e., diffusion) of the negative charge of the main counterion to other cavity residue positions, to mimic what observed in Fig. 4E. This can be achieved by defining a scalar function of the cavity residue charge vector (q) which returns $\Delta E_{TIDIR\text{-}FS}$. Such $\Delta E_{TIDIR\text{-}FS}(q)$ function is differentiated numerically to study how $\Delta E_{TIDIR\text{-}FS}$ responds to q. As detailed in Supplementary Methods 11, the problem of determining q can be formulated as a constrained optimization. In Fig. 5, we show how the optimization modifies the charge distribution of Arch3 to reproduce the $\Delta E_{TIDIR\text{-}FS}$ value computed for Arch7 at the zeroth-order level. The resulting q shows that, for Arch3 to reproduce Arch7 excited state properties, a relocation of ca. 50% of D222 negative charge to other cavity residues is necessary, thus supporting the conclusion that a relocation and diffusion of the counterion is indeed a determinant of the TIDIR destabilization in the Arch-set.

Conclusions

The engineering of efficient, rhodopsin-based action-potential reporters (also called Genetically Encodable Voltage Indicators or GEVIs in the literature) is a challenge currently addressed by experimentally investigating the fluorescence and voltage sensitivity mechanisms[15] and by tuning the protein fluorescence via several rounds of random mutagenesis and/or directed evolution. Above we have investigated the increase in fluorescence intensity in a set of Arch3 variants using MCQC models with second order perturbative corrections. The key results are: (i) the FQY is primarily determined by the competition between S_1 emission and isomerization extending the mechanism reported for DRONPA2[20] to retinal proteins, (ii) the isomerization rate is governed by the stability of the rather exotic TIDIR intermediate with respect to the fluorescent state, (iii) the TIDIR stability is modulated by the electrostatics of the protein that increasingly offset the S_1 isomerization energy barrier when going from Arch7 to Arch3.

The atomistic mechanism for the TIDIR stabilization appears to be a consequence of the cavity electrostatics on the distinct charge distributions of the FS and TIDIR chromophores. As schematically (top schemes) and computationally (bottom plots) displayed in Fig. 4E, the chromophore positive charge is largely delocalized in the FS but confined on the C_{14}-C_{15}-N fragment in the TIDIR. In this situation the described change in electrostatics from Arch3 to Arch7 leads to a reduced stabilization of the Schiff base confined with respect to the delocalized charge. In simple terms, the effect of the amino acid (and atomic point charges) replacement resulting in the Arch3 to Arch7 progressive TIDIR destabilizing along the Arch-set, can be interpreted as the gradual relocation of a negative charge from the Schiff base region to a region closer to the β-ionone ring.

The presented mechanism has both methodological and biological implications. The first is related to the fact that TIDIR and FS, being energy minima on the S_1 potential energy surface, are computationally fast to locate. It is thus possible to envision the development of a tool for the in silico selection of highly fluorescent Arch3 variants. Such tool would be based on the automated[38] construction of MCQC models and the calculation of the corresponding $\Delta E_{TIDIR\text{-}FS}$ value to be maximized via in silico mutational experiments. The biological implication is instead related to the hypothesis that the changes in sequence, and thus electrostatics, leading to a negligible Arch3 fluorescence could have occurred also in nature through a natural selection process aimed at increasing the protein photoisomerization rate and improve the protein function. We cannot exclude that the same mechanism plays a role in other microbial rhodopsin evolution. On the other hand, an inverse natural selection process leading to the suppression of the photoisomerization, and maximization of the fluorescence output may have generated the recently discovered Neorhodopsin from *Rhizoclosmatium globosum*[39] that displays an intense fluorescence with a ca. 0.2 FQY and therefore much closer to that of DRONPA2 and other optimized green fluorescent protein variants.

Methods

The three-dimensional structures of the Arch-set proteins were obtained either from the corresponding crystallographic structures deposited in the Protein Data Bank (Arch2 and Arch3 with PDB ID:3WQJ and 6GUX, respectively) or via the comparative model (the Arch3 variants) approach described in Supplementary Methods 1. The corresponding QM/MM models were constructed from the obtained three-dimensional structures according to the *a*-ARM based protocol[38] discussed in Supplementary Methods 2 with the QM and MM subsystems interacting via an electrostatic embedding scheme. The retinal chromophore atoms, the chromophore-bound lysine side chain atoms starting from the N-terminal to the Cε and the hydrogen link atom (HLA) are included in the QM subsystem and treated at the multiconfigurational (CASSCF) level, while the rest of the atoms of the model are treated at the MM level using the AMBER94 force field[40].

The ground state optimized geometries are obtained via geometry optimization at the single-state CASSCF(12,12)/6-31G*/AMBER level of theory. Further refinement was performed introducing a perturbative geometrical correction to produce XMS-CASPT2/SA3-CASSCF(12,12)/ANO-L-vDZP/AMBER geometries (Supplementary Methods 4 and 5). Both the CASSCF and XMS-CASPT2 models were then used to compute reaction paths along the isomerization coordinate via a relaxed scan taking, as the reacting coordinate, the twisting of the reactive double bond. The CASSCF geometries were also employed to probe the excited state dynamics of three models via non adiabatic deterministic surface-hop trajectories propagated from the Franck-Condon (FC) point on S_1 with zero initial velocities (FC trajectories) at the SA2-CASSCF(12,12)/6-31G*/AMBER level using the Tully algorithm[41]. All the QM/MM calculations were performed using the [Open]Molcas v19.11/TINKER and MOLCAS v8.1/TINKER packages[42–44].

Data availability

Cartesian coordinates of the FC, FS, CoIn and TIDIR geometries of the QM/MM models calculated at the SA2-CASSCF(12,12)/6-31 G*/AMBER level of theory for the models of the Arch set are provided as Supplementary Data 1.

Cartesian coordinates of the FC, FS and TIDIR geometries of the QM/MM models calculated at the XMS-CASPT2/SA3-CASSCF(12,12)/ANO-L-vDZP/AMBER level of theory for the models of the Arch set are provided as Supplementary Data 2. Source data is provided with this paper. Source data are provided with this paper.

Code availability

All the calculations discussed in this work are based on the MOLCAS/TINKER interface.

References

1. Xu, Y., Zou, P. & Cohen, A. E. Voltage imaging with genetically encoded indicators. *Curr. Opin. Chem. Biol.* **39**, 1–10 (2017).
2. Li, X. et al. Fast noninvasive activation and inhibition of neural and network activity by vertebrate rhodopsin and green algae channelrhodopsin. *Proc. Natl Acad. Sci.* **102**, 17816–17821 (2005).
3. Chow, B. Y. et al. High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* **463**, 98–102 (2010).
4. Saint Clair, E. C., Ogren, J. I., Mamaev, S., Kralj, J. M. & Rothschild, K. J. Conformational changes in the archaerhodopsin-3 proton pump: detection of conserved strongly hydrogen bonded water networks. *J. Biol. Phys.* **38**, 153–168 (2012).
5. Saint Clair, E. C. et al. Near-IR resonance Raman spectroscopy of archaerhodopsin 3: effects of transmembrane potential. *J. Phys. Chem. B* **116**, 14592–14601 (2012).
6. Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R. & Piston, D. W. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys. J.* **73**, 2782–2790 (1997).
7. Heim, R. & Tsien, R. Y. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.* **6**, 178–182 (1996).
8. Zhao, Y., Hochbaum, D., Harrison, D. J., Cohen, A. E. & Campbell, R. E. A comprehensive live cell screening approach for developing improved microbial rhodopsin-based voltage biosensors. *Biophys. J.* **106**, 415a (2014).

9. Flytzanis, N. C. et al. Archaerhodopsin variants with enhanced voltage-sensitive fluorescence in mammalian and Caenorhabditis elegans neurons. *Nat. Commun.* **5**, 1–9 (2014).
10. 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).
11. Maclaurin, D., Venkatachalam, V., Lee, H. & Cohen, A. E. Mechanism of voltage-sensitive fluorescence in a microbial rhodopsin. *Proc. Natl Acad. Sci.* **110**, 5939–5944 (2013).
12. Hochbaum, D. R. et al. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat. Methods* **11**, 825–833 (2014).
13. McIsaac, R. S. et al. Directed evolution of a far-red fluorescent rhodopsin. *Proc. Natl Acad. Sci.* **111**, 13034–13039 (2014).
14. Piatkevich, K. D. et al. A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat. Chem. Biol.* **14**, 352–360 (2018).
15. Silapetere, A. et al. QuasAr Odyssey: the origin of fluorescence and its voltage sensitivity in microbial rhodopsins. *Nat. Commun.* **13**, 1–20 (2022).
16. Dempsey, G. T. et al. Cardiotoxicity screening with simultaneous optogenetic pacing, voltage imaging and calcium imaging. *J. Pharm. Toxicol. Methods* **81**, 240–250 (2016).
17. Azimi Hashemi, N. et al. Rhodopsin-based voltage imaging tools for use in muscles and neurons of Caenorhabditis elegans. *Proc. Natl Acad. Sci.* **116**, 17051–17060 (2019).
18. Lou, S. et al. Genetically targeted all-optical electrophysiology with a transgenic cre-dependent optopatch mouse. *J. Neurosci.* **36**, 11059–11073 (2016).
19. Ernst, O. P. et al. Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. *Chem. Rev.* **114**, 126–163 (2014).
20. Romei, M. G., Lin, C.-Y., Mathews, I. I. & Boxer, S. G. Electrostatic control of photoisomerization pathways in proteins. *Science* **367**, 76–79 (2020).
21. Granovsky, A. A. Extended multi-configuration quasi-degenerate perturbation theory: The new approach to multi-state multi-reference perturbation theory. *J. Chem. Phys.* **134**, 214113 (2011).
22. 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).
23. Battaglia, S. & Lindh, R. Extended Dynamically Weighted CASPT2: The Best of Two Worlds. *J. Chem. Theory Comput.* **16**, 1555–1567 (2020).
24. del Carmen Marín, M. et al. Fluorescence enhancement of a microbial rhodopsin via electronic reprogramming. *J. Am. Chem. Soc.* **141**, 262–271 (2018).
25. Schapiro, I. et al. The ultrafast photoisomerizations of rhodopsin and bathorhodopsin are modulated by bond length alternation and HOOP driven electronic effects. *J. Am. Chem. Soc.* **133**, 3354–3364 (2011).
26. Schnedermann, C. et al. Evidence for a vibrational phase-dependent isotope effect on the photochemistry of vision. *Nat. Chem.* **10**, 449–455 (2018).
27. Frutos, L. M., Andruniów, T., Santoro, F., Ferré, N. & Olivucci, M. Tracking the excited-state time evolution of the visual pigment with multiconfigurational quantum chemistry. *Proc. Natl Acad. Sci.* **104**, 7764–7769 (2007).
28. Atchity, G. J., Xantheas, S. S. & Ruedenberg, K. Potential energy surfaces near intersections. *J. Chem. Phys.* **95**, 1862–1876 (1991).
29. Laricheva, E. N. et al. Origin of fluorescence in 11-cis locked bovine rhodopsin. *J. Chem. Theory Comput.* **8**, 2559–2563 (2012).
30. Penzkofer, A., Silapetere, A. & Hegemann, P. Photocycle dynamics of the Archaerhodopsin 3 based fluorescent voltage sensor Archon2. *J. Photochem Photobio. B* **225**, 112331 (2021).
31. Penzkofer, A., Silapetere, A. & Hegemann, P. Photocycle dynamics of the Archaerhodopsin 3 based fluorescent voltage sensor QuasAr1. *Int J. Mol. Sci.* **21**, 160 (2019).
32. Mayer, I. Bond orders and valences from ab initio wave functions. *Int J. Quantum Chem.* **29**, 477–483 (1986).
33. Gozem, S., Schapiro, I., Ferré, N. & Olivucci, M. The molecular mechanism of thermal noise in rod photoreceptors. *Science* **337**, 1225–1228 (2012).
34. Bonačić-Koutecký, V., Koutecký, J. & Michl, J. Neutral and charged biradicals, zwitterions, funnels in S_1, and proton translocation: their role in photochemistry, photophysics, and vision. *Angew. Chem. Int. Ed. Engl.* **26**, 170–189 (1987).
35. Hammond, G. S. A correlation of reaction rates. *J. Am. Chem. Soc.* **77**, 334–338 (1955).
36. Leffler, J. E. Parameters for the description of transition states. *Science (1979)* **117**, 340–341 (1953).
37. Manathunga, M. et al. Probing the photodynamics of rhodopsins with reduced retinal chromophores. *J. Chem. Theory Comput.* **12**, 839–850 (2016).
38. Pedraza-González, L., De Vico, L., del Carmen Marín, M., Fanelli, F. & Olivucci, M. a-ARM: automatic rhodopsin modeling with chromophore cavity generation, ionization state selection, and external counterion placement. *J. Chem. Theory Comput.* **15**, 3134–3152 (2019).
39. Broser, M. et al. NeoR, a near-infrared absorbing rhodopsin. *Nat. Commun.* **11**, 1–10 (2020).
40. Case, D. A. et al. The Amber biomolecular simulation programs. *J. Comput Chem.* **26**, 1668–1688 (2005).
41. Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.* **93**, 1061–1071 (1990).
42. Fdez. Galván, I. et al. OpenMolcas: from source code to insight. *J. Chem. Theory Comput* **15**, 5925–5964 (2019).
43. Aquilante, F. et al. Molcas 8: New capabilities for multi-configurational quantum chemical calculations across the periodic table. J Comput Chem. **37**, 506–41 (2016).
44. Ponder, J. W. et al. *TINKER: Software tools for molecular design*, vol. 3 (Washington University School of Medicine, Saint Louis, MO, 2004).

Acknowledgements

The research has been partially supported by the following grants: NSF CHE-CLP-1710191, NIH GM126627-01, USIAS 2015, the Ohio Supercomputer Center, the MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca) for a "Dipartimento di Eccellenza 2018-2022" and the Fondazione Banca d'Italia to M.O. The MIUR is also acknowledged for a Rita Levi Montalcini grant to D.P. N.F. and M.H.-R. acknowledge the financial support by Agence Nationale de la Recherche (project ULTRArchea, grant ANR-21-CE11-0029-03). L.B. and M.O. acknowledge partial support from European Union, Next Generation EU, MIUR Italia Domani Progetto mRNA Spoke 6 del "National Center for Gene Therapy and Drugs based on RNA Technology". CUP di progetto B63C22000610006. The authors are grateful to Xuchun Yang and María del Carmen Marín for fruitful discussion.

Author contributions

Conceptualization and Project administration: M.O., Methodology: L.P.G., E.M., L.B., D.P., L.D.V., N.F., M.H.R., M.F., D.K., A.B.G., Investigation and Visualization: L.B., E.M., M.O., D.K., Funding acquisition: M.O., D.P., Supervision: D.P., L.D.V., M.O., Writing and Reviewing: M.O., L.B.

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/s41467-022-33993-4.

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

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

Publication: Palombo, R., Barneschi, L., Pedraza-González, L., Padula, D., Schapiro, I. and Olivucci, M., 2022. Retinal chromophore charge delocalization and confinement explain the extreme photophysics of Neorhodopsin. Nature communications, 13(1), pp.1-9.

Contribution: Theoretical calculations, discussion, graphics, revision.

Current Status: Published, November 4, 2022.

nature communications

Article

https://doi.org/10.1038/s41467-022-33953-y

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

Received: 21 April 2022

Accepted: 7 October 2022

Riccardo Palombo[1,2], Leonardo Barneschi[1], Laura Pedraza-González[1], Daniele Padula[1], Igor Schapiro[3] & 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 unprecedent case study that could potentially expand our comprehension of red-shifted and highly fluorescent rhodopsins. NeoR is a rhodopsin guanylyn-cyclase (RGC) expressed in the *Rhizoclosmatium globosum* from Chytridiomycota, the only phylium 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 (λ^{a}_{max}), respectively. In contrast, NeoR displays the strongest bathochromic shift among all known microbial rhodopsins, yielding an extremely red-shited ($\lambda^{a}_{max} = 690$ nm) absorption band. Such a band is mirrored by an intense emission band with a maximum (λ^{f}_{max}) 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 protein[8] (GFP), represents an anomaly in the rhodopsin superfamily and suggests an evolution-driven origin. More specifically,

---

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

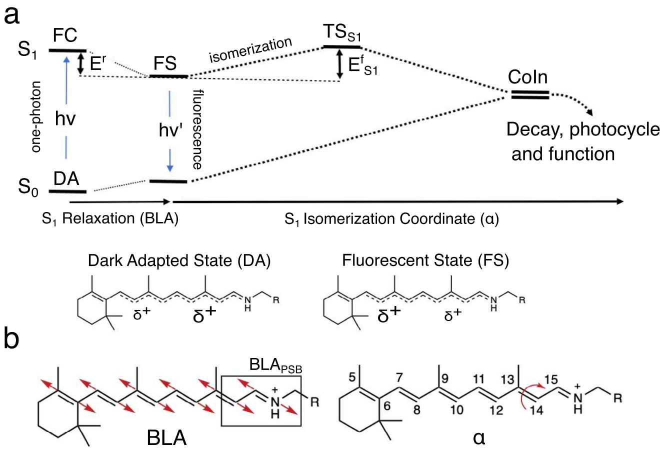


Fig. 1 | Geometrical and electronic structure changes in NeoR. a Schematic representation of the hypothetic S_0 and S_1 energy changes occurring along the S_1 relaxation that involves the bond length alternation (BLA, quantified by the difference between the average of the double-bond lengths and the 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^f_{S1}) 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).

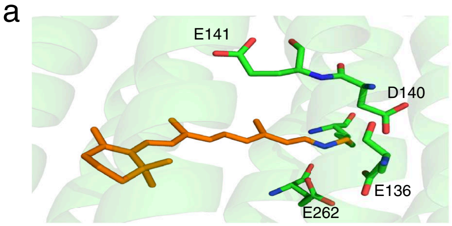
In order to pursue the objectives above, we construct a quantum-mechanics/molecular mechanics (QM/MM) model of NeoR based on multiconfigurational quantum chemistry. Since a crystallographic structure is not available, we employ, for the model construction, a previously reported comparative model[7]. 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^f_{S1}. 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^f_{S1} 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^a_{max} 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 ambiguous[28]. 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^a_{max}) and emission (λ^f_{max}) maxima as well as the relaxation energy (E^r) defined by the basic mechanism of Fig. 1a. The λ^a_{max} and λ^f_{max} 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^r 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^a_{max}, λ^f_{max}, and E^r 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.



b

Counterion ID	λ^{a}_{max} (nm)	λ^{f}_{max} (nm)	Stokes shift (nm)	E^{r} (kcal mol^{-1})
No counterion[a]	678	NP	NP	NP
E136	582	892	310	8.96
D140	568	896	328	10.84
E141	644	693	49	3.40
E262	602	892	290	9.05
E136 D140	364	596	232	10.44
E136 E141	539	815	276	9.62
E136 E262	439	597	158	10.41
E141 E262	545	833	288	10.24
D140 E141	522	795	273	9.30
D140 E262	423	677	246	11.50
E136 D140 E141	387	602	215	15.01
E136 D140 E262	347	674	327	27.53
E136 E141 E262	393	607	214	14.14
D140 E141 E262	396	596	200	15.85

Experimental values: λ^{a}_{max}= 690 nm, λ^{f}_{max}= 707 nm, Stokes shift = 17 nm

[a] This model does not exhibit a stable FS

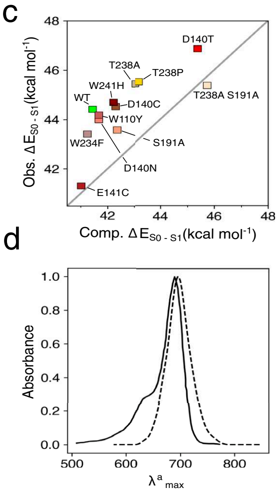


Fig. 2 | Choice of the NeoR chromophore counterion and model assessment. **a** Overview of the structure of the all-trans rPSB chromophore (orange) and its four potential residue counterions (in green). The lysine residue (in green) bounded to the rPSB chromophore is also displayed. **b** Computed (CASPT2 level) maximum absorption wavelength (λ^{a}_{max}), maximum emission wavelength (λ^{f}_{max}) and relaxation energy (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^{a}_{max} 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 *a*-ARM$_{E141}$) is the most accurate. In fact, *a*-ARM$_{E141}$ yields λ^{a}_{max} and λ^{f}_{max} 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^{a}_{max} 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 bands[7], the *a*-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 *a*-ARM$_{E141}$, we constructed the models of a set of NeoR variants whose λ^{a}_{max} values have been experimentally measured (see Supplementary Table 3). As shown in Fig. 2c the models reproduce the observed trend indicating that a-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 *a*-ARM models[20,21,23,25–27,29]. We also used *a*-ARM$_{E141}$ to simulate the WT NeoR absorption band at room temperature by computing the ΔE_{S0-S1} 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^{a}_{max} value and the computed band half-width is close to that seen experimentally. *a*-ARM$_{GLU141}$ only appears to miss a shoulder at 640 nm that is, likely, of vibronic origin[7] 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 *a*-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

nature communications

Article

https://doi.org/10.1038/s41467-022-33953-y

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

Received: 21 April 2022

Accepted: 7 October 2022

Riccardo Palombo[1,2], Leonardo Barneschi[1], Laura Pedraza-González[1],
Daniele Padula[1], Igor Schapiro[3] & 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 unprecedent case study that could potentially expand our comprehension of red-shifted and highly fluorescent rhodopsins. NeoR is a rhodopsin guanylyn-cyclase (RGC) expressed in the *Rhizoclosmatium globosum* from Chytridiomycota, the only phylium 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 (λ^{a}_{max}), respectively. In contrast, NeoR displays the strongest bathocromic shift among all known microbial rhodopsins, yielding an extremely red-shited ($\lambda^{a}_{max} = 690$ nm) absorption band. Such a band is mirrored by an intense emission band with a maximum (λ^{f}_{max}) 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 protein[8] (GFP), represents an anomaly in the rhodopsin superfamily and suggests an evolution-driven origin. More specifically,

---

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

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^f_{S1} values possibly represent upper limits as a sloped CoIn 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^f_{S1} 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 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 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^f_{S1} 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 CoIn) 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^f_{S1} 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 CoIn 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^f_{S1} 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 CoIn 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^f_{S1} 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

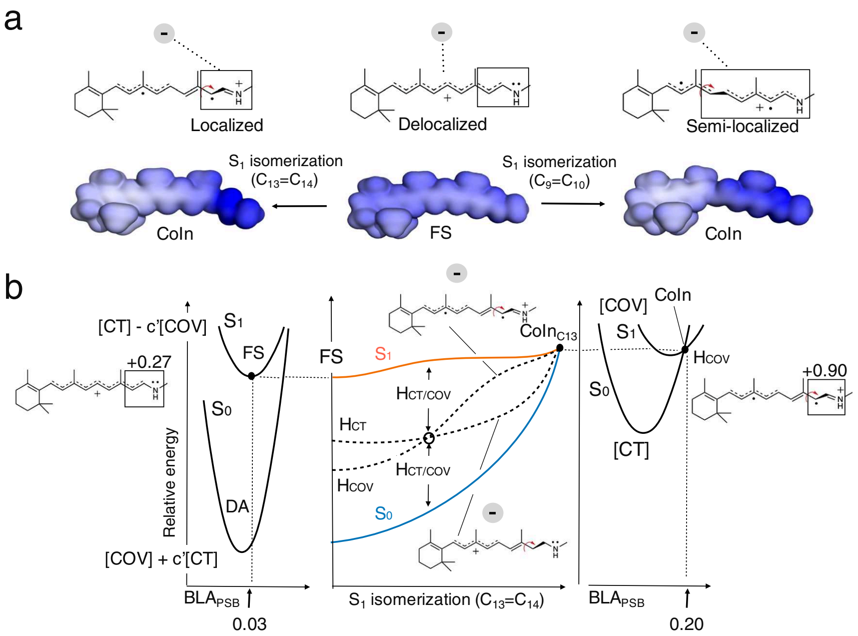


Fig. 5 | Origin of the E^f_{S1} barrier in NeoR. a Representation of the increase in the distance between the negative and positive charge centroids due to the positive charge confinement along the C13=C14 and C9=C10 photoisomerization paths. Comparison between electrostatic potential (ESP) maps indicates 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 (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 CoIn. 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 CoIn (right) and are in line with the presented FC→FS and CoIn 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 CoIn 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 CoIn 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 CoIn 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 NeoR[2]. 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^a_{max} changes observed in a set of rhodopsin mimics based on the human cellular retinol-binding protein II (hCRBPII)[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^a_{max} 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 Fig. 3a) or, in a different language, to a near cyanine limit situation[43].

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 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^{a}_{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 a-ARM protocol[20–22] and based on the comparative model structure built and validated by S. Adam et al.[7]. Further details about the a-ARM protocol are given as Supplementary Informations (Supplementary Section 1). After initially producing, automatically, default a-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 Mocas/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 v8.2[49]. The collection of geometries connecting the FS to the different CoIn'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^{f}_{S1} is estimated after revaluating 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 CoIn) 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

1. Xu, Y., Zou, P. & Cohen, A. E. Voltage imaging with genetically encoded. indicators. *Curr Opin Chem Biol.* **39**, 1–10 (2017).
2. Lehtinen, K., Nokia, M. S. & Takala, H. Red light optogenetics in neuroscience. Front. Cell. Neurosci. **15**, (2022).
3. 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).
4. 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).
5. 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).
6. James, T. Y. et al. A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). *Mycologia* **98**, 860–871 (2006).
7. Broser, M. et al. NeoR, a near-infrared absorbing rhodopsin. *Nat. Commun.* **11**, 5682 (2020).
8. 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).
9. Andruniów, T.; O. M. QM/MM studies of light-responsive biological systems. (Springer, 2021).
10. 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).
11. 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).
12. Kouyama, T., Kinosita, K. & Ikegami, A. Excited-state dynamics of bacteriorhodopsin. *Biophys. J.* **47**, 43–54 (1985).
13. Engqvist, M. K. M. et al. Directed evolution of gloeobacter violaceus rhodopsin spectral properties. *J. Mol. Biol.* **427**, 205–220 (2015).
14. McIsaac, R. S. et al. Directed evolution of a far-red fluorescent rhodopsin. *Proc. Natl Acad. Sci. USA* **111**, 13034–13039 (2014).
15. Silapetere, A. et al. QuasAr Odyssey: the origin of fluorescence and its voltage sensitivity in microbial rhodopsins. *Nat. Commun.* **13**, 5501 (2022).
16. Ganapathy, S. et al. Retinal-based proton pumping in the near infrared. *J. Am. Chem. Soc.* **139**, 2338–2344 (2017).
17. 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).
18. Hochbaum, D. R. et al. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat. Methods* **11**, 825–833 (2014).
19. Piatkevich, K. D. et al. A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat. Chem. Biol.* **14**, 352–360 (2018).
20. Melaccio, F. et al. Toward automatic rhodopsin modeling as a tool for high-throughput computational photobiology. *J. Chem. Theory Comput.* **12**, 6020–6034 (2016).
21. Pedraza-González, L. et al. *a* -ARM: automatic rhodopsin modeling with chromophore cavity generation, ionization state selection, and external counterion placement. *J. Chem. Theory Comput.* **15**, 3134–3152 (2019).
22. 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. Mroginski, M. et al. Frontiers in multiscale modeling of photoreceptor 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 Enzymerhodopsin. *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.

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.

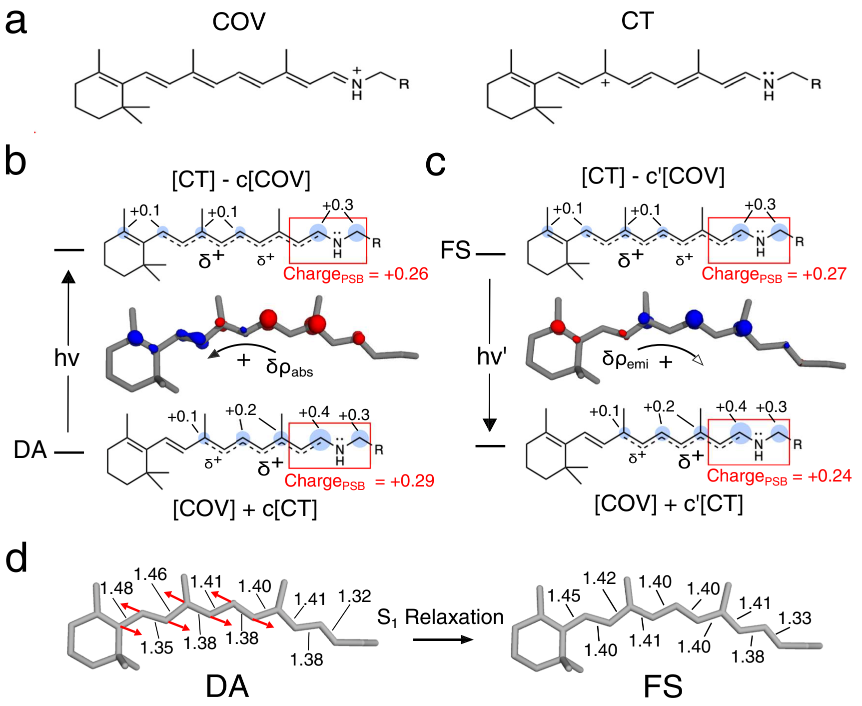


Fig. 3 | Electronic and geometrical character of S_1 relaxation in NeoR chromophore. a Representation of the two limiting resonance formulas adopted to describe the electronic character of the rPSB chromophore. **b** Electron density variation ($\delta\rho_{abs}$) characterizing the vertical S_0→S_1 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-Cε- rPSB fragment is also given. **c** Same data for the S_1→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-Cε 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 diabatics (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 proteins[17].

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→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→S_1 excitation[32,33].

Notice that the chromophore charge delocalization seen in the selected a-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^a_{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→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 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 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 $E^f \approx 3.5$ kcal mol^{-1}

PAPER III Pro219 is an electrostatic color determinant in the light-driven sodium pump KR2

Publication: Nakajima, Y., Pedraza-González, L., Barneschi, L., Inoue, K., Olivucci, M. and Kandori, H., 2021. Pro219 is an electrostatic color determinant in the light-driven sodium pump KR2. Communications biology, 4(1), pp.1-15.

Contribution: Theoretical calculations, code development, discussion, graphics, revision.

Current Status: Published, October 13, 2021

communications
biology

ARTICLE

https://doi.org/10.1038/s42003-021-02684-z OPEN

Pro219 is an electrostatic color determinant in the light-driven sodium pump KR2

Yuta Nakajima[1,6], Laura Pedraza-González[2,6], Leonardo Barneschi[2], Keiichi Inoue [3], Massimo Olivucci [2,4]✉ & Hideki Kandori [1,5]✉

Color tuning in animal and microbial rhodopsins has attracted the interest of many researchers, as the color of their common retinal chromophores is modulated by the amino acid residues forming the chromophore cavity. Critical cavity amino acid residues are often called “color switches”, as the rhodopsin color is effectively tuned through their substitution. Well-known color switches are the L/Q and A/TS switches located in the C and G helices of the microbial rhodopsin structure respectively. Recently, we reported on a third G/P switch located in the F helix of the light-driven sodium pumps of KR2 and *Js*NaR causing substantial spectral red-shifts in the latter with respect to the former. In order to investigate the molecular-level mechanism driving such switching function, here we present an exhaustive mutation, spectroscopic and computational investigation of the P219X mutant set of KR2. To do so, we study the changes in the absorption band of the 19 possible mutants and construct, semi-automatically, the corresponding hybrid quantum mechanics/molecular mechanics models. We found that the P219X feature a red-shifted light absorption with the only exception of P219R. The analysis of the corresponding models indicate that the G/P switch induces red-shifting variations via electrostatic interactions, while replacement-induced chromophore geometrical (steric) distortions play a minor role. However, the same analysis indicates that the P219R blue-shifted variant has a more complex origin involving both electrostatic and steric changes accompanied by protonation state and hydrogen bond networks modifications. These results make it difficult to extract simple rules or formulate theories for predicting how a switch operates without considering the atomistic details and environmental consequences of the side chain replacement.

---

[1] Department of Life Science and Applied Chemistry, Nagoya Institute of Technology, Showa-ku, Nagoya 466-8555, Japan. [2] Dipartimento di Biotecnologie, Chimica e Farmacia, Università degli Studi di Siena, Via Aldo Moro, 2, I-53100 Siena, Italy. [3] The Institute for Solid State Physics, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8581, Japan. [4] Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403, USA. [5] OptoBioTechnology Research Center, Nagoya Institute of Technology, Showa-ku, Nagoya 466-8555, Japan. [6] These authors contributed equally: Yuta Nakajima, Laura Pedraza-González. ✉email: massimo.olivucci@unisi.it; kandori@nitech.ac.jp

Microbial or animal rhodopsins contain either an all-*trans* or 11-*cis* retinal chromophore respectively[1]. In both cases the chromophore is located inside the seven-helix (TM1 to TM7) transmembrane structure of diverse opsins where it binds, covalently, a lysine residue to form a Schiff base linkage[2]. In such an environment the chromophore spectroscopic and reactivity properties are modulated by its molecular environment as demonstrated by the variety of displayed colors and, therefore, variations in the wavelength (λ_{max}) of the maximum of the absorption bands[3–9]. The mechanism allowing such "color tuning" effect is an important topic in rhodopsin research since the λ_{max} value of the light captured by a specific rhodopsin represents a biological functions[10–14]. While the color tuning mechanism[7–9] is still not fully understood, it is apparent that it must be determined by the interactions between the chromophore and the surrounding amino acid residues featuring side-chains which may be charged, dipolar, aromatic and capable of hydrogen-bonding and steric contact effects. Accordingly, learning the precise "rules" for controlling such interaction appears to be of basic importance for the understanding of different facets of rhodopsin biology including their evolution, ecology, biophysics and the laboratory engineering required for optogenetic applications[15–17].

In most cases, the all-*trans* retinal chromophore of microbial rhodopsins features a protonated, Schiff base linkage (rPSBAT) and such -CH = NH(+)- state is stabilized by a negatively charged -COO(−) counterion, a glutamate or aspartate, placed in its vicinity. While the electrostatic interaction between chromophore and counter-ion is prominent in color tuning, other specific color determining residues have been reported, which are sometimes called "color switches". In the case of microbial rhodopsin, a famous color switch is the "L/Q switch" in proteorhodopsin[10,18]. In fact, green-absorbing (GPR) and blue-absorbing (BPR) proteorhodopsins contain Leu and Gln at position 105 of C-helix (TM3), respectively. It has been proposed that the L/Q selection provides a mechanism for optimizing light absorption with respect to specific environmental light conditions[10,19]. In other words, the color regulation would enable bacteria living in shallow and deep ocean waters to use green and blue light respectively which is abundant in these respective habitats. Although Leu is hydrophobic and Gln is hydrophilic, comprehensive mutation study of L105 in GPR indicated that molecular volume (i.e., steric interactions), not hydrophobicity, is correlated with the λ_{max} value[20].

Another color switch is the "A/T switch"[3]. The alanine residue at position of 215 in bacteriorhodopsin (BR) is known to partly contribute to the spectral difference between the BR and *Natronomonas pharaonis* sensory rhodopsin II (*Np*SRII, also *pharaonis* phoborhodopsin) as it is replaced with Thr in *Np*SRII[3]. This is considered to be related to the evolution from BR to *Np*SRII[21]. A previous study indicates that the S254 of a bacterial light-driven sodium pump from *Krokinobacter eikastus* (KR2) also plays a role similar to *Np*SRII and BR[22]. Hence, here we refer to it as the "A/TS switch", as it appears to represent an additional example of naturally occurring color determining residue. This switch satisfies the general principle of color tuning, where introduction of a polar residue in the vicinity of the β-ionone ring or the Schiff base moiety of the chromophore causes spectral red and blue shift, respectively[14,23–25]. This observation is in line with the present theoretical understanding of color tuning based on the electronic structure and, therefore, opposite positive charge distribution, of the ground (S_0) and first excited (S_1) states of rPSBAT. In short, while in S_0 the positive charge is mainly localized in the -CH = NH- moiety of the chromophore, in the S_1 the positive charge is delocalized towards the β-ionone ring. Accordingly, negatively charge atoms located in the vicinity of the Schiff base moiety would stabilize S_0 with respect to S_1 leading to a blue-shift of the λ_{max} value. In contrast, if the negative atoms are located in the vicinity of the β-ionone ring they would stabilize S_1 with respect to S_0 leading to a red-shift. Of course, positively charged atoms will have an opposite effect[7].

Recently, a combined experimental and computational study carried out by some of the authors revealed that certain mutations at P219 in KR2 led to a spectral red-shift with no loss of its sodium pump function[22]. As far as color tuning is concerned, Pro is an unusual residue. In fact, this residue is highly conserved in most microbial rhodopsins, but two light-driven sodium pumping rhodopsins (NaRs) that do not conserve it were identified from *Parvularcula oceani*[26] (*Po*NaR) and *Jannaschia seosinensis*[27] (*Js*NaR). These rhodopsins display a Thr and Gly residue at the P219 position of KR2, respectively (see Fig. 1). Although the former was reported in the mention previous study, it was not expressed in *E. coli* cells[28]. In contrast, it was possible to express *Js*NaR, and a mutation study showed that it represents a third color switch that was named "G/P switch"[22]. This role of residue 219 is supported by the fact that the λ_{max} value of KR2 (525 nm) is red-shifted to 535 nm in its P219G mutant and the fact that the *Js*NaR λ_{max} value (550 nm) is blue-shifted to 523 nm in its G216P mutant. However, the exact mechanism that is at the basis of such color tuning effect is unknown.

We find that KR2 represents a suitable laboratory model for an in-depth investigation of the G/P switching mechanism. At the same time a molecular-level understanding of such a phenomenon must necessarily be based on the construction of a light-responsive computational model of the protein. Such a model has to incorporate a description of the electronic structure of rPSBAT. Accordingly, in this work we present an exhaustive investigation of the color tuning induced by point mutation at the P219 position of KR2. We prepare all possible P219X mutants where X stands for all alternative 19 natural amino acids. The resulting color variations, spanning a red-to-blue range going from 545 nm to 515 nm, provides a unique basis for molecular-level mechanistic studies performed using hybrid quantum mechanics/molecular mechanics (QM/MM) models generated with the Automatic Rhodopsin Modeling (*a*-ARM) protocol[9], equipped with multiconfigurational second order perturbation theory (see "Materials and Methods" section). By systematically

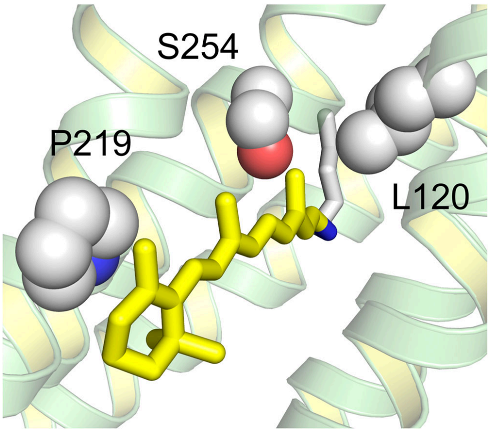


Fig. 1 Color switches of microbial rhodopsins, shown in the structure of a light-driven sodium pump KR2. KR2 has L120, S254, and P219 in the L/Q switch, A/TS switch, and G/P switch, respectively.

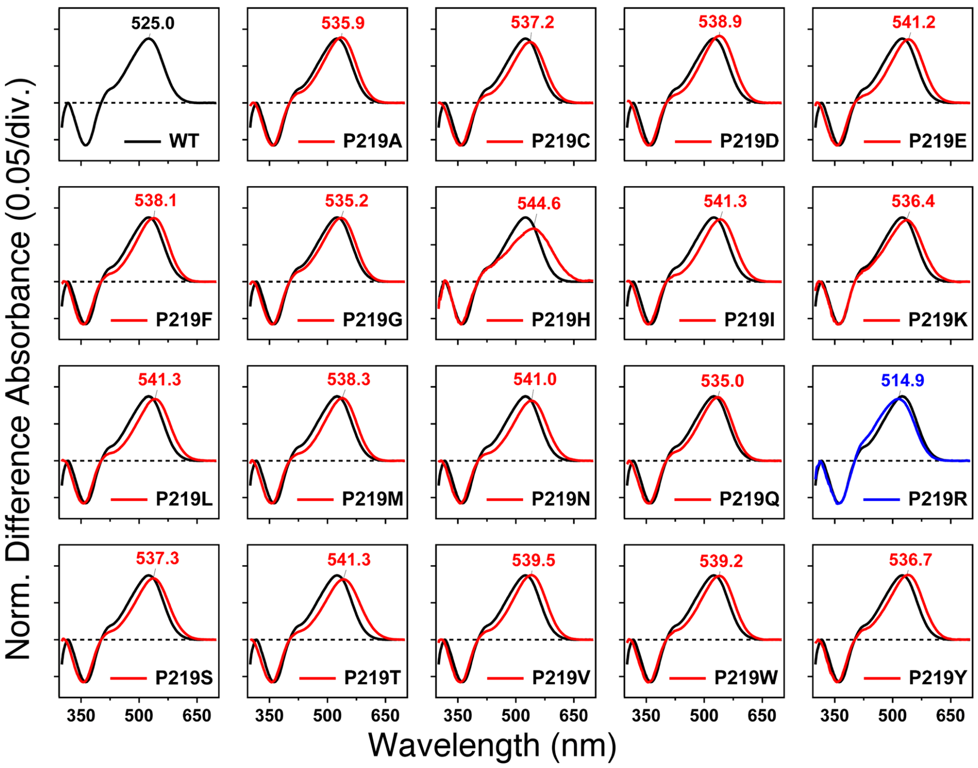


Fig. 2 Light-induced difference absorption spectra of the WT-KR2 (black curves) and 19 P219X mutants (colored curves) of KR2 in the presence of 500 mM hydroxylamine. Positive and negative signals show the spectra before and after illumination corresponding to those of the rhodopsin and retinal oxime, respectively. Red and blue curves represent red- and blue-shifted mutants, respectively. Mutant and WT-KR2 spectra were normalized by use of a negative peak at 361 nm.

building and analyzing congruously built QM/MM models of the wild-type and mutants (20 models in total), we support the hypothesis that the 219 position induces color changes through, essentially, electrostatic effects while steric effects contribute to a single, blue-shifting P219R mutant. Thus, the putative P/X color switch would be, mostly, electrostatically driven when causing a red-shift but uses a more complex mechanism for a blue-shift. To the best of our knowledge, this manuscript represents the first study of site-saturation mutagenesis for a color switch that has been carried out both experimentally and computationally on the same mutant set.

Results and discussion

Absorption spectra of P219X mutants of KR2. We attempted to express the 19 different P219X mutants of KR2 in *E. coli* where all-*trans* retinal was added at the induction period to produce the rPSBAT chromophore. Supplementary Note 1 shows pictures of WT and 19 P219 mutants (see Supplementary Fig. 1). As is seen from the color, expression was much lower for basic amino acids such as His, Lys and Arg. The expression levels are quantitatively compared in Supplementary Fig. 2, where all the P219X mutants are enough expressed to test their absorption maxima.

To quantify the absorption properties of each mutant, the sample was illuminated in the presence of hydroxylamine. This process converts protein-bound retinal chromophore into retinal oxime by light, so that one can easily obtain the λ_{max} of each protein without purification. Figure 2 shows the change in absorption (before-minus-after illumination) representing the photobleaching of WT-KR2 and 19 mutants in the presence of 500 mM hydroxylamine. In WT-KR2 (black curve), positive and negative peaks appeared at 525 and 361 nm, corresponding to the unphotolyzed protein and retinal oxime, respectively. The mutant spectra were normalized to the WT-KR2 spectrum by use of a negative peak at 361 nm.

Blue and red curves in Fig. 2 show light-induced difference absorption spectra (before-minus-after illumination) of mutants, which exhibit spectral blue and red shifts, respectively. While negative peaks due to retinal oxime were identical for WT-KR2 and mutants, all mutants showed spectral red-shifts, corresponding to λ_{max} in the 535–541 nm range. The only exception was P219R, whose λ_{max} is located at 515 nm. It is likely that Arg is positively charged even in the hydrophobic environment surrounding position 219. Consistently with the color tuning theory mentioned above, such positive charge, placed near the β-ionone ring of rPSBAT, would cause a spectral blue shift. On the other hand, in Fig. 3 we report the observed spectral shift in energy (in wavenumber) versus volume (Fig. 3a) and hydropathy (Fig. 3b) of amino acids. The figure shows that for 18 mutants the

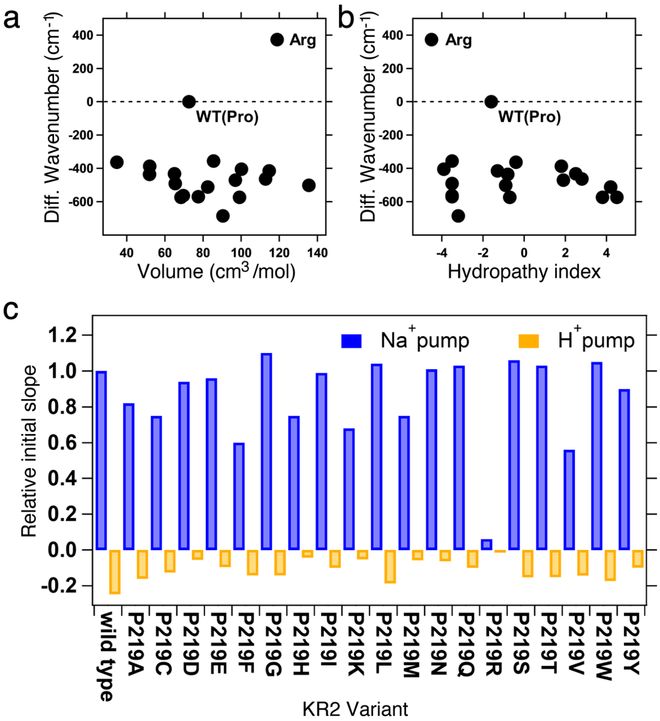


Fig. 3 Experimental Data. Correlation between the absorption light energy and (**a**) the volume or (**b**) hydropathy index of the amino acid residue at position 219 in KR2. The y-axis represents the difference in wavenumber from that of the wild type (reciprocal λ_{max}; cm^{-1}), where positive and negative values correspond to the spectral blue- and red-shifts, respectively. (**c**) Quantitative comparison of pump activities of the WT-KR2 and mutant KR2. The numbers of protons taken in or released from the cells by the pump activity of KR2 and each mutant per one protein molecule in one second are shown. The values were estimated from the initial slope of light induced pH changes and the expression levels of the proteins. *E. coli* cells suspensions expressing KR2 mutants are illuminated at >520 nm light, and quantified proton release and uptake are measured in the solution containing 100 mM NaCl (blue bars) and CsCl (orange bars). Positive values indicate the numbers of protons which are taken into cells by one protein molecule per second, which originates from outward Na^+ pump. Negative values indicate the numbers of protons which are released from cells by one protein molecule per second, which originates from outward proton pump.

λ_{max} correlate with neither volume nor hydropathy, whereas the blue-shifted Arg is unique as it has both higher volume and lower hydropathy than Pro. The higher volume of Arg would suggest that the observed blue-shift is related to a steric effect. Below we show that, according to our computational analysis, this effect alone would not explain the observed P219R spectral change.

Figure 3c compares ion-transport activity using a pH electrode. The sodium and proton pump activities of the WT-KR2 is maintained except for P219R. This fact suggests that the addition of a positive charge at that specific position is responsible for the lack of ion transport. Both absorption spectra (Fig. 2) and transport activity (Fig. 3c) suggest that other amino acid replacements lead to neutral side-chain. In fact, it is reasonable for Asp, Glu, and His to be neutral in the hydrophobic environment. In addition, Lys may also be neutral in the P219K mutant. All these assigned protonation states are consistent with those of the constructed *a*-ARM QM/MM models, as described in next section.

Construction of the WT model. The QM/MM models (see Fig. 4a) were generated by using the *a*-ARM protocol, whose workflow is illustrated in Fig. 4b, c and described in the "Materials and Methods" section (see Supplementary Note 2 and Supplementary Note 3). The *a*-ARM$_{default}$ model for the WT-KR2 was automatically generated taking, as the only input, the X-ray crystallographic of the pentameric form resolved at 2.2 Å (PDB ID 6REW[29]). The following parameters were automatically selected by the Input file generator: Chain A of the X-ray crystallographic structure; rotamer N84 with occupancy number of 0.50 (step 1 in Fig. 4b); 24 chromophore cavity residues defined based on Voronoi tessellation and alpha spheres theory, and including the K255 residue covalently linked to the chromophore, plus the D116 main (MC), D251 secondary (SC) counter-ion residues and 3 molecules of water (step 2 in Fig. 4b); protonation states predicted at pH 5.2, as: neutral E160 (step 4 in Fig. 4b); and the inclusion of 8 Cl^- inner (IS) 7 Na^+ outer (OS) external counter-ions, with positions optimized with respect to an electrostatic potential grid constructed around each charged target residue. The Na^+ ion present in the X-ray structure was kept, for a total of 8 Na^+ ions in the model (step 5 in Fig. 4b). Such parameters ultimately led to an *a*-ARM$_{default}$ model producing a λ_{max} (in terms of the average vertical excitation energy of 10

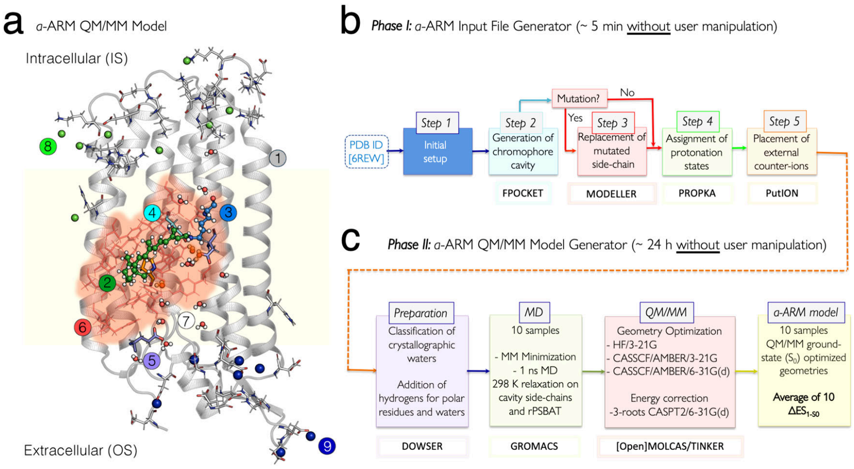


Fig. 4 Structure of *a*-ARM protocol. (**a**) General scheme of a monomeric, gas-phase and globally uncharged QM/MM model for the WT-KR2 rhodopsin, generated by the *a*-ARM rhodopsin model building protocol. This is composed of: (1) environment subsystem (silver cartoon), (2) retinal chromophore (green tubes), (3) Lys side-chain covalently linked to the retinal chromophore (blue tubes), (4) main counter-ion MC (cyan tubes), (5) protonated residues, (6) residues of the chromophore cavity subsystem (red tubes), (7) water molecules, and external (8) Cl^- (green balls) and (9) Na^+ (blue balls) counterions. Parts 2 and 3 form the Lys-QM subsystem which includes the H-link atom located along the only bond connecting blue and green atoms. Parts 4 and 6 form the cavity subsystem. The water molecules (Part 7) may be part of the environment or cavity subsystems. The external OS and IS charged residues are shown in frame representation. The residue P219 is presented as orange tubes. (right) General workflow of the *a*-ARM rhodopsin model building protocol for the generation of QM/MM models of wild-type and mutant rhodopsins. The *a*-ARM protocol comprises two phases: (**b**) input file generator phase and (**c**) QM/MM model generator phase. The different software used in each step are also specified.

replicas, see the "Materials and Methods" section) of about 127 nm (0.75 eV in terms of ΔE_{S1-S0}) blue-shifted with respect to the experimental data. In fact, as documented elsewhere[9,30], the default models do not always replicate the correct electrostatics of the chromophore counterion complex in ion-pumping rhodopsins[9,30]. For this reason, a customized model was produced (see Supplementary Fig. 3 and Supplementary Table 1). The model customization employs the a-ARM$_{\text{customized}}$ approach of ref. [9]. (see also Supplementary Note 2) that guarantees reproducibility. While the default model predicts that both aspartic acid residues forming the counterion complex of the rPSBAT, namely D116 and D251[31], are negatively charged, we have observed that the presence of two negative charges in the counterion complex would outbalance the positive charge of the rPSBAT, generating a large, blue-shifting effect[9] in the $\lambda_{\max}$ value. The a-ARM$_{\text{customized}}$ approach allows to reassign the protonation states (step 4 of the input file generator in Fig. 4b) of D116 and D251. More specifically, as illustrated in Supplementary Fig. 3 and Supplementary Table 2, these protonation states are systematically scanned. It was found that a model featuring a protonated (i.e., neutral) secondary counterion (SC) D251 counterbalanced the charge in the vicinity of the rPSBAT by mitigating the overstabilization of the S_0 positive charge of the chromophore Schiff base region yielding a smaller, blue-shifted error of about 14 nm (0.07 eV). Notice that in QM/MM modeling, it is a common practice to evaluate the protonation states of the rPSB counterion complex by looking, as a guidance, at the reproducibility of the experimental $\lambda_{\max}$ (see for instance refs. [13,32,33]).

Construction of the P219X mutant models. The customized protonation states of the WT-KR2 a-ARM$_{\text{customized}}$ model were employed for the construction of the "default" P219X mutant models. Notice that, as in the WT-KR2 case, the initial structure provided to the Input file generator phase for each mutant was the X-ray structure. As a result,16 out of 19 models featured ΔE_{S1-S0} values with an error, with respect to the observed value and therefore lower than 3.0 kcal mol^{-1} (0.13 eV) threshold. Only 3 models, P219R, P219K, and P219H, were further customized. The default P219R model features a 67 nm (0.36 eV) blue-shifted ΔE_{S1-S0} with respect to the observed value. This indicates that the positively charged arginine side-chain, located near the β-ionone ring of rPSBAT, leads to a too strong S_1 destabilization (a positively charge β-ionone region is typical of the S_1 state of the chromophore). In order to moderate such an effect, (i) the D116 was protonated (i.e., neutral) and the D251 was modeled as negatively charged, accordingly with the customization protocol reported in Supplementary Note 2, and (ii) the E160 residue in the R219 region, assigned to be neutral by step 4 of the default approach, was deprotonated assuming that the positive charge in position 219 increases the group pKa (see Supplementary Table 2). Such customization yields a model just 14 nm (0.07 eV) blue shifted. In the case of P219K, the lysine residue had to be assigned as neutral since the protonated lysine selected by the default protocol yielded a strongly blue-shifted ΔE_{S1-S0} value with respect to the experiment. Finally, for P219H, the histidine residue had to be modeled with the neutral tautomer having the ε nitrogen unprotonated (contrary to the default selection).

The final protonation states of the twenty QM/MM models are: neutral D251, E160 for WT-KR2 and P219X with X = A, C, F, G, I, L, M, N, Q, S, T, V, W, Y; neutral D116 for P219R; neutral D251, E160, H219 (with hydrogen in the ε nitrogen) for P219H; neutral D251, E160, K219 for P219K; neutral D251, E160, E219 for P219E; and neutral D251, E160, D219 for P219D. Notice that, when appropriate, the external counter-ions were automatically updated in step 5 (see Fig. 4b), where an additional Na^+ was

Table 1 *a*-ARM QM/MM models for the wild-type KR2 (WT-KR2) rhodopsin and 19 of its mutants (P219X, with X = A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W, Y). First vertical excitation energy (ΔE_{S1-S0}), maximum absorption wavelength (λ^{a}_{max}), transition oscillator strength (f_{Osc}), and difference between calculated and experimental data ($\Delta\Delta E^{Exp,a-ARM}_{S1-S0}$), for the representative QM/MM model.

Variant	Experimental			*a*-ARM (*N* = 1)[a]				Experimental vs *a*-ARM		
	ΔE^{Exp}_{S1-S0} (kcal mol^{-1})	(eV)	$\lambda^{a,Exp}_{max}$ (nm)	ΔE^{a-ARM}_{S1-S0} (kcal mol^{-1})	(eV)	$\lambda^{a,a-ARM}_{max}$ (nm)	f_{Osc}	$\Delta\Delta E^{Exp,a-ARM}_{S1-S0}$ (kcal mol^{-1})	(eV)	(nm)
WT	54.5	2.36	525	56.0	2.43	511	1.15	1.5	0.07	−14
P219A	53.4	2.31	536	55.8	2.42	512	1.16	2.4	0.11	−23
P219C	53.2	2.31	537	55.4	2.40	516	1.20	2.2	0.09	−21
P219D	53.1	2.30	539	55.0	2.39	520	1.22	2.0	0.09	−19
P219E	52.8	2.29	541	54.8	2.37	522	1.24	1.9	0.08	−19
P219F	53.1	2.30	538	54.9	2.38	521	1.22	1.8	0.08	−18
P219G	53.4	2.32	535	55.8	2.42	513	1.17	2.4	0.10	−23
P219H	52.5	2.28	545	54.5	2.36	525	1.22	2.0	0.09	−20
P219I	52.8	2.29	541	55.4	2.40	516	1.22	2.6	0.11	−25
P219K	53.3	2.31	536	54.8	2.38	521	1.22	1.5	0.07	−15
P219L	52.8	2.29	541	55.1	2.39	519	1.20	2.3	0.10	−22
P219M	53.1	2.30	538	54.8	2.37	522	1.21	1.7	0.07	−16
P219N	52.8	2.29	541	54.5	2.36	524	1.21	1.7	0.07	−17
P219Q	53.4	2.32	535	54.9	2.38	521	1.23	1.5	0.06	−14
P219R	55.5	2.41	515	57.0	2.47	501	1.12	1.5	0.06	−13
P219S	53.2	2.30	538	54.7	2.37	523	1.24	1.5	0.07	−15
P219T	52.8	2.29	541	55.5	2.40	516	1.19	2.6	0.11	−26
P219V	53.0	2.30	540	54.6	2.37	523	1.31	1.6	0.07	−16
P219W	53.0	2.30	539	55.1	2.39	519	1.22	2.1	0.09	−20
P219Y	53.3	2.31	537	55.2	2.39	518	1.19	1.9	0.08	−18
							MAE	1.9		
							AD_{max}	2.6		
							MAD	0.3		

[a] Replica with ΔE^{a-ARM}_{S1-S0} closest to the average.
MAE: Mean absolute error, AD_{max}: Maximum absolute error, MAD: Mean absolute deviation. See definitions in Supplementary Note 6.

included in the OS. The other parameters for the customized inputs (i.e., chain, rotamers, cavity) remain the same as the described for the customized WT-KR2 model in Supplementary Table 1.

Simulation of the λ_{max} variation. As mentioned above, the *a*-$ARM_{customized}$ model of WT-KR2 has been selected as a suitable template for mutant modeling as it reproduces the observed λ_{max} (525 nm) equivalent to a vertical excitation energy ΔE^{Exp}_{S1-S0} = 54.5 kcal mol^{-1}[22] (2.36 eV). The discrepancy falls within the 3.0 kcal mol^{-1} (0.13 eV) error bar established for the *a*-ARM protocol[9,34,35]. This result is reported in Table 1, together with both the computed (ΔE^{a-ARM}_{S1-S0}) and observed (ΔE^{Exp}_{S1-S0}) vertical excitation energies, the corresponding oscillator strengths (f_{Osc}), and their difference $\Delta\Delta E^{Exp,a-ARM}_{S1-S0}$ for the representative model (i.e., replica with ΔE^{a-ARM}_{S1-S0} closest to the average). This is calculated as $\Delta\Delta E^{Exp,a-ARM}_{S1-S0} = \Delta E^{a-ARM}_{S1-S0} - \Delta E^{Exp}_{S1-S0}$. The WT-KR2 ΔE^{a-ARM}_{S1-S0} value is 56.0 ± 0.1 kcal mol^{-1} (2.42 eV, 511 nm) and differs from the experimental data of just 1.5 kcal mol^{-1} (0.07 eV, −14 nm). However, as discussed below, such accuracy limit does not necessarily impact λ_{max} trends that are not affected by systematic errors. Detailed information of the row data is provided in Supplementary Note 4 (see Supplementary Table 3).

The achieved WT-KR2 model shows that P219 is a suitable position for color tuning. Indeed, as displayed in Fig. 5a, it is close to the β-ionone ring of rPSBAT, so it is expected that the replacement of Pro by residues with different steric hindrance and/or polarity can affect the vertical excitation energy by stabilizing or destabilizing S_1 with respect to S_0, as has been previously observed by some of the authors when modeling the mutants P219G and P219T using the *original* version of the ARM protocol[22]. Although the conclusions derived from the study of Inoue et. al. can be qualitatively compared with the results presented in this work (see Supplementary Fig. 4 and Supplementary Table 4), notice that they cannot be quantitatively compared since (i) the QM/MM models were constructed from a different X-ray structure (i.e., 3X3C), using the *original* ARM[9] that featured (ii) manual input file generation (i.e., handmade and not reproducible counterion placement, different chromophore cavity, etc.) and (iii) a different methodological approach for the generation of the mutant side-chain, as described in Supplementary Note 5.

We begin the analysis of the experimental data by comparing the ΔE^{Exp}_{S1-S0} trend with the corresponding computed values (see Supplementary Table 4), in the hope to learn how the P219X mutants shift their colors with respect to WT-KR2 (see blue down-triangles in Fig. 5c). As shown in Fig. 5b and reported in Table 1, the observed values vary from 545 nm to 515 nm. Remarkably, as reported in Supplementary Table 4, the P219R is the only variant that exhibit a blue-shifted effect with respect to the WT-KR2 of ca. 1.1 kcal mol^{-1}, whereas the other 18 mutants exhibit a red-shifting effect raging from −1.0 to −2.0 kcal mol^{-1} suggesting that an electrostatic destabilization of the chromophore positive charge near its β-ionone is a key factor in the observed color tuning.

It is apparent that the WT-KR2 and P219R define a correlation line which is parallel to the perfect correlation line between observed and computed values, thus indicating a general systematic blue shifted error of ca. 15 nm (see Fig. 5d). On the other hand, the <1.0 kcal mol^{-1} observed ΔE^{Exp}_{S1-S0} variations among the 18 red-shifted mutants are, in most cases, far too small

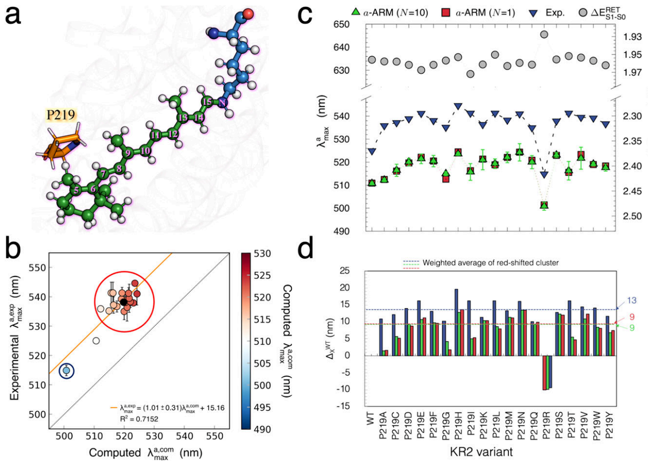


Fig. 5 Computed vs experimental trend in maximum absorption wavelength. (**a**) Scheme of the retinal proton Schiff base of WT- KR2. The residue P219 is located near the β-ionone ring. (**b**) Correlation plot between computed (ΔE^{a-ARM}_{S1-S0}) and measured (ΔE^{Exp}_{S1-S0}) vertical excitation energy for WT-KR2 rhodopsin and 19 of its mutants (P219X, with X = A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W, Y). The red circle indicates the red-shifted mutant cluster whose weighted average value is marked with a black circle. The blue circle indicates the only blue-shifted mutant of the set. (**c**) Maximum absorption wavelength (λ^a_{max}, nm) and (ΔE^{a-ARM}_{S1-S0}, eV) for wild-type KR2 and 19 of its P219X variants. The experimental values (blue down triangles) are presented along with the *a*-ARM values predicted by using the average of *n* = 10 replicas (green triangles), and the representative replica with ΔE^{a-ARM}_{S1-S0} closest to the average (red squares). The values for the retinal chromophore (gray circles) calculated in vacuum (ΔE^{RET}_{S1-S0}), outside the protein environment, are also presented. Row data is provided in Supplementary Table 2. (**d**) Relative observed and computed ΔE^{a-ARM}_{S1-S0} changes ($\Delta_{,X}^{WT}$) with respect to WT. Row data is provided in Supplementary Table 3. The dashed lines show the values corresponding the weighted averages of the red-shifted cluster. S_0 and S_1 energy calculations were performed at the CASPT2(12,12)//CASSCF(12,12)/AMBER level of theory using the 6-31G(d) basis set.

for their trend to be reproduced by a *a*-ARM QM/MM model. These mutants form a cluster whose weighted average (the black circle in Fig. 5b which represents the average of the measured and computed values weighted according to the more frequent deviation from the wild-type value, see Supplementary Note 6 and Supplementary Note 7) aligns with WT-KR2 and P219R supporting the general validity of the constructed models and confirming the presence of a systematic error (the row data reported in Table 1shows a systematic blue-shifted error of ca. 2.0 kcal mol^{-1}). Notice that for these models the mean absolute error (MAE) and the mean absolute deviation (MAD), calculated as indicated in Supplementary Note 6, are 1.9 and 0.3 kcal mol^{-1} respectively, consistently with those established for the *a*-ARM protocol[9,30,35–37]. For this reason, in the present work we avoid a detailed analysis of the members of the red-shifted cluster and focus on their weighted average features. In other words, we primarily focus on the effects that make: i) the P219R and ii) the center of the red-shifted cluster different from WT-KR2 in terms of ΔE^{a-ARM}_{S1-S0}.

The observed ΔE^{Exp}_{S1-S0} and calculated ΔE^{a-ARM}_{S1-S0} values are reported in Fig. 5c, while their difference relative to WT-KR2 are given in Fig. 5d (see also Supplementary Table 5). These quantities were computed as the difference between observed $\Delta E^{Exp}_{S1-S0}$for each P219X mutant with respect to the observed value of the WT-KR2 ($\Delta_{max,X}{}^{WT,Exp}$), as well as the difference between computed ΔE^{a-ARM}_{S1-S0} of each of the P219X mutants with respect to the corresponding WT-KR2 ($\Delta_{max,X}{}^{WT,a\text{-}ARM}$) values. Notice that in this figure the average values (see dashed horizontal lines) for the red-shifted cluster are also provided. In all the cases, blue or red direction of the shift is reproduced. More interestingly, in Fig. 5d it is evident that, in line with the observations (blue bars), the computed data (green and red bars) show that P219R is the only mutant presenting a blue-shifted effect and P219H is the most red-shifted mutant.

In order to quantify the parallelism between the computed and experimental trends in excitation energy (see Supplementary Note 7), we have calculated the trend deviation ($\|\text{Trend Dev.}\| = |\Delta_{max,X}{}^{WT,Exp} - \Delta_{max,X}{}^{WT,a-ARM}|$) as 0.4 ± 0.3 kcal mol^{-1} (0.02 ± 0.01 eV), using the data reported in Supplementary Table 3. Such value is close to the value reported for the *a*-ARM protocol in ref. [9]. (see "Materials and methods" section), further supporting the general validity of our QM/MM models.

The results above indicate that the 20 generated QM/MM models may be employed to investigate the color tuning mechanism operating in KR2. This is done by analyzing the differences between the P219R model and the center of the red-shifted cluster (18 red-shifted models) with respect to the WT-KR2 model. A first question to be answered is: why R is the only residue, out of the three canonical positively charged residue (K, H and R), causing a blue-shift in spite of its not dramatically larger volume?

Color tuning analysis in terms of steric and electrostatic effects. In order to gain insight into the color tuning mechanism inducing ΔE^{a-ARM}_{S1-S0}red- and blue-shifting, we looked at the steric and/or electrostatic effects that modulate the energy of either the S_0 or the S_1 states and, consequently, the excitation energy. As mentioned above, such analyses are primarily focused on the red-shifted (18 mutants) cluster center and P219R mutant (see Fig. 5b). In Fig. 6, we give a visual representation of three fundamental quantities ($\Delta\Delta E^{TOT}{}_{S1-S0}$, $\Delta\Delta E^{STR}{}_{S1-S0}$, and $\Delta\Delta E^{X219\text{-}OFF}{}_{S1-S0}$) whose values are a function of either the structural (both at the chromophore and protein cavity levels) or electrostatic changes of each mutant with respect to WT-KR2. $\Delta\Delta E^{TOT}{}_{S1-S0}$ is

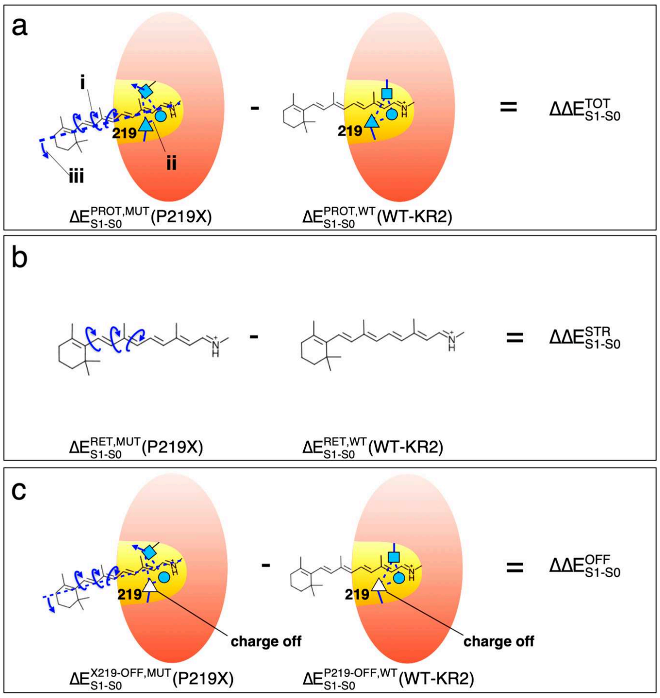
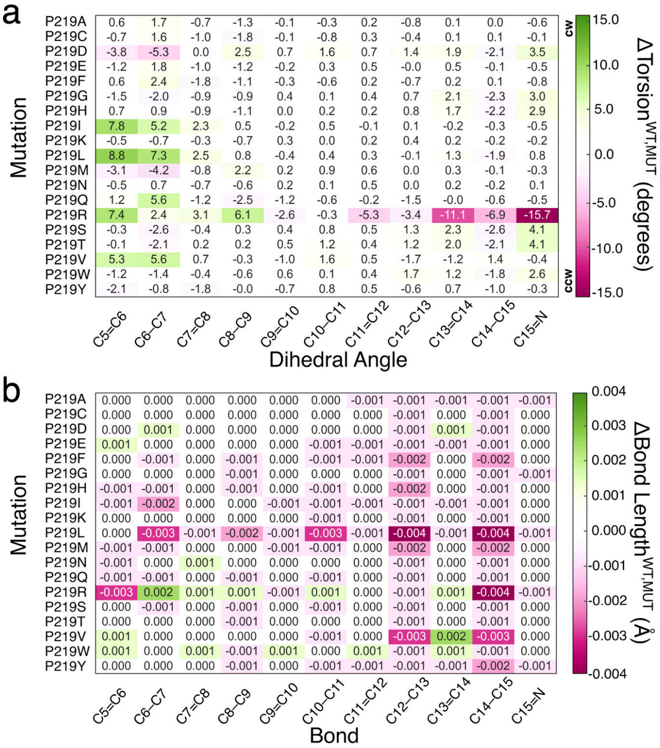


Fig. 6 Pictorial illustration of the corresponding excitation energy analysis terms. (**a**) Difference between the *a*-ARM models excitation energies (in protein environment, $\Delta E^{PROT}{}_{S1\text{-}S0} = \Delta E^{a-ARM}_{S1-S0}$) of a P219X mutant with respect to the wild-type due to (i) structural deformation (see curly arrows) and (ii) structural deformation of the cavity residues (see straight arrow and dotted lines indicating the hydrogen bond network) and (iii) rPSBAT reorientation (see dashed line along the chromophore axis). (**b**) Difference between the excitation energies of the chromophores (in vacuum, $\Delta\Delta E^{STR}{}_{S1\text{-}S0}$) of P219X mutant and wild-type, isolated at their equilibrium structures from the protein environment. (**c**) Difference between the excitation energies (in protein environment) of a P219X mutant with respect to the wild-type (i.e., same as in part (**a**)) but calculated after setting to zero the charges of residue X in the mutant and the charges of residue P in the wild-type. Red ovals and yellow areas represent the protein environment and the chromophore cavity, respectively.

the “Total” excitation energy change (see Fig. 6a). It is directly computed as the difference between the QM/MM computed vertical excitation energies (see red squares in Fig. 5c) of the mutant and WT-KR2 protein models. $\Delta\Delta E^{STR}{}_{S1-S0}$ is the “Steric component” of $\Delta\Delta E^{TOT}{}_{S1-S0}$ (see Fig. 6b). It is directly computed as the difference between the QM/MM vertical excitation energies of the isolated retinal chromophores, taken with their protein environment geometries, (see gray circles in Fig. 5c) of the mutant and WT-KR2. As we will see in the following, using these quantities we can compute three additional components. $\Delta\Delta E^{ELE(t)}{}_{S1-S0}$ is the “Total electrostatic component” that is indirectly computed as the difference between the Total and the Steric components above ($\Delta\Delta E^{ELE(t)}{}_{S1-S0} = \Delta\Delta E^{TOT}{}_{S1-S0} - \Delta\Delta E^{STR}{}_{S1-S0}$) for each mutant. As specified below, $\Delta\Delta E^{ELE(t)}{}_{S1-S0}$ can be decomposed into two parts. $\Delta\Delta E^{ELE(i)}{}_{S1-S0}$ is the “Indirect electrostatic component” that is indirectly computed in two steps by first computing the differences between the vertical excitation energy of the mutant and WT-KR2 obtained after having switched off (turned to zero) the charges of residue 219 (see Fig. 6c), component $\Delta\Delta E^{OFF}{}_{S1-S0}$) and then by subtracting from such difference the steric effect $\Delta\Delta E^{STR}_{S1-S0}$ defined above. Finally, $\Delta\Delta E^{ELE(d)}{}_{S1-S0}$ is the “Direct electrostatic component” that is indirectly computed as $\Delta\Delta E^{ELE(d)}{}_{S1-S0} = \Delta\Delta E^{ELE(t)}{}_{S1-S0} - \Delta\Delta E^{ELE(i)}{}_{S1-S0}$. Further details are provided in Supplementary Note 8 (see Supplementary Table 6 and Supplementary Table 7).

Steric effects. We begin the discussion on steric effects (here by “steric effects” we mean “indirect” or “geometrical” effects, i.e., the change in excitation energy of a chromophore due to a change in the minimum geometry. The change in geometry could be induced by both steric and electrostatic factors) by investigating how the retinal chromophore is structurally modified by the mutations (i) near the β-ionone ring and (ii) near the Schiff base linkage (see Fig. 5a). As discussed below, such structural rearrangements of rPSBAT rather than being a simple effect induced by the side-chain replacement, could be also attributed to a different charge distribution due to changes in protonation states for ionizable residues as well as water addition/removal. Notice that we evaluated steric effect through an “atomistic” approach focused on the changes in rPSBAT geometrical and electronic structure and therefore not directly related to steric effects evaluated on the basis of the changes in residue volume addressed



a

Mutation	C5=C6	C6–C7	C7=C8	C8–C9	C9=C10	C10–C11	C11=C12	C12–C13	C13=C14	C14–C15	C15=N
P219A	0.6	1.7	-0.7	-1.3	-0.1	-0.3	0.2	-0.8	0.1	0.0	-0.6
P219C	-0.7	1.6	-1.0	-1.8	-0.1	-0.8	0.3	-0.4	0.1	0.1	-0.1
P219D	-3.8	-5.3	0.0	2.5	0.7	1.6	0.7	1.4	1.9	-2.1	3.5
P219E	-1.2	1.8	-1.0	-1.2	-0.2	0.3	0.5	-0.0	0.5	-0.1	-0.5
P219F	0.6	2.4	-1.8	-1.1	-0.3	-0.6	0.2	-0.7	0.2	0.1	-0.8
P219G	-1.5	-2.0	-0.9	-0.9	0.4	0.1	0.4	0.7	2.1	-2.3	3.0
P219H	0.7	0.9	-0.9	-1.1	0.0	0.2	0.2	0.8	1.7	-2.2	2.9
P219I	7.8	5.2	2.3	0.5	-0.2	0.5	-0.1	0.1	-0.2	-0.3	-0.5
P219K	-0.5	-0.7	-0.3	-0.7	0.3	0.0	0.2	0.4	0.2	-0.2	-0.2
P219L	8.8	7.3	2.5	0.8	-0.4	0.4	0.3	-0.1	1.3	-1.9	0.8
P219M	-3.1	-4.2	-0.8	2.2	0.2	0.9	0.6	0.0	0.3	-0.1	-0.3
P219N	-0.5	0.7	-0.7	-0.6	0.2	0.1	0.5	0.0	0.2	-0.2	0.1
P219Q	1.2	5.6	-1.2	-2.5	-1.2	-0.6	-0.2	-1.5	-0.0	-0.6	-0.5
P219R	7.4	2.4	3.1	6.1	-2.6	-0.3	-5.3	-3.4	-11.1	-6.9	-15.7
P219S	-0.3	-2.6	-0.4	0.3	0.4	0.8	0.5	1.3	2.3	-2.6	4.1
P219T	-0.1	-2.1	0.2	0.2	0.5	1.2	0.4	1.2	2.0	-2.1	4.1
P219V	5.3	5.6	0.7	-0.3	-1.0	1.6	0.5	-1.7	-1.2	1.4	-0.4
P219W	-1.2	-1.4	-0.4	-0.6	0.6	0.1	0.4	1.7	1.2	-1.8	2.6
P219Y	-2.1	-0.8	-1.8	-0.0	-0.7	0.8	0.5	-0.6	0.7	-1.0	-0.3

b

Mutation	C5=C6	C6–C7	C7=C8	C8–C9	C9=C10	C10–C11	C11=C12	C12–C13	C13=C14	C14–C15	C15=N
P219A	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	-0.001	-0.001	-0.001	-0.001
P219C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000
P219D	0.000	0.001	0.000	0.000	0.000	0.000	0.000	-0.001	0.001	-0.001	0.000
P219E	0.001	0.000	0.000	0.000	0.000	-0.001	-0.001	-0.001	-0.001	-0.001	0.000
P219F	0.000	-0.001	0.000	-0.001	0.000	-0.001	-0.001	-0.002	0.000	-0.002	0.000
P219G	0.000	0.000	0.000	-0.001	0.000	0.000	0.000	-0.001	0.000	-0.001	-0.001
P219H	-0.001	-0.001	0.000	-0.001	0.000	-0.001	0.000	-0.002	0.000	-0.001	0.000
P219I	-0.001	-0.002	0.000	0.000	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	0.000
P219K	0.000	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
P219L	0.000	-0.003	-0.001	-0.002	-0.001	-0.003	-0.001	-0.004	-0.001	-0.004	-0.001
P219M	-0.001	-0.001	0.000	0.000	-0.001	-0.001	0.000	-0.002	0.000	-0.002	0.000
P219N	-0.001	0.000	0.001	0.000	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000
P219Q	-0.001	-0.001	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
P219R	-0.003	0.002	0.001	0.001	-0.001	0.001	0.000	-0.001	0.001	-0.004	-0.001
P219S	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
P219T	0.000	0.000	0.000	-0.001	0.000	0.000	0.000	-0.001	0.000	-0.001	0.000
P219V	0.001	0.000	0.000	0.000	0.000	-0.001	0.000	-0.003	0.002	-0.003	0.000
P219W	0.001	0.000	0.001	-0.001	0.001	0.000	0.001	-0.001	0.001	-0.001	0.000
P219Y	0.000	0.000	0.000	-0.001	0.000	-0.001	-0.001	-0.001	-0.001	-0.002	-0.001

Fig. 7 Heatmap representation of the variation of structural parameters of P219X (X= A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W, Y) mutants with respect to the wild-type KR2 (WT-KR2). Difference between WT-KR2 and mutant (**a**) dihedral angles, $\Delta\text{Torsion}^{WT,MUT}$ and (**b**) bond lengths, $\Delta\text{BondLength}^{WT,MUT}$. Counterclockwise (CCW) and clockwise (CW) orientation of the rPSBAT.

above. To do so, we select a representative QM/MM structure for each mutant model (i.e., the replica with ΔE^{a-ARM}_{S1-S0} closest to the average, see red squares in Fig. 5c) and compute the difference between the magnitude of its most relevant structural parameters (e.g., torsional dihedral angles and bond lengths) with respect to those corresponding to the representative structure of the WT-KR2 model. A visual structural comparation between each mutant and WT-KR2 is provided in Supplementary Figure 5. In Fig. 7 we report a heatmap visualization of such differences in terms of skeletal dihedral angles ($\Delta\text{Torsion}^{WT,MUT}$, Fig. 7a) and carbon-carbon bond lengths ($\Delta\text{BondLength}^{WT,MUT}$, Fig. 7b). Based on the range in which the latest quantities change, we establish an arbitrary threshold of 2.0 degrees and 0.01 Å, respectively, as meaningful variations. The dihedrals with the largest variation are C5 = C6 and C6-C7 that, as shown in Fig. 5a, belong to the rPSBAT framework geometrically closer to the 219 residue. In addition, for specific cases the "reactive" C13 = C14 dihedral as well as the C14-C15 and C15 = N dihedrals show variability. We start the analysis discussing the variants belonging to the red-shifted cluster. These are almost constantly accompanied by structural effect near the β-ionone ring (among them, P219I, P219L, P219M, P219V, P219F, P219Y featuring hydrophobic side-chain, and P219Q featuring large polar uncharged side-chain present the largest structural effects) making such a structural deformation a characteristic of the cluster center. On the other hand, P219G, P219H and P219W show a limited change in the C15 = N region and P219D (in its neutral form), P219S and P219T (i.e., polar uncharged side-chains) present changes in both regions. As expected, the blue-shifted P219R mutant featuring a positively charged side-chain, induces a significant, more than 2.0 degrees, change of its dihedral angles near the β-ionone ring. More interestingly, this mutant exhibits a particular large variation of the "reactive" C13 = C14 dihedral as well as the C14-C15 and C15 = N dihedrals (see Fig. 7a).

Although at a first glance none of the bond lengths are significantly altered with respect to the WT-KR2 reference (see Fig. 7b), it is well-known that the excitation energy of a conjugated chromophore in its protein environment is sensitive to the delocalization of the π-electron and, consequently, to bond length alternation (BLA)[38]. The latter is computed as the difference between the average single bond length and the average double bond length of the π-conjugated chromophore[39,40]. Therefore, we computed the BLA for each of the P219X variants as well as their difference with respect to the WT-KR2. The results, reported in Supplementary Table 9 and Supplementary Note 11, show that the 18 variants of the red-shifted cluster present values of BLA lower than the WT-KR2, indicating that this is a common feature shared by the red-shifted cluster and its center. In contrast, the only blue-shifted variant presents a larger BLA value. Such results are consistent

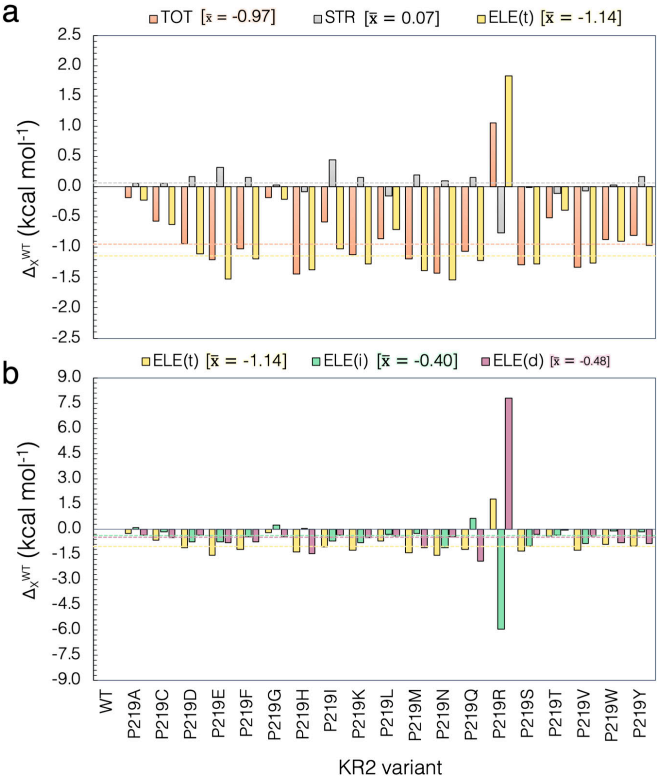


Fig. 8 Steric and electrostatic contributions to the vertical excitation energy. (**a**) Total steric (STR) and electrostatic (ELE(t)) contributions of the interaction of the retinal with the protein environment for the P219X (X = A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W, Y). (**b**) Decomposition of the total ELE(t) electrostatic effects on its indirect (ELE(i)) and direct (ELE(d)) components. Row data is provided in Supplementary Table 5. The dashed lines and corresponding numerical values refer to the weighted average values (**x̄**) of the 18 residues of the red-shifted cluster exclusively, presented in square parenthesis.

with a previous study reported for some of the authors, where it is discussed that more conjugation of the double bonds implies less BLA and red-shift[22].

We now look at the impact of the described geometrical changes on the λ^{a}_{max} value. To this aim, we compare the computed $\Delta E^{RET,MUT}_{S1-S0}$ values with the corresponding wild-type $\Delta E^{RET,WT}_{S1-S0}$ for the isolated chromophore (in kcal mol^{-1}), as reported in Supplementary Table 7. This analysis consists in extracting, for WT-KR2 and for each variant, the rPSBAT structure from the protein and computing its vertical excitation energy without relaxing the chromophore structure (see Fig. 6b). A small $\Delta\Delta E^{STR}_{S1-S0} = \Delta E^{RET,MUT}_{S1-S0} - \Delta E^{RET,WT}_{S1-S0}$ value indicates that the geometrical distortion of the retinal due to the point mutation has only a limited effect on the excitation energy change. As mentioned above, we analyze the $\Delta\Delta E^{STR}_{S1-S0}$ contribution for the red-shifted cluster center and P219R. To this aim, we computed the weighted average of the $\Delta\Delta E^{STR}_{S1-S0}$ contribution for members of the red-shifted cluster (see gray dashed line in Fig. 8a) and compare this value with the $\Delta\Delta E^{STR}_{S1-S0}$ contribution of the blue-shifted one. Consistently with the small variations in $\Delta E^{RET,WT}_{S1-S0}$ reported in Fig. 5c (see gray circles) for the red-shifted cluster, the weighted $\Delta\Delta E^{STR}_{S1-S0}$ average of about 0.07 kcal mol^{-1} reported in Fig. 8a and Supplementary Table 6, show that the impact of the rPSBAT geometrical deformation on the change in their excitation energy is very limited. On the other hand, as also confirmed by the geometrical data discussed above, the blue-shifted P219R variant features a considerable $\Delta\Delta E^{STR}_{S1-S0}$ contribution of −0.8 kcal mol^{-1}. These results, also reported in Supplementary Table 7, give a first indication about the different behavior on the color tuning mechanism exhibited for the red-shifted clusters and P219R mutant.

Electrostatic effects. The QM/MM models also allow to investigate electrostatic effects. As anticipated above, the total electrostatic effect ($\Delta\Delta E^{ELE(t)}{}_{S1-S0}$) can be decomposed in two parts: (i) the first can be considered as a *direct* component ($\Delta\Delta E^{ELE(d)}{}_{S1-S0}$) due to the variation in number, magnitude, and position of the point charges of residue 219 caused by the P to X replacement and (ii) a more *indirect* component ($\Delta\Delta E^{ELE(i)}{}_{S1-S0}$) produced from the reorganization of the local environment and hydrogen bond network induced by the same replacement and due to the fact that conserved residues and water molecules change in position or orientation. Moreover, as discussed below, possible changes in protonation states of conserved residues, induced by P to X replacement, have a major contribution to the *indirect* component.

Figure 8a presents the relative vertical excitation energy for each mutant with respect to WT-KR2 value ($\Delta\Delta E^{TOT}_{S1-S0}$), and its steric ($\Delta\Delta E^{STR}_{S1-S0}$) and total electrostatic ($\Delta\Delta E^{ELE(t)}_{S1-S0}$) contributions (see Supplementary Note 8). Again, it is apparent that steric effects due to the variation in geometry of the retinal chromophore (see gray bars in Fig. 8a), are limited and do not compete with the electrostatic effect, except for the case of the blue-shifted P219R. In general, it can be claimed that, on the basis of the *a*-ARM QM/MM models the electrostatic effect is the one driving the changes in the computed and, therefore, observed λ_{max} values. Accordingly, in each case $\Delta\Delta E^{ELE(t)}_{S1-S0}$ (see gold bars in Fig. 8a) follows the trend of $\Delta\Delta E^{TOT}_{S1-S0}$, being a positive value for the cases that present a red-shifting effect, and a negative value in the only blue-shifted case.

We now discuss the role of the two components of the electrostatic effects that are calculated using the definitions given above (see Fig. 8b and Supplementary Note 8). A comparison of the contribution of both direct and indirect electrostatic components is plotted in Fig. 8b and the values are reported in Supplementary Table 7. Again, we performed the analyses for both $\Delta E^{ELE(d)}_{S1-S0}$ and $\Delta E^{ELE(i)}_{S1-S0}$ components in terms of weighted average value for the red-shifted cluster and value of the blue-shifted P219R. As observed, the direct $\Delta E^{ELE(d)}_{S1-S0}$ component of the red-shifted cluster is of about −0.48 kcal mol^{-1}, while the blue-shifted P219R mutant displays a large value of 7.8 kcal mol^{-1}. The latter value is attributed to the disappearance of a S_1 destabilizing Arg positive charge at position 219. The differences in both sign and magnitude of the $\Delta E^{ELE(d)}_{S1-S0}$ in both clusters, show how they operate in a different fashion. Furthermore, the analysis of the indirect $\Delta E^{ELE(i)}_{S1-S0}$ component shows that the weighted average of the red-shifted cluster of about −0.40 kcal mol^{-1} is only comparable in sign but not in magnitude with the value of the blue-shifted variant, computed as −6.0 kcal mol^{-1}. Such effect is expected since the blue-shifted P219R is the only variant that exhibits a considerable change in the factors defined above as the cause of the indirect electrostatic component. More specifically, the QM/MM model of P219R suggests that when full positive charge is introduced, the general structure gets naturally counterbalance by the deprotonation of the Glu residue in position 160 even if this residue is located ca. 11 Å away from the β-ionone ring (see Fig. 9a). This change, that justifies the large value of the indirect electrostatic component, is also accompanied by a substantial rearrangement of the molecules of water near the chromophore protonated Schiff base group and therefore a variation of the hydrogen bond network. Furthermore, the same model indicates that, in contrast to WT-KR2 and the red-shifted cluster, the P219R model can only reproduce the observed λ_{max} value after changing the protonation states of the rPSBAT counterion complex making D116 neutral and D251 negatively charged (see Fig. 9a). We hypothesize that this reorganization of the protonation states is connected to the observed lack of sodium pumping activity in this mutant. This relatively important alterations appear reasonable when considering, as also stressed above, that a full localized charge is introduced in the KR2 cavity upon mutation. In fact, such large changes, including the changes in protonation states, are found to be not necessary in the mutant models P219K and P219H (see Fig. 9d) as the residues Lys and His are in their deprotonated (i.e., neutral) forms and, thus, the general arrangement of the charges in the cavity is not altered with respect to WT-KR2 (see Fig. 9b) or P219G (see Fig. 9c).

Conclusions

Above we have presented a combined experimental and computational investigation of the P219X mutants of KR2 supporting the existence of a G/P switch in microbial rhodopsins. We have also reported on a computational analysis of the mutant vertical excitation energy indicating that G/P is, essentially, an electrostatic switch capable of a ca. 30 nm modulation (going from R to H). More specifically, the decomposition of the simulated excitation energy changes using multiconfigurational quantum chemistry-based QM/MM models, suggests that the switch operates by inducing variations in the electrostatic interaction of the chromophore with its environment while replacement-induced geometrical distortions only contribute to the P219R blue-shifted variant.

The reported QM/MM modeling studies provide a mechanistic interpretation of the G/P switch color tuning mechanism. In fact, it is found that P219R features an increased planarity of the β-ionone and Schiff base moieties leading to a red-shifting steric (ΔE^{STR}_{S1-S0}) contribution of ca. −0.8 kcal mol^{-1}. Such steric red-shifting effect, which is the result not only of the P/R replacement side-chain but also of a change in the HBN pattern, is counterbalanced by a larger blue-shifting electrostatic contribution ($\Delta E^{ELE(t)}_{S1-S0}$). The decomposition of $\Delta E^{ELE(t)}_{S1-S0}$ into direct ($\Delta E^{ELE(d)}_{S1-S0}$) and indirect ($\Delta E^{ELE(i)}_{S1-S0}$) components in P219R, reveals that: (i) $\Delta E^{ELE(d)}_{S1-S0}$ is significantly larger than in the other variants, inducing a large blue-shifting effect (ca. 7.8 kcal mol^{-1}); (ii) $\Delta E^{ELE(i)}_{S1-S0}$ is almost of the same magnitude of $\Delta E^{ELE(d)}_{S1-S0}$ (ca. −6.0 kcal mol^{-1}) but induces a red-shifting effect. These large and contrasting effects including structure, counterion and HBN pattern reorganization, are justified by the introduction of a positively charged arginine side-chain necessary to reproduce the observed P219R blue-shifting. No change in structure and counterion and only limited changes in HBN are instead seen in the members of the red-shifted cluster (i.e., when we represent it with a weighted average) or in the specific P219G and P219H variants representing the cluster minimum and maximum values. Indeed, all red-shifted variants conserve the main features of WT-KR2 and display a color tuning mechanism mainly controlled by the direct electrostatic changes associated with the replaced side-chain. These conclusions agree with those reported in Inoue et. al. the red-shifting mechanism of P219G and P219T was investigated.

The fact that in the investigated set of KR2 mutants both direct (the change in the electrostatic field due to the residue replacement) and indirect (the changes due to all other cavity reorganization induced by the replacement and including chromophore reorientation, side-chain and water relocations and the modification of the hydrogen bond network) electrostatic effects and specific changes in the chromophore structure contribute to determine the color variability, cast doubts on the possibility to extract simple rules for predicting how a switch operates without understanding the molecular-level details of the side chain replacement. In other words, the analysis indicates that the color tuning mechanism seen in KR2 has a complex origin.

The possibility to carry out, for the first-time, a systematic modeling of mutants (i.e., comprising all 19 possible replacements) underscores the importance of automated (or semi-automated) computational tools for the fast building of congruous (i.e. comparable) QM/MM models thus allowing comparative (i.e. trend) studies. More specifically, the employed *a*-ARM building protocol avoid errors and biases likely to impact the congruity of a generated QM/MM model set and, therefore, allows well defined comparative studies and mechanistic

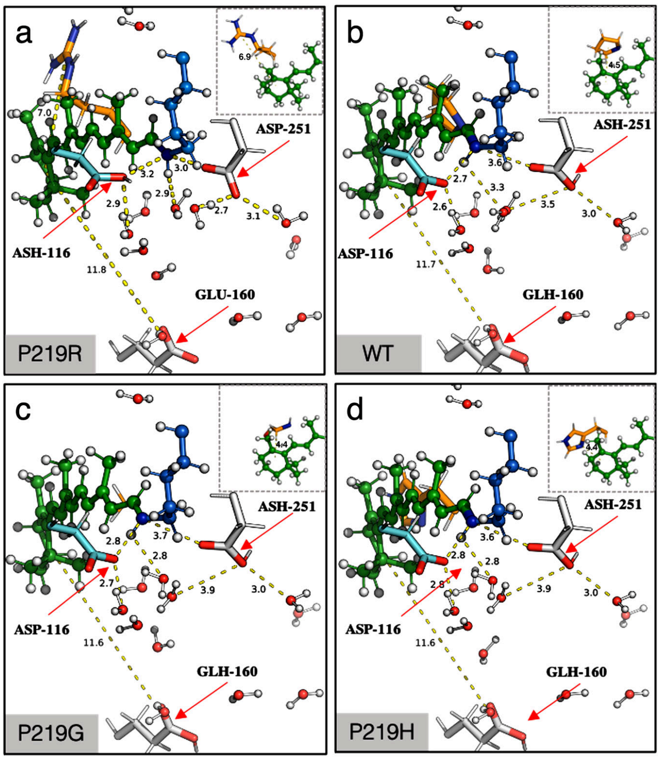


Fig. 9 *a*-ARM QM/MM-optimized models for P219R, WT-KR2, P219G and P219H. Differences between Hydrogen Bond Network (HBN) presented for (**a**) the only blue-shifted variant (P219R), (**b**) the wild-type (WT-KR2), (**c**) the least red-shifted variant (P219G) and (**d**) the most red-shifted variant (P219H). Hydrogen bonds are represented as dashed lines. The red arrows indicate the protonation state of the relevant residues.

interpretations within the limits imposed by the model structure. Future work will be devoted to the study of other potential color switches present in the microbial rhodopsins with the hope to support the conclusions of the present work or document distinct types of color tuning mechanisms.

Materials and methods

Mutagenesis and protein expression. The synthesized genes of KR2 (GenBank Accession number: BAN14808.1) codon-optimized for *E. coli* were incorporated into the pET21a(+) vector (Novagen, Merck KGaA, Germany). The site-directed mutation was conducted using a QuikChange site-directed mutagenesis kit (Agilent, CA). The plasmids carrying the genes of the wild type of KR2 (WT-KR2), and mutants were transformed into the *E. coli* C43(DE3) strain (Lucigen, WI). The protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) in the presence of 10 µM all-*trans* retinal for 4 h. See Supplementary Note 9 and Supplementary Table 8.

Measurement of λ_{max} by hydroxylamine bleach. The λ_{max} of WT-KR2 and mutants was determined by bleaching the protein with hydroxylamine according to the previously reported method[22]. The *E. coli* cells expressing the rhodopsins were washed with a solution containing 100 mM NaCl, 50 mM Na_2HPO_4 (pH 7) for three times. The washed cells were treated with 1 mM lysozyme for 1 h at room temperature, and then disrupted by sonication. To solubilize the rhodopsins, 3% DDM was added, and the samples were stirred for overnight at 4 °C. The rhodopsins were bleached with 500 mM hydroxylamine and subjected to illumination of yellow light ($\lambda > 500$ nm) from the output of 1 kW tungsten−halogen projector lamp (Master HILUX-HR, Rikagaku) through a glass filter (Y-52, AGC Techno Glass, Japan). The absorption change upon the bleaching was measured by UV-visible spectrometer (V-730, JASCO, Japan).

Ion-transport assay. The ion transport activity assay in *E. coli* cells was conducted according to the previously reported method[41]. The *E. coli* cells carrying the expressed rhodopsin were washed for three times and resuspended in unbuffered 100 mM NaCl or CsCl to assay the Na^+ or H^+ pump activity, respectively. The sample was illuminated after adjusting the pH to ~7 by the addition of a small amount of HCl or NaOH. The pH change upon light illumination was monitored with a pH electrode (9618S-10D, HORIBA, Japan). The wavelength of the illuminating light was changed by placing different color filters (Y-52, Y-54, O-55 and O-56, AGC Techno Glass, Japan) with a heat absorbing filter (HAF-50S-50H, SIGMAKOKI, Japan) in front of the light source (1 kW tungsten−halogen projector lamp, Master HILUX-HR, Rikagaku) for KR2 WT ($\lambda > 500$ nm), and P219X ($\lambda > 520$ nm) to correct for the change in the degrees of light absorption by their spectral shift.

QM/MM modeling. Congruous and reproducible computational models for WT-KR2 and its P219X (X = A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W, Y) mutants were generated using the *a*-ARM version[9] of the Automatic Rhodopsin Modeling protocol[34], a specialized user-friendly command-line oriented computational tool (i.e., Python3-based software package) for either the fully-automated (default) or semi-automated (i.e., customized) construction of basic hybrid QM/MM models of rhodopsins[30,35–37,42], called *a*-ARM models. The produced *a*-ARM models illustrated in Fig. 4a are monomeric, "gas-phase", and globally uncharged, based on electrostatic embedding and the hydrogen link-atom frontier between the QM and MM subsystems. The term "gas-phase" refers to the fact that the protein membrane and solvation effects are not explicitly modeled during the calculations. Nevertheless, the electrostatic effect of the protein external environment is effectively accounted for by an asymmetric distribution of external ions (i.e., Cl^- or Na^+) whose charge and position are determined, automatically, by a specific algorithm, near the most positively and/or negatively charged surface amino acids in both the intracellular (IS) and extracellular (OS) protein surfaces. Moreover, to account the water-mediated hydrogen-bond network (HBN) in the protein cavity, the internal crystallographic waters are retained in the model, while the waters that are not experimentally detected are assumed to be extremely mobile or just absent in the chromophore hydrophobic protein cavity. Considering these approximations, during the construction of the *a*-ARM model the rhodopsin structure is subdivided into three sub-systems requiring layers of increasing level of accuracy for their description. The first MM-based sub-system, called protein environment, features residues (backbone and side-chain atoms) fixed at the crystallographic or comparative (homology) structure and incorporates external Cl^- and/or Na^+ counterions fixed at preoptimized positions as well as some crystallographic/comparative waters (see below). The second sub-system, namely the chromophore cavity, constitutes the second MM layer of the model and contains amino acid residues with fixed backbone and relaxed side-chains, including the covalently linked lysine residue atoms (excluding the terminal NH_2-C_ϵ moiety), as well as flexible crystallographic/comparative waters in the vicinity (<4 Å) of the rPSB. The third sub-system, called Lys-QM, contains the MM atoms of the covalently linked lysine side-chain in contact (through C_ϵ) with the QM/MM frontier (i.e., 9 atoms) and the entire QM sub-system which corresponds to a N-methylated retinal chromophore (i.e., 53 atoms). As illustrated in Fig. 4b, c, to obtain the QM/MM model automatically the only required input is either the PDB ID or a template PDB file (X-ray structure or comparative model) of the rhodopsin, that is processed by two subsequent phases called input file generator (Phase I, Fig. 4b) and QM/MM model generator (Phase II, Fig. 4c) correspondingly, previously reported in refs. [9,30,34,35]. and summarized in Supplementary Note 2 and Supplementary Note 3, respectively. Briefly, Phase II produces 10 model replicas of the *a*-ARM model via S_0 geometry optimization of the input generated through Phase I starting from distinct molecular dynamics (MD) runs (see Supplementary Note 3). Such optimization relies on the use of the multi-configurational complete active space self-consistent field (CASSCF) at the 2-roots single-state, for modeling the QM sub-system while the MM subsystem is treated using the AMBER force field (CASSCF(12,12)/6-31 G(d)/AMBER level). The AMBER force field is not polarizable and only includes polarizability effect in a mean-field fashion possibly contributing the method systematic error.

The CASSCF(12,12) level of theory is followed by an energy correction at the multi-configurational second-order perturbation level (CASPT2) to recover the missing dynamical electron correlation. Thus, a 3-roots state-average CASPT2 that uses the three-root stage-average CASSCF(12,12)/6-31 G(d)/AMBER as the zero-order reference wavefunction, is computed (CASTP2(12,12)/6-31 G(d)/AMBER). Ultimately, each of the 10 model replicas correspond to an equilibrated gas-phase and globally uncharged monomer QM/MM model and it is associated with a ΔE_{S1-S0} calculated between $S_0 \rightarrow S_1$. Finally, the average ΔE_{S1-S0} is reported along with the corresponding standard deviation (see Section Supplementary Note 3).

Validation, capabilities, and potential applications of *a*-ARM. The *a*-ARM protocol, in its current version, does not represent a predictive tool but rather is designed to produce models useful for investigating the origin of trends in spectroscopic/photochemical properties (e.g., between sequence variability and function) emerging from different sets of experimental data. Furthermore, the models are used assuming that single amino acid replacements will change the protein conformation at the level of the chromophore cavity. Accordingly, after demonstrating that the models reproduce, within the protocol error bar (see below), the trend in observed vertical excitation energy ΔE_{S1-S0} (equivalent to the $\lambda^a{}_{max}$ value), the QM/MM models are employed to study the factors (i.e., steric and electrostatic contributions to ΔE_{S1-S0}) determining the ΔE_{S1-S0} change of each mutant with respect to the WT-KR2. This is only possible because the level of protocol automation guarantees, regardless of the user ability or computational facility, reproducible *a*-ARM QM/MM models when the same parameters are employed in Phases I and II of Fig. 4b, c. Indeed, the production of the default *a*-ARM model (see *a*-$ARM_{default}$ approach in Supplementary Note 2) is fully automated and does not require human intervention thus preventing the generation of biased models. Furthermore, the absence of parameters required at the input stage, prevents overfitting. On the other hand, in specific cases, mostly related to the difficulties in predicting the correct ionization state of ionizable groups surrounding the chromophore it is necessary to customize the default model by following a well-defined protocol (see *a*-$ARM_{customized}$ approach in Supplementary Note 2). Such a systematic (non-arbitrary) protocol, documented in refs. [9,30]. and summarized in Supplementary Note 2, involves three steps which only concern the ionization states of the ionizable residues that, as reported in refs. [9,30]. and anticipated above, is the most frequent cause of inconsistencies between the computed and observed excitation energies. The customization procedure does not represent an overfitting of the model but a systematic ionization state scan also used by other groups working on rhodopsin simulations (see for instance refs. [13,32,33]).

The *a*-ARM models described above (see Supplementary Note 10 and Supplementary Fig. 6) have been successfully benchmarked by reproducing the trends in ΔE_{S1-S0} of a set of 44 rhodopsins (25 of wild type and 19 mutants)[9,30,35–37,42]. More specifically, the *a*-$ARM_{default}$ approach proved to be capable of reproducing the observed ΔE_{S1-S0} values for 79% (35/44) of the models, whereas the other 21% were successfully obtained with the *a*-$ARM_{customized}$ approach (i.e., customizing the protonation states pattern). The estimated trend deviation, relative to the probed set using as a reference bovine rhodopsin (Rh), has been reported as 0.7 ± 0.5 kcal mol^{-1} (0.03 ± 0.02 eV) and the mean absolute error (MAE) (see definition in Supplementary Note 6) as 1.0 kcal mol^{-1} (0.04 eV)[9]. Both computed and observed trends in ΔE_{S1-S0} for the benchmark set, that comprises microbial and animal rhodopsins whose structures were obtained from either X-ray crystallography or comparative modeling, is presented in Supplementary Figure 7.

Considering that in this study we are interested in trends in vertical excitation energies, and more specifically red- or blue-shifts of these values, the results are presented and discussed in terms of mutation-induced shifts with respect to WT-KR2 rather than differences between observed and calculated values. On the other hand, the observed ΔE_{S1-S0} displays a remarkable experimental "asymmetry" with 18 over 19 single site variants featuring a red-shifted ΔE_{S1-S0} values compressed in a 1.0 kcal mol^{-1} (0.04 eV or ca. 10 nm) wide range. Such relatively small difference has prevented a robust individual quantitative analysis of the color variation of these mutants. For these reasons the red-shifted mutants have been collected in a single cluster and the analysis is mainly performed by focusing on the origin of the ΔE_{S1-S0} change when going from WT-KR2 to the center of the cluster. Still, the less and most red-shifted mutants falling parallel to the correlation line (P219G and P219H respectively) are individually discussed. The limitations and pitfalls of the protocol are summarized in Supplementary Note 12.

Automation of the mutant rotamer selection. The QM/MM modeling of the mutants required a second step of experimentally-driven customization dealing with the residue side-chain conformation (i.e., rotamer selection). That procedure consisted in evaluating the model performance for a set of automatically chosen rotamers and select the one that better reproduces the observed ΔE_{S1-S0} value (we assume $\Delta E^{Exp}_{S1-S0} = hc/\lambda^{a,Exp}_{max}$ and use ΔE^{Exp}_{S1-S0} and $\lambda^{a,Exp}_{max}$ interchangeably in the following). To this aim, an update version of the mutation routine of the input generator (see step 3 in Fig. 4b), that uses Modeller[43] instead of the default rotamer library SCWRL4[44], is proposed as part of the present work. A detailed description of the new approach is provided in Supplementary Note 13. Briefly, in order to explore the performance of different rotamers of the mutated side-chain, three *a*-ARM QM/MM models featuring the three highest scored mutated side-chain rotamers selected by Modeller[43], are produced and their ΔE_{S1-S0} evaluated (see Supplementary Fig. 7). Then, the model that better reproduces the observed ΔE_{S1-S0} is selected (see Supplementary Fig. 8). To perform such selection, we use, as a baseline, the difference between the computed and observed ΔE_{S1-S0} of the WT-KR2, hereafter referred to as $\Delta\Delta E^{Exp,WT}{}_{S1-S0}$. The equivalent quantity calculated for each rotamer ($\Delta\Delta E^{Exp,rotX}{}_{S1-S0}$, with X=1, 2, 3) is then contrasted with the $\Delta\Delta E^{Exp,WT}{}_{S1-S0}$ via the equation $rotX = (\Delta\Delta E^{Exp,rotX}{}_{S1-S0} - \Delta\Delta E^{Exp,WT}{}_{S1-S0})$ (see Supplementary Fig. 9). The rotamer that features the lower rotX value (preferring blue-shifted values) is chosen as the representative *a*-ARM model (see Supplementary Fig. 10). Although this approach relies on experimental information and does not represent a predictive tool, it automates the side-chain conformation selection during the construction of mutant models aimed at the reproduction of experimental trends in properties. Some limitations and pitfalls of the proposed approach are provided in Supplementary Note 13.

Statistics and reproducibility. As further described in "QM/MM Modeling" section and Supplementary Note 3, each *a*-ARM QM/MM model is composed of $n = 10$ independent replicas that correspond to an equilibrated gas-phase and globally uncharged monomer, and it is associated with a ΔE_{S1-S0} calculated between $S_0 \rightarrow S_1$. As shown in Fig. 5c, for each model the average ΔE_{S1-S0} is reported (see green triangles) along with the corresponding standard deviation (see green error bars). Finally, the replica with ΔE_{S1-S0} closest to the average (see red squares) is selected as a representative model for the color tuning analyses. The production of the models, as well as the computation of the average and standard deviation, and the posterior selection of the representative replica is performed automatically by the *a*-ARM protocol illustrated in Fig. 4. Therefore, the reproducibility of the data is guaranteed when selecting the same initial seeds for the $n = 10$ MD runs. On the other hand, the statistical parameters employed for the different computational analyses, *i.e.*, mean absolute error (MAE), mean absolute deviation (MAD), and weighted average ($\bar{x}$) are described in Supplementary Note 6 and the equations are provided. Furthermore, the methodology employed for the calculation of the weighted average is provided in Supplementary Note 6, along with an illustrative example.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All experimental data shown in main figures were deposited in Supplementary Data 1. XYZ Coordinates for the optimized S_0 *a*-ARM QM/MM models are provided as Supplementary Data 2-21 as follows: P219A (Supplementary Data 2.xyz), P219C (Supplementary Data 3.xyz), P219D (Supplementary Data 4.xyz), P219E (Supplementary Data 5.xyz), P219F (Supplementary Data 6.xyz), P219G (Supplementary Data 7.xyz), P219H (Supplementary Data 8.xyz), P219I (Supplementary Data 9.xyz), P219K (Supplementary Data 10.xyz), P219L (Supplementary Data 11.xyz), P219M (Supplementary Data 12.xyz), P219N (Supplementary Data 13.xyz), P219Q

(Supplementary Data 14.xyz), P219R (Supplementary Data 15.xyz), P219S (Supplementary Data 16.xyz), P219T (Supplementary Data 17.xyz), P219V (Supplementary Data 18.xyz), P219W (Supplementary Data 19.xyz), P219Y (Supplementary Data 20.xyz), and WT-KR2 (Supplementary Data 21.xyz). Any remaining information can be obtained from the corresponding author upon reasonable request.

Code availability

Code for the *a*-ARM protocol used in the computational experiments for the QM/MM modeling and color tuning analysis of P219X mutants of KR2 is currently only installed in our cluster facilities but can be distributed via GitLab repository under user request. The manuscript concerning the publication of the computational package that implements all the *a*-ARM protocol/features is work in progress.

Received: 3 January 2021; Accepted: 19 September 2021;
Published online: 13 October 2021

References

1. Ernst, O. P. et al. Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. *Chem. Rev.* **114**, 126–163 (2013).
2. Devine, E. L., Oprian, D. D. & Theobald, D. L. Relocating the active-site lysine in rhodopsin and implications for evolution of retinylidene proteins. *Proc. Natl. Acad. Sci. USA.* **110**, 13351–13355 (2013).
3. Shimono, K., Ikeura, Y., Sudo, Y., Iwamoto, M. & Kamo, N. Environment around the chromophore in pharaonis phoborhodopsin: Mutation analysis of the retinal binding site. *Biochim. Biophys. Acta* **1515**, 92–100 (2001).
4. Yokoyama, S. Evolution of dim-light and color vision pigments. *Annu. Rev. Genomics Hum. Genet.* **9**, 259–282 (2008).
5. Sudo, Y. E. A. A blue-shifted light-driven proton pump for neural silencing. *J. Biol. Chem.* **288**, 20624–20632 (2013).
6. Engqvist, M. K. E. A. Directed evolution of Gloeobacter violaceus rhodopsin spectral properties. *J. Mol. Biol.* **427**, 205–220 (2015).
7. Hayashi, S. E. A. Structural determinants of spectral tuning in retinal proteins-bacteriorhodopsin vs sensory rhodopsin II. *J. Phys. Chem. B* **105**, 10124–10131 (2001).
8. Fujimoto, K., Hayashi, S., Hasegawa, J. Y. & Nakatsuji, H. Theoretical studies on the color-tuning mechanism in retinal proteins. *J. Chem. Theory Comput.* **3**, 605–618 (2007).
9. Pedraza-González, L., De Vico, L., Marín, M. D. C., Fanelli, F. & Olivucci, M. *a*-ARM: automatic rhodopsin modeling with chromophore cavity generation, ionization state selection, and external counterion placement. *J. Chem. Theory Comput.* **15**, 3134–3152 (2019).
10. Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M. & DeLong, E. F. Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786–789 (2001).
11. Katayama, K. & Sekharan, S. S. Y. Optogenetics (eds. Yawo, H., Kandori, H. & Koizumi, A.) vol. 7, 89–107 (Springer, 2015).
12. Inoue, K., Tsukamoto, T. & Sudo, Y. Molecular and evolutionary aspects of microbial sensory rhodopsins. *Biochim. Biophys. Acta* **1837**, 562–577 (2013).
13. Broser, M. et al. Neor, a near-infrared absorbing rhodopsin. *Nat. Commun.* **11**, 5682 (2020).
14. 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).
15. Kato, H. E. et al. Atomistic design of microbial opsin-based blue-shifted optogenetics tools. *Nat. Commun.* **6**, 1–10 (2015).
16. Lin, J. Y., Knutsen, P. M., Muller, A., Kleinfeld, D. & Tsien, R. Y. Reachr: a red-shifted variant of channelrhodopsin enables deep transcranial optogenetic excitation. *Nat. Neurosci.* **16**, 1499–1508 (2013).
17. Oda, K. E. A. Crystal structure of the red light-activated channelrhodopsin Chrimson. *Nat. Commun.* **9**, 3949 (2018).
18. Man, D. E. A. Diversification and spectral tuning in marine proteorhodopsins. *EMBO J.* **22**, 1725–1731 (2003).
19. Bielawski, J. P., Dunn, K. A., Sabehi, G. & Beja, O. Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. *Proc. Natl. Acad. Sci. USA.* **101**, 14824–14829 (2004).
20. Ozaki, Y., Kawashima, T., Abe-Yoshizumi, R. & Kandori, H. A color-determining amino acid residue of proteorhodopsin. *Biochemistry* **53**, 6032–6040 (2014).
21. Sudo, Y. E. A. A microbial rhodopsin with a unique retinal composition shows both sensory rhodopsin II and bacteriorhodopsin-like properties. *J. Biol. Chem.* **286**, 5967–5976 (2011).
22. Inoue, K. et al. Red-shifting mutation of light-driven sodium-pump rhodopsin. *Nat. Commun.* **10**, 1993 (2019).
23. Melaccio, F., Ferré, N. & Olivucci, M. Quantum chemical modeling of rhodopsin mutants displaying switchable colors. *Phys. Chem. Chem. Phys.* **14**, 12485–12495 (2012).
24. Huntress, M. E. A. Towards an understanding of the retinal chromophore in rhodopsin mimics. *J. Phys. Chem. B* **117**, 10053–10070 (2013).
25. Orozco-Gonzalez, Y., Kabir, M. P. & Gozem, S. Electrostatic spectral tuning maps for biological chromophores. *J. Phys. Chem. B* **129**, 4813–4824 (2019).
26. Tang, K., Lin, D. L. K. & Jiao, N. Draft genome sequence of *Parvularcula oceani* JLT2013T, a rhodopsin-containing bacterium isolated from deep-sea water of the southeastern Pacific. *Mar. Genomics* **24**, 211–213 (2015).
27. Choi, D. H. Y. H., Chun, J. & Cho, B. C. *Jannaschia seosinensis* sp. Nov., isolated from hypersaline water of a solar saltern in Korea. *Int J. Syst. Evol. Microbiol.* **56**, 45–49 (2006).
28. Inoue, K. E. A. A natural light-driven inward proton pump. *Nat. Commun.* **7**, 13415 (2016).
29. Kovalev, K. E. A. Structure and mechanisms of sodium-pumping KR2 rhodopsin. *Sci. Adv.* **5**, eaav2671 (2019).
30. Pedraza-González, L. et al. On the automatic construction of QM/MM models for biological photoreceptors: rhodopsins as model systems, In *QM/MM Studies of Light-responsive Biological Systems.* vol 31 (Springer, Cham, 2021).
31. Skopintsev, P. E. A. et al. Femtosecond-to-millisecond structural changes in a light-driven sodium pump. *Nature* **583**, 314–318 (2020).
32. Adam, S., Wiebeler, C. & Schapiro, I. Structural factors determining the absorption spectrum of channelrhodopsins: A Case study of the chimera C1C2. *J. Chem. Theory Comput.* (2021).
33. Kaufmann, J. C. et al. Modulation of light energy transfer from chromophore to protein in the channelrhodopsin ReaChR. *Biophys. J.* **119**, 705–716 (2020).
34. Melaccio, F. et al. Toward automatic rhodopsin modeling as a tool for high-throughput computational photobiology. *J. Chem. Theory Comput.* **12**, 6020–6034 (2016).
35. Aquilante, F. et al. Modern quantum chemistry with [Open]Molcas. *J. Chem. Phys.* **152**, 214117 (2020).
36. Pedraza-González, L. et al. Web-ARM: a web-based interface for the automatic construction of QM/MM models of rhodopsins. *J. Chem. Inf. Model.* **60**, 1481–1493 (2020).
37. Mroginski, M. A . et al. Frontiers in multiscale modeling of photoreceptor proteins. *Photochem. Photobiol.* **97**, 243–269 (2021).
38. Mao, J. et al. Structural basis of the green-blue color switching in proteorhodopsin as determined by NMR spectroscopy. *J. Am. Chem. Soc.* **136**, 17578–17590 (2014).
39. Gieseking, R. L., Risko, C. & Bredas, J. L. Distinguishing the effects of bond-length alternation versus bond-order alternation on the nonlinear optical properties of π-conjugated chromophores. *J. Phys. Chem. Lett.* **6**, 2158–2162 (2015).
40. Ho Choi, C., Kertesz, M. & Karpfen, A. The effects of electron correlation on the degree of bond alternation and electronic structure of oligomers of polyacetylene. *J. Chem. Phys.* **107**, 6712–6721 (1997).
41. Inoue, K. E. A. A light-driven sodium ion pump in marine bacteria. *Nat. Commun.* **4**, 1678 (2013).
42. Gholami, S. et al. Multistatemulti- configuration quantum chemical computation of the two-photon absorption spectra of bovine rhodopsin. *J. Phys. Chem. Lett.* **11**, 6293–6300 (2019).
43. Web, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **54**, 5–6 (2016).
44. Krivov, G. G., Shapovalov, M. V. & Dunbrack, R. L. J. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins Struct. Funct. Bioinformatics* **77**, 778–795 (2009).

Acknowledgements

Dr. M.O. is grateful for partial support provided by grants NSF CHE-CLP-1710191, NIH 1R15GM126627 01, Banca D'Italia and MIUR, Dipartimento di Eccellenza, 2017–2022 (L.P-G.). We also thank the financial supports by JSPS KAKENHI grants 25104009, 15H02391, 18H03986 (H.K.), and 17H03007 (K.I.), and by JST PRESTO (JPMJPR15P2) and CREST (JPMJCR1753 and JPMJCR17N5).

Author contributions

Y.N. constructed the P219X mutants of KR2, performed the expression of the protein with *E. coli*, purified the protein, and performed measurement of $\lambda_{max}^{a,Exp}$ of the WT-KR2 and mutant KR2. M.O. and L.P-G. designed the computational experiments for the QM/MM modeling and color tuning analysis of P219X mutants of KR2. L.P-G. performed the methodological development and code implementation of the employed QM/MM protocols, executed the QM/MM calculations, and produced the computational data. L.P-G. and L.B. designed and implemented the routine for the side-chain mutant generator. H.K., K.I., M.O. and L.P-G. wrote the manuscript. All authors contributed to critical editing of 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/s42003-021-02684-z.

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

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Krishnananda Chattopadhyay and Anam Akhtar. Peer reviewer reports are available.

Bibliography

1. Deisseroth, K. Optogenetics. *Nat Methods* **8**, 26–29 (2011).
2. Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K. Optogenetics in neural systems. *Neuron* **71**, 9–34 (2011).
3. Salinas, F., Rojas, V., Delgado, V., Agosin, E. & Larrondo, L. F. Optogenetic switches for light-controlled gene expression in yeast. *Appl Microbiol Biotechnol* **101**, 2629–2640 (2017).
4. Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K. Optogenetics in neural systems. *Neuron* **71**, 9–34 (2011).
5. Packer, A. M., Roska, B. & Häusser, M. Targeting neurons and photons for optogenetics. *Nat Neurosci* **16**, 805–815 (2013).
6. Crick, F. H. C. Thinking about the brain. *Sci Am* **241**, 219–233 (1979).
7. Lozier, R. H., Bogomolni, R. A. & Stoeckenius, W. Bacteriorhodopsin: a light-driven proton pump in Halobacterium Halobium. *Biophys J* **15**, 955 (1975).
8. Zhang, F., Wang, L.-P., Boyden, E. S. & Deisseroth, K. Channelrhodopsin-2 and optical control of excitable cells. *Nat Methods* **3**, 785–792 (2006).
9. Holland, E. M., Harz, H., Uhl, R. & Hegemann, P. Control of phobic behavioral responses by rhodopsin-induced photocurrents in Chlamydomonas. *Biophys J* **73**, 1395–1401 (1997).
10. Li, X. *et al.* Fast noninvasive activation and inhibition of neural and network activity by vertebrate rhodopsin and green algae channelrhodopsin. *Proceedings of the National Academy of Sciences* **102**, 17816–17821 (2005).
11. 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).
12. Deisseroth, K. Optogenetics: 10 years of microbial opsins in neuroscience. *Nat Neurosci* **18**, 1213–1225 (2015).
13. Adam, Y. All-optical electrophysiology in behaving animals. *J Neurosci Methods* **353**, 109101 (2021).
14. Rost, B. R., Wietek, J., Yizhar, O. & Schmitz, D. Optogenetics at the presynapse. *Nat Neurosci* **25**, 984–998 (2022).
15. Miesenböck, G. The optogenetic catechism. *Science (1979)* **326**, 395–399 (2009).
16. Xu, Y., Zou, P. & Cohen, A. E. Voltage imaging with genetically encoded indicators. *Curr Opin Chem Biol* **39**, 1–10 (2017).
17. Storace, D. *et al.* Toward better genetically encoded sensors of membrane potential. *Trends Neurosci* **39**, 277–289 (2016).
18. Xu, Y., Zou, P. & Cohen, A. E. Voltage imaging with genetically encoded indicators. *Curr Opin Chem Biol* **39**, 1–10 (2017).
19. Lou, S. *et al.* Genetically targeted all-optical electrophysiology with a transgenic cre-dependent optopatch mouse. *Journal of Neuroscience* **36**, 11059–11073 (2016).
20. Hochbaum, D. R. *et al.* All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat Methods* **11**, 825–833 (2014).
21. Broussard, G. J., Liang, R. & Tian, L. Monitoring activity in neural circuits with genetically encoded indicators. *Front Mol Neurosci* **7**, 97 (2014).
22. Rodriguez, E. A. *et al.* The growing and glowing toolbox of fluorescent and photoactive proteins. *Trends Biochem Sci* **42**, 111–129 (2017).

23. Flytzanis, N. C. *et al.* Archaerhodopsin variants with enhanced voltage-sensitive fluorescence in mammalian and Caenorhabditis elegans neurons. *Nat Commun* **5**, 1–9 (2014).
24. 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).
25. Kouyama, T. & Murakami, M. Structural divergence and functional versatility of the rhodopsin superfamily. *Photochemical & Photobiological Sciences* **9**, 1458–1465 (2010).
26. Govorunova, E. G., Sineshchekov, O. A., Li, H. & Spudich, J. L. Microbial rhodopsins: diversity, mechanisms, and optogenetic applications. *Annu Rev Biochem* **86**, 845 (2017).
27. Sharma, A. K., Spudich, J. L. & Doolittle, W. F. Microbial rhodopsins: functional versatility and genetic mobility. *Trends Microbiol* **14**, 463–469 (2006).
28. Spudich, J. L. The multitalented microbial sensory rhodopsins. *Trends Microbiol* **14**, 480–487 (2006).
29. Palczewski, K. G protein–coupled receptor rhodopsin. *Annu. Rev. Biochem.* **75**, 743–767 (2006).
30. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res* **28**, 235–242 (2000).
31. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol* **18**, 581–586 (2008).
32. Bada Juarez, J. F. *et al.* Structures of the archaerhodopsin-3 transporter reveal that disordering of internal water networks underpins receptor sensitization. *Nat Commun* **12**, 1–10 (2021).
33. Pedraza-González, L., de Vico, L., del Carmen Marín, M., Fanelli, F. & Olivucci, M. a-ARM: automatic rhodopsin modeling with chromophore cavity generation, ionization state selection, and external counterion placement. *J Chem Theory Comput* **15**, 3134–3152 (2019).
34. Melaccio, F. *et al.* Toward automatic rhodopsin modeling as a tool for high-throughput computational photobiology. *J Chem Theory Comput* **12**, 6020–6034 (2016).
35. Kandori, H. Ion-pumping microbial rhodopsins. *Front Mol Biosci* **2**, 52 (2015).
36. Subramaniam, S. & Henderson, R. Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. *Nature* **406**, 653–657 (2000).
37. Sharma, A. K., Spudich, J. L. & Doolittle, W. F. Microbial rhodopsins: functional versatility and genetic mobility. *Trends Microbiol* **14**, 463–469 (2006).
38. Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proceedings of the National Academy of Sciences* **100**, 13940–13945 (2003).
39. Vogeley, L. *et al.* Anabaena sensory rhodopsin: a photochromic color sensor at 2.0 Å. *Science (1979)* **306**, 1390–1393 (2004).
40. Nakao, S., Kojima, K. & Sudo, Y. Phototriggered Apoptotic Cell Death (PTA) Using the Light-Driven Outward Proton Pump Rhodopsin Archaerhodopsin-3. *J Am Chem Soc* **144**, 3771–3775 (2022).
41. Ernst, O. P. *et al.* Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. *Chem Rev* **114**, 126–163 (2014).
42. Ihara, K. *et al.* Evolution of the archaeal rhodopsins: evolution rate changes by gene duplication and functional differentiation. *J Mol Biol* **285**, 163–174 (1999).
43. Silapetere, A. *et al.* QuasAr Odyssey: the origin of fluorescence and its voltage sensitivity in microbial rhodopsins. *Nat Commun* **13**, 1–20 (2022).
44. Maclaurin, D., Venkatachalam, V., Lee, H. & Cohen, A. E. Mechanism of voltage-sensitive fluorescence in a microbial rhodopsin. *Proceedings of the National Academy of Sciences* **110**, 5939–5944 (2013).

45. McIsaac, R. S., Bedbrook, C. N. & Arnold, F. H. Recent advances in engineering microbial rhodopsins for optogenetics. *Curr Opin Struct Biol* **33**, 8–15 (2015).
46. McIsaac, R. S. *et al.* Directed evolution of a far-red fluorescent rhodopsin. *Proceedings of the National Academy of Sciences* **111**, 13034–13039 (2014).
47. Engqvist, M. K. M. *et al.* Directed evolution of Gloeobacter violaceus rhodopsin spectral properties. *J Mol Biol* **427**, 205–220 (2015).
48. Piatkevich, K. D. *et al.* A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat Chem Biol* **14**, 352–360 (2018).
49. saint Clair, E. C. *et al.* Near-IR resonance Raman spectroscopy of archaerhodopsin 3: effects of transmembrane potential. *J Phys Chem B* **116**, 14592–14601 (2012).
50. Penzkofer, A., Silapetere, A. & Hegemann, P. Photocycle dynamics of the Archaerhodopsin 3 based fluorescent voltage sensor Archon2. *J Photochem Photobiol B* **225**, 112331 (2021).
51. Penzkofer, A., Silapetere, A. & Hegemann, P. Photocycle dynamics of the Archaerhodopsin 3 based fluorescent voltage sensor QuasAr1. *Int J Mol Sci* **21**, 160 (2019).
52. Hochbaum, D. R. *et al.* All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat Methods* **11**, 825–833 (2014).
53. Mure, L. S. *et al.* Melanopsin bistability: a fly's eye technology in the human retina. *PLoS One* **4**, e5991 (2009).
54. Lythgoe, J. N. & Partridge, J. C. Visual pigments and the acquisition of visual information. *Journal of Experimental Biology* **146**, 1–20 (1989).
55. Sephus, C. D. *et al.* Earliest photic zone niches probed by ancestral microbial rhodopsins. *bioRxiv* 2011–2021 (2022).
56. Bamann, C., Bamberg, E., Wachtveitl, J. & Glaubitz, C. Proteorhodopsin. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1837**, 614–625 (2014).
57. Pedraza-González, L., Mar\'\in, M. del C., de Vico, L., Yang, X. & Olivucci, M. On the automatic construction of QM/MM models for biological photoreceptors: rhodopsins as model systems. in *QM/MM studies of light-responsive biological systems* 1–75 (Springer, 2021).
58. Melaccio, F., Ferré, N. & Olivucci, M. Quantum chemical modeling of rhodopsin mutants displaying switchable colors. *Physical Chemistry Chemical Physics* **14**, 12485–12495 (2012).
59. Ryazantsev, M. N., Altun, A. & Morokuma, K. Color tuning in rhodopsins: the origin of the spectral shift between the chloride-bound and anion-free forms of halorhodopsin. *J Am Chem Soc* **134**, 5520–5523 (2012).
60. Luecke, H., Schobert, B., Lanyi, J. K., Spudich, E. N. & Spudich, J. L. Crystal structure of sensory rhodopsin II at 2.4 angstroms: insights into color tuning and transducer interaction. *Science (1979)* **293**, 1499–1503 (2001).
61. Hoffmann, M. *et al.* Color tuning in rhodopsins: the mechanism for the spectral shift between bacteriorhodopsin and sensory rhodopsin II. *J Am Chem Soc* **128**, 10808–10818 (2006).
62. Katayama, K., Sekharan, S. & Sudo, Y. Color tuning in retinylidene proteins. *Optogenetics* 89–107 (2015).
63. Wang, W., Geiger, J. H. & Borhan, B. The photochemical determinants of color vision: revealing how opsins tune their chromophore's absorption wavelength. *Bioessays* **36**, 65–74 (2014).
64. Honig, B. *et al.* An external point-charge model for wavelength regulation in visual pigments. *J Am Chem Soc* **101**, 7084–7086 (1979).

65. Motto, M. G., Sheves, M., Tsujimoto, K., Balogh-Nair, V. & Nakanishi, K. Opsin shifts in bovine rhodopsin and bacteriorhodopsin. Comparison of two external point-charge models. *J Am Chem Soc* **102**, 7947–7949 (1980).
66. del Carmen Marín, M. *et al.* Fluorescence enhancement of a microbial rhodopsin via electronic reprogramming. *J Am Chem Soc* **141**, 262–271 (2018).
67. Michl, J., Bonacic-Koutecky, V. & others. *Electronic aspects of organic photochemistry*. (Wiley, 1990).
68. 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).
69. Barneschi, L. *et al.* On the fluorescence enhancement of arch neuronal optogenetic reporters. *Nat Commun* **13**, 1–9 (2022).
70. Palombo, R. *et al.* Retinal chromophore charge delocalization and confinement explain the extreme photophysics of Neorhodopsin. *Nat Commun* **13**, 1–9 (2022).
71. Jensen, F. *Introduction to computational chemistry*. (John wiley & sons, 2017).
72. Nelson, E. Derivation of the Schrödinger equation from Newtonian mechanics. *Physical review* **150**, 1079 (1966).
73. Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *J Chem Phys* **98**, 1372–1377 (1993).
74. Malmqvist, P.-Å. & Roos, B. O. The CASSCF state interaction method. *Chem Phys Lett* **155**, 189–194 (1989).
75. 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).
76. Ghigo, G., Roos, B. O. & Malmqvist, P.-Å. A modified definition of the zeroth-order Hamiltonian in multiconfigurational perturbation theory (CASPT2). *Chem Phys Lett* **396**, 142–149 (2004).
77. Battaglia, S. & Lindh, R. Extended dynamically weighted CASPT2: The best of two worlds. *J Chem Theory Comput* **16**, 1555–1567 (2020).
78. Finley, J., Malmqvist, P.-Å., Roos, B. O. & Serrano-Andrés, L. The multi-state CASPT2 method. *Chem Phys Lett* **288**, 299–306 (1998).
79. Andersson, K., Malmqvist, P. A., Roos, B. O., Sadlej, A. J. & Wolinski, K. Second-order perturbation theory with a CASSCF reference function. *Journal of Physical Chemistry* **94**, 5483–5488 (1990).
80. Levitt, M. & Warshel, A. Computer simulation of protein folding. *Nature* **253**, 694–698 (1975).
81. Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol* **103**, 227–249 (1976).
82. Dapprich, S., Komáromi, I., Byun, K. S., Morokuma, K. & Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *Journal of Molecular Structure: THEOCHEM* **461**, 1–21 (1999).
83. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **15**, 5–6 (2006).
84. Krivov, G. G., Shapovalov, M. v & Dunbrack Jr, R. L. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins: Structure, Function, and Bioinformatics* **77**, 778–795 (2009).

85. Bas, D. C., Rogers, D. M. & Jensen, J. H. Very fast prediction and rationalization of pKa values for protein–ligand complexes. *Proteins: Structure, Function, and Bioinformatics* **73**, 765–783 (2008).
86. Morozenko, A. & Stuchebrukhov, A. A. Dowser++, a new method of hydrating protein structures. *Proteins: Structure, Function, and Bioinformatics* **84**, 1347–1357 (2016).
87. van der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *J Comput Chem* **26**, 1701–1718 (2005).
88. Aquilante, F. *et al.* Molcas 8: New capabilities for multiconfigurational quantum chemical calculations across the periodic table. Preprint at (2016).
89. Ponder, J. W. & others. TINKER: Software tools for molecular design. *Washington University School of Medicine, Saint Louis, MO* **3**, (2004).
90. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668–1688 (2005).
91. Broser, M. *et al.* NeoR, a near-infrared absorbing rhodopsin. *Nat Commun* **11**, 1–10 (2020).
92. Sugiura, M. *et al.* Unusual Photoisomerization Pathway in a Near-Infrared Light Absorbing Enzymerhodopsin. *J Phys Chem Lett* **13**, 9539–9543 (2022).
93. Broser, M. Far-red absorbing rhodopsins, insights from heterodimeric Rhodopsin-cyclases. *Front Mol Biosci* 1384 (2022).
94. Hammond, G. S. A correlation of reaction rates. *J Am Chem Soc* **77**, 334–338 (1955).
95. Romei, M. G., Lin, C.-Y., Mathews, I. I. & Boxer, S. G. Electrostatic control of photoisomerization pathways in proteins. *Science (1979)* **367**, 76–79 (2020).
96. Broser, M. *et al.* NeoR, a near-infrared absorbing rhodopsin. *Nat Commun* **11**, 1–10 (2020).
97. Yan, B. *et al.* Spectral Tuning in Bacteriorhodopsin in the Absence of Counterion and Coplanarization Effects (∗). *Journal of Biological Chemistry* **270**, 29668–29670 (1995).
98. Liu, R. S. H. *et al.* Analyzing the red-shift characteristics of azulenic, naphthyl, other ring-fused and retinyl pigment analogs of bacteriorhodopsin. *Photochem Photobiol* **58**, 701–705 (1993).
99. Ottolenghi, M. & Sheves, M. Synthetic retinals as probes for the binding site and photoreactions in rhodopsins. *J Membr Biol* **112**, 193–212 (1989).
100. Kim, S.-H. *et al.* Color-tuning of natural variants of heliorhodopsin. *Sci Rep* **11**, 1–9 (2021).
101. Kim, S. Y., Waschuk, S. A., Brown, L. S. & Jung, K.-H. Screening and characterization of proteorhodopsin color-tuning mutations in Escherichia coli with endogenous retinal synthesis. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1777**, 504–513 (2008).
102. Nakajima, Y. *et al.* Pro219 is an electrostatic color determinant in the light-driven sodium pump KR2. *Commun Biol* **4**, 1–15 (2021).
103. Lin, S. W. *et al.* Mechanisms of spectral tuning in blue cone visual pigments: visible and raman spectroscopy of blue-shifted rhodopsin mutants. *Journal of Biological Chemistry* **273**, 24583–24591 (1998).
104. Soppa, J. *et al.* Bacteriorhodopsin Mutants of Halobacterium sp. GRB: II. Characterization of mutants. *Journal of Biological Chemistry* **264**, 13049–13056 (1989).
105. Sakmar, T. P., Menon, S. T., Marin, E. P. & Awad, E. S. Rhodopsin: insights from recent structural studies. *Annu Rev Biophys Biomol Struct* **31**, 443–484 (2002).
106. Krebs, M. P., Mollaaghababa, R. & Khorana, H. G. Gene replacement in Halobacterium halobium and expression of bacteriorhodopsin mutants. *Proceedings of the National Academy of Sciences* **90**, 1987–1991 (1993).

107. Luecke, H., Schobert, B., Lanyi, J. K., Spudich, E. N. & Spudich, J. L. Crystal structure of sensory rhodopsin II at 2.4 angstroms: insights into color tuning and transducer interaction. *Science (1979)* **293**, 1499–1503 (2001).
108. Royant, A. *et al.* X-ray structure of sensory rhodopsin II at 2.1-Å resolution. *Proceedings of the National Academy of Sciences* **98**, 10131–10136 (2001).
109. Nakanishi, K., Balogh-Nair, V., Arnaboldi, M., Tsujimoto, K. & Honig, B. An external point-charge model for bacteriorhodopsin to account for its purple color. *J Am Chem Soc* **102**, 7945–7947 (1980).
110. Wang, W., Geiger, J. H. & Borhan, B. The photochemical determinants of color vision: revealing how opsins tune their chromophore's absorption wavelength. *Bioessays* **36**, 65–74 (2014).
111. Wang, W. *et al.* Tuning the electronic absorption of protein-embedded all-trans-retinal. *Science (1979)* **338**, 1340–1343 (2012).
112. Clair, E. C. saint, Ogren, J. I., Mamaev, S., Kralj, J. M. & Rothschild, K. J. Conformational changes in the archaerhodopsin-3 proton pump: detection of conserved strongly hydrogen bonded water networks. *J Biol Phys* **38**, 153–168 (2012).
113. Flytzanis, N. C. *et al.* Archaerhodopsin variants with enhanced voltage-sensitive fluorescence in mammalian and Caenorhabditis elegans neurons. *Nat Commun* **5**, 1–9 (2014).
114. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261–272 (2020).
115. Fdez. Galván, I. *et al.* OpenMolcas: From source code to insight. *J Chem Theory Comput* **15**, 5925–5964 (2019).

Leonardo Barneschi

Signed 01/03/2023