Topics in Current Chemistry (2022) 380:21
https://doi.org/10.1007/s41061-022-00374-w

REVIEW

Evolution of the Automatic Rhodopsin Modeling (ARM) Protocol

Laura Pedraza-González[1,3] · Leonardo Barneschi[1] · Daniele Padula[1] · Luca De Vico[1] · Massimo Olivucci[1,2]

Received: 3 December 2021 / Accepted: 29 January 2022

Abstract

In recent years, photoactive proteins such as rhodopsins have become a common target for cutting-edge research in the field of optogenetics. Alongside wet-lab research, computational methods are also developing rapidly to provide the necessary tools to analyze and rationalize experimental results and, most of all, drive the design of novel systems. The Automatic Rhodopsin Modeling (**ARM**) protocol is focused on providing exactly the necessary computational tools to study rhodopsins, those being either natural or resulting from mutations. The code has evolved along the years to finally provide results that are **reproducible** by any user, **accurate** and **reliable** so as to replicate experimental trends. Furthermore, the code is **efficient** in terms of necessary computing resources and time, and **scalable** in terms of both number of concurrent calculations as well as features. In this review, we will show how the code underlying ARM achieved each of these properties.

Keywords QM/MM · Rhodopsins · Photochemistry · Photobiology · Python package

Laura Pedraza-González, Luca De Vico and Massimo Olivucci contributed equally to this work.

This article is part of the Topical Collection "New Horizon in Computational Chemistry Software"; edited by Michael Filatov, Cheol. H. Choi and Massimo Olivucci.

✉ Laura Pedraza-González
laura.pedraza@unisi.it

✉ Luca De Vico
Luca.DeVico@unisi.it

✉ Massimo Olivucci
olivucci@unisi.it

Extended author information available on the last page of the article

Published online: 15 March 2022

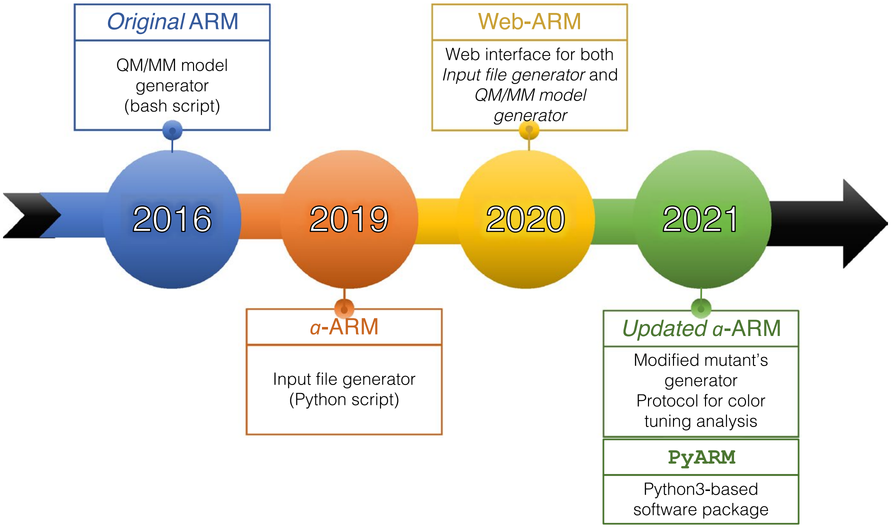

Original ARM
QM/MM model generator (bash script)
Web-ARM
Web interface for both *Input file generator* and *QM/MM model generator*
2016
2019
2020
2021
a-ARM
Input file generator (Python script)
Updated a-ARM
Modified mutant's generator
Protocol for color tuning analysis
PyARM
Python3-based software package

Fig. 1 Timeline for Automatic Rhodopsin Modeling (ARM) development detailing various parts of the ARM protocol and the years of the corresponding publications

1 Introduction: Contents and Scope

This review deals with the development of the Automatic Rhodopsin Modeling protocol (ARM), as seen in the last 5 years (Fig. 1). ARM represents a protocol capable of reliably building and analyzing computational models of rhodopsins, a superfamily of photoactive proteins. Given their major role in many aspects of life, from the basic act of ion-gating and ion-pumping in Archaea and Eubacteria to vision in animals, rhodopsins are currently studied extensively. Furthermore, rhodopsins and their engineered mutants represent powerful tools for potential biotechnological applications, the most prominent of which is optogenetics. Thus, researchers are constantly looking for new rhodopsins with particular photochemical properties, those being, among others, a specific absorption wavelength, a long or short excited state lifetime, and a strong fluorescence. Section 2 gives a panoramic view of rhodopsins and discusses current and future technological applications.

ARM was developed to provide a basic quantum-mechanics molecular mechanics (QM/MM) computational model, but sufficiently accurate so that differences in the model would reflect actual changes in the behavior of a different/mutated rhodopsin, and vice versa. This first iteration of ARM, called ARM$_{\text{original}}$, was created to generate models capable of accurately reproducing the spectral trends observed for a limited set of rhodopsins, and will be described in Sect. 3. In particular, ARM$_{\text{original}}$ models have been shown to be capable of reproducing trends in light absorbance maximum values in rhodopsin of different origins, and provide effective tools to discern the causes of effects such as blue- or red-shifting of the absorbed wavelength.

The natural extension of the ARM code has been to extend the general accuracy and applicability of the models and, most importantly, the level of automation in building protocol. Section 4 presents the *advanced* Automatic Rhodopsin Model building protocol (*a*-ARM), which meant completely rewriting the previous ARM code, and incorporating the possibility to automatically prepare inputs for the protocol itself. The completely new input preparation phase removed the need for user files manipulation and possible source of errors, hence achieving **reliability** and complete **reproducibility** of the results. The code behind *a*-ARM has also been used to power a web-interface, which allows, in principle, any user to obtain rhodopsin models of the same quality (Web-ARM, Sect. 5). This meant the extension of the benchmarking pool to more rhodopsins, thus increasing the applicability of the protocol, which, in turn, made it possible to use ARM models to guide the rational design of rhodopsins. More specifically, *a*-ARM can now be used to suggest which residue to mutate to obtain, for instance, a desired spectroscopic effect.

Finally, the ARM code was encapsulated inside a python-based package, called PyARM and described in Sect. 6. While *a*-ARM was already efficient in producing a single rhodopsin model, PyARM allows hundreds of concurrent rhodopsin models to be obtained automatically. This high level of efficiency is obtained by allowing the code to completely take care of all necessary calculations (complete automation), through the clever use of a highly modular code structure. Different types of analyses are now possible, all through the use of automatic, user-friendly, command-line Python drivers. Finally, the modular nature of PyARM allows the easy implementation of additional features, thus scaling-up the usability of the code with additional features.

2 Rhodopsins: a Family of Biological Photoreceptors

2.1 Structure and Diversity

Rhodopsins constitute a vast family of photoreceptive, membrane-embedded, seven-helix proteins, structurally formed by an opsin apoprotein and a retinal that serves as a chromophore[1] to absorb photons for either energy conversion or the initiation of intra- or intercellular signaling. As illustrated in Fig. 2, the opsin features an internal pocket hosting the retinal chromophore, which is covalently attached by a Schiff base linkage to the ε-amino group of a lysine side-chain in the middle of helix VII; the resulting retinal Schiff base (*r*SB) is protonated in most cases, constituting the so-called retinal protonated Schiff base (*r*PSB), illustrated in Fig. 3. Changes in protonation state of the *r*SB are crucial for both signaling and transport activity of rhodopsins [5–7].

Recent genomic advances have revealed that tens of thousands of rhodopsin genes are distributed widely in all domains of life (i.e., *Eukaryotes*, *Bacteria*, and *Archaea*) [5–10], and reside in many diverse organisms such as animals (e.g., vertebrates and invertebrates), microorganisms, and even viruses [11, 12]. Based on their

[1] Part of a molecular entity consisting of an atom or moiety in which the electronic transition responsible for a given spectral band above 200 nm is approximately localized [4].

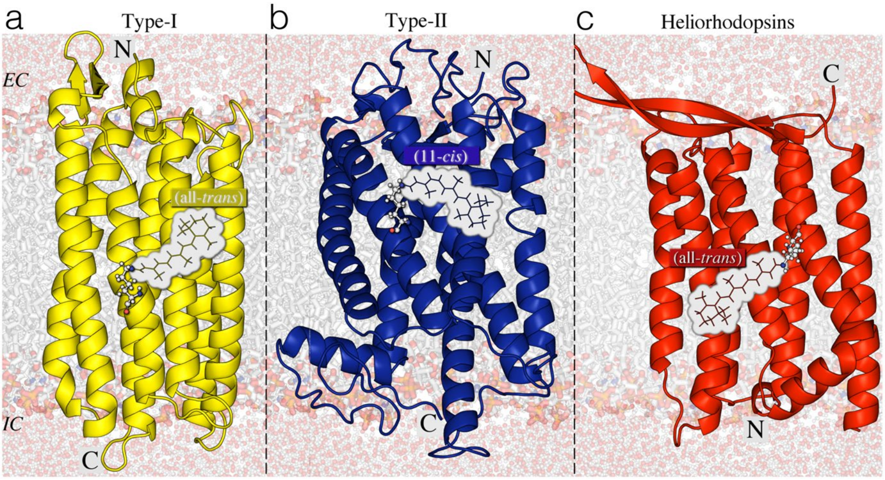


Fig. 2 **a**–**c** Rhodopsin types: structural similarities and differences. Rhodopsin proteins are classified into three types: yellow cartoon microbial or type-I, blue cartoon animal or type-II, red cartoon heliorhodopsins. For each type, the retinal chromophore in the dark adapted state (DA) is presented as lines and the covalently linked lysine is shown as ball-and-sticks. The orientation of the N terminus and C terminus residues with respect to the extracellular (EC) and intracellular (IS) surfaces of the membrane is specified. The structures correspond to **a** KR2 (type-I) [6REW [1]], **b** Rh (Type-II) [1U19 [2]] and **c** TaHeR (heliorhodopsin) [6IS6 [3]]

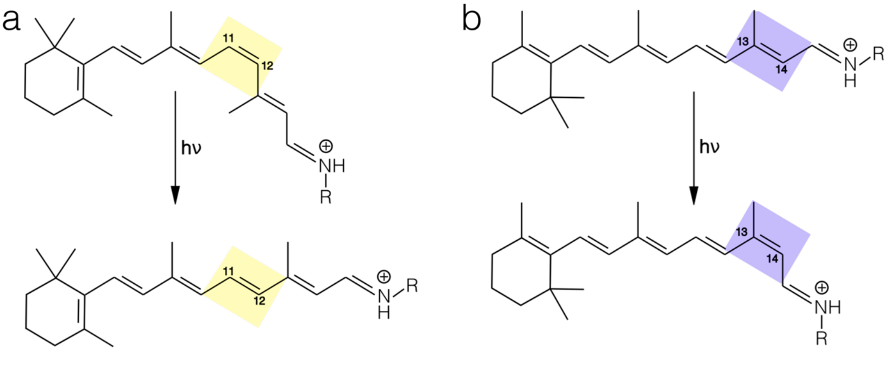


Fig. 3 Primary photoreaction in animal, microbial and helio-rhodopsins. Retinal isomerization from **a** the 11 – *cis* to the all-*trans* form and **b** from the all-*trans* to the 13 – *cis* form is the primary reaction in animal and microbial/helio-rhodopsins, respectively

host organism, rhodopsins are divided into different types (see Fig. 2): animal (type II) rhodopsins, a specialized subset of G-protein-coupled receptor (GPCR); microbial (type I) rhodopsins [6]; and heliorhodopsins—a recently discovered new type of light-sensing microbial rhodopsins [13].

Although members of the three types display a remarkably constant general architecture, they exhibit large differences in amino acid sequence (i.e., they have almost no sequence homology), as reflected in different chromophore cavities as well as in certain structural features [5, 7, 8, 13, 14]. In animal rhodopsins, the equilibrium dark adapted state (DA) shows a 11-*cis* (i.e., C11–C12 double bond) *r*PSB chromophore, which photoisomerizes to its all-*trans* configuration (Fig. 3a). In both microbial and heliorhodopsin families the DA state is dominated by an all − *trans* (i.e., C13–C14 double bond) *r*PSB chromophore, which is usually transformed into the 13 − *cis* configuration (see Fig. 3b) upon light absorption. Finally, in heliorhodopsins, the N-terminus amino acid is exposed to the cytoplasm or intracellular (IC) part, and the C-terminus residue faces the extracellular (EC) part of the cell membrane, whereas the opposite happens for the other two families (Fig. 2) [15].

2.2 Biological Functions

Rhodopsins exhibit an extensive pool of biological functions [5, 6, 8, 9, 16, 17]. For instance, animals use the photosensory functions (i.e., visual responses) of type II rhodopsins, lower organisms utilize type I for light energy conversion and intracellular signaling, while the function of heliorhodopsins is still not well elucidated [18–20].

In particular, microbial rhodopsins (type I) are represented in many diverse microorganisms, spanning the three domains of cellular life, as well as in giant viruses [12]. As such, modifications of a single protein scaffold through evolution produced many novel, chemical, light-dependent biological functions [5–7, 9, 10, 15, 17, 21]. Such functions can be roughly divided into two categories: (1) photoenergy transducers able to convert light into electrochemical potential to energize cells (e.g., light-driven ion pumps catalyzing outward active transport of protons [H^+], inward chloride transport [Cl^-], and outward sodium transport [Na^+]); and (2) photosensory receptors that make use of light to gain information about the environment to regulate inter- or intracellular processes (e.g., photosensors with membrane-embedded or soluble transducers, ion channels [anion and cation channel rhodopsins (ChRs)], and enzyme rhodopsins) [5–7, 9, 17, 22].

2.3 Photoreactivity

Rhodopsin functions are triggered by highly stereo-selective photoisomerization events, i.e., generally *cis* → *trans* of the C11=C12 bond for type II and *trans* → *cis* of the C13=C14 for type I and heliorhodopsins (Fig. 3). The specific wavelength (hereinafter referred to as Maximum absorption wavelength ($\lambda^{a}_{\mathrm{max}}$)) is due to a phenomenon called “opsin-shift”, which consists of spectral shifts, in a wide range of the UV/visible spectra, modulated by the interaction between the retinal chromophore and the opsin residues.

Figure 4 depicts the inter-state photoisomerization pathways of rhodopsins, representing the potential energy change, i.e., the progression along the potential energy surface (PES), along a reaction coordinate possibly involving a combination of electronic transitions between the Ground-state (S_0) and both first (S_1) and second (S_2) excited electronic states. Such a coordinate is usually considered as the

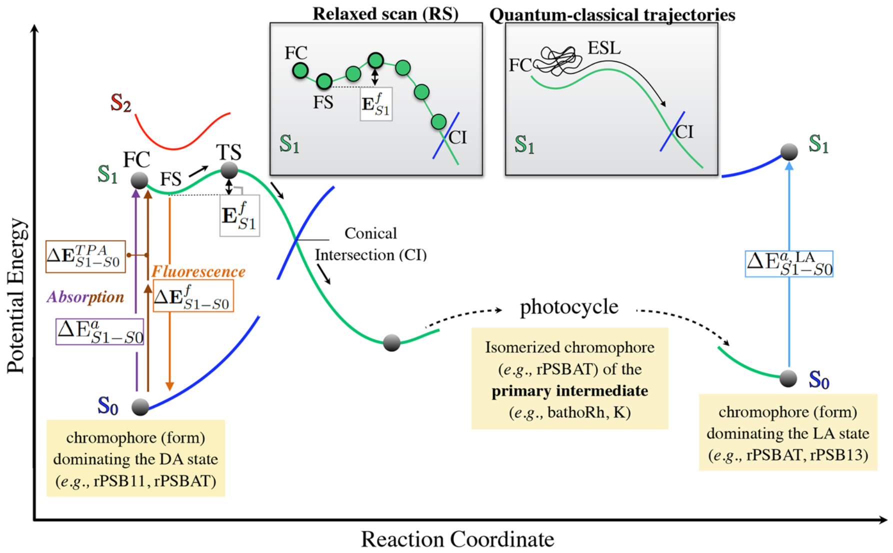

Relaxed scan (RS)
FC
FS
E^f_{S1}
S_1
CI
Quantum-classical trajectories
FC
ESL
S_1
CI
S_2
S_1
FC
FS
TS
E^f_{S1}
ΔE^{TPA}_{S1-S0}
Fluorescence
Absorption
ΔE^f_{S1-S0}
ΔE^a_{S1-S0}
S_0
Conical Intersection (CI)
photocycle
Isomerized chromophore (e.g., rPSBAT) of the primary intermediate (e.g., bathoRh, K)
chromophore (form) dominating the DA state (e.g., rPSB11, rPSBAT)
S_1
$\Delta E^{a,LA}_{S1-S0}$
S_0
chromophore (form) dominating the LA state (e.g., rPSBAT, rPSB13)
Potential Energy
Reaction Coordinate

Fig. 4 Light-induced and light-emission properties of rhodopsin proteins, investigated using computational modeling. Schematic diagram displaying the photoiomerization path, including the relevant S_0, S_1 and S_2 energy profiles of a generic rhodopsin. QM/MM models are used to compute the vertical excitation energy for absorption ($\Delta E^a_{S1-S0} \equiv \lambda^a_{\max}$), fluorescence ($\Delta E^f_{S1-S0} \equiv \lambda^f_{\max}$) and two-photon absorption ($\Delta E^{TPA}_{S1-S0} \equiv \lambda^{TPA}_{\max}$) as well as the excited state energy isomerization barrier (E^f_{S1}) associated with emission, computed as the energy difference between the fluorescent excited-state (FS) structure and the transition state (TS). *Inset* (top-center) Schematic illustration of the calculation of an excited state isomerization path providing the E^f_{S1} value via a Relaxed scan (RS) and of a Franck–Condon (FC) quantum-classical trajectory (this provides, in case of an ultrafast reaction, an estimate of the ESL associated to the double bond energy isomerization barrier). QM/MM models can also been used to investigate the structure and spectroscopy of primary photocycle intermediates (batho and K intermediates) and of photocycle intermediates corresponding to light-adapted states

skeletal dihedral angle of the double-bond highlighted in Fig. 3. The involved electronic states commonly change their electronic character and can cross [23, 24]. Different rhodopsin "functions" may take advantage of different features of the rhodopsin PES, illustrated in the following.

The primary event of light absorption of one photon of energy $h\nu$ prompts the vertical electronic π-π^* transition of the retinal chromophore from the ground to the first excited Franck–Condon[2] (FC) state. The length of the π-conjugated polyene chain in the retinal chromophore, and the protonation state of the *r*SB linkage, determine the energy gap, hereafter called Vertical Excitation energy (ΔE^a_{S1-S0}), of this process [5]. The computational evaluation of ΔE^a_{S1-S0} allows the prediction of $\lambda^a_{\max}$ for a given rhodopsin, and, in general, for the modeling of light absorption. Considering that the wavelength dependence of the absorption efficiency defines the colors of the rhodopsin proteins, the modulation of the ΔE^a_{S1-S0} gap makes possible the engineering

[2] Classically, the Franck–Condon principle is the approximation that an electronic transition is most likely to occur without changes in the positions of the nuclei in the molecular entity and its environment. The resulting state is called a Franck–Condon state, and the transition, a vertical transition [4].

of either blue- or red-shifted rhodopsin variants, absorbing in specific regions of the UV-Vis spectrum (i.e., color tuning, see Sect. 6.4) [19–21, 25–35]. Photoexcitation to the S_1 state can also be achieved by simultaneous absorption of two infrared photons (i.e., Two-Photon Absorption (TPA) process) of the same energy $\Delta E^{TPA}_{S1\text{-}S0}$, corresponding to half of the energy necessary for One-Photon Absorption (OPA) [23, 36–40].

After photoexcitation, the retinal chromophore leaves the FC region by relaxing along stretching and torsional modes, and starts the exploration of the S_1 PES. Depending on the surface topography, it may (1) quickly encounter a minimum, hereinafter called fluorescent excited-state (FS), or (2) continue visiting other regions of the S_1 PES. In (1), the chromophore remains in a long-lived excited state that eventually decays back to the ground state via fluorescence, i.e., spontaneous emission of radiation (luminescence) [4]. In this case, a photon of a different wavelength can be emitted after a short while (10^{-9} to 10^{-5} s). The energy gap associated to this process is denominated Vertical Emission energy (ΔE^{f}_{S1-S0}). Therefore, the computational evaluation of ΔE^{f}_{S1-S0} allows the prediction of the Maximum emission wavelength ($\lambda^{f}_{\max}$) for a given rhodopsin, and in general the modeling of emission properties, such as fluorescence. In (2), that is, if there is a shallow or no FS, the retinal chromophore twists around the reactive C=C bond and reaches the Conical intersection (CI) region, where it decays to S_0. As illustrated in Fig. 4, "reacting" rhodopsins then trigger a series of sequential protein moiety conformational changes (required for their biological functions) [8, 10, 23] and returns to the initial state (e.g., process known as photocycle), whereas the "non-reacting" molecules relax back to the original S_0 state without entering such a photocycle. The described cycle allows microbial rhodopsins to repeat their functions every light stimulation since the chromophore is ultimately regenerated through the photocycle. This type of photocycle (in chemistry one would talk about type 1 photochromics) is remarkably different from that of vertebrate visual opsins (i.e., vertebrate rhodopsin and cone visual pigments), in which the retinal chromophore dissociates after the photoreaction and, therefore, additional retinal is required to regenerate the pigments [9].

2.4 Applications of Natural and Engineered Rhodopsins: Optogenetics

The modeling of primary photoproducts, photoisomerization reaction paths and bistable states in animal and microbial rhodopsins is of great interest for engineering photoswitchable fluorescent probes [41–43]. Bistable rhodopsins are rhodopsins featuring two stable isomeric forms (i.e., characterized by two chromophore isomers such all − *trans* and 13-*cis*) and thus require the sequential and independent absorption of two photons (often of different wavelengths) to complete the photocycle [43]. Certain bistable rhodopsins can be interconverted using light of different wavelengths (type-2 or P-type photochromism[3])

[3] Reversible transformation of a molecular entity between two forms, A and B, having different absorption spectra, induced in one or both directions by absorption of electromagnetic radiation. The spectral change produced is typically, but not necessarily, of visible color and is accompanied by differences in other physical properties [4].

[44]. However, more frequently the light adapted state (LA) state reverts back to the DA state state thermally (type-1 or T-type photochromism), in which case (i.e., for monostable microbial rhodopsins) the photocycle is completed after the absorption of only one photon [44]. Applications of bistable rhodopsins are related to different properties/functions of the DA and LA states and the use of light irradiation to change the rhodopsin isomeric composition passing from a DA-dominated to a LA-dominated equilibrium. Since the efficiency of such conversion is proportional to the difference between the two $\lambda^{a}_{\max}$ values, it is apparent that achieving bistable rhodopsins featuring one form with a $\lambda^{a}_{\max}$ value significantly shifted to the red, may facilitate applications where it is important to switch on-and-off (i.e., control) the rhodopsin biological function using light irradiation [43].

The photoisomerization event that initiates each rhodopsin function (see Sect. 2.2), has been studied widely in scientific areas such as physics, chemistry, and biology [5–7], with the aim of allowing for novel and modern applications in fields as diverse as medicine [45–47], bioenergetics [9, 48], biotechnology and neurosciences [7–10, 16], among others [9]. Particularly, specific microbial rhodopsins that function as either ion transporters or channels are at the heart of a new biotechnology called optogenetics.

Optogenetics (i.e., a combination of “optics” and “genetics”) uses mainly genetically encoded, and often specifically engineered, microbial rhodopsins for the optical control of physiological processes [7–10, 16, 41]. The development of so-called optogenetic tools leads to the investigation of the nervous system at the cellular and tissue level, without noticeable tissue damage as well as side effects. Currently, rhodopsins with potential application as optogenetic tools are used as light-driven actuators (i.e., action potential triggers), light-driven silencers (i.e., action potential quenchers) and as fluorescent reporters (i.e., action potential probes) [7].

In order to either improve the current or develop new optogenetic tools, it is imperative to gain further insight into the different factors controlling color tuning in rhodopsins, to then be able to increase the variety of $\lambda^{a}_{\max}$, thus enabling simultaneous optical control by different colors of light [19–21, 25–35]. As such, various rhodopsin genes have been screened in order to find additional colors [49, 50]. In particular, while many blue-absorbing rhodopsin at $\lambda^{a}_{\max}$ < 500 nm have been reported [51] and even applied to optogenetics [49], the longer absorption maxima are limited in $\lambda^{a}_{\max}$ < 600 nm. In this regard, there is presently an interest in screening rhodopsin variants exhibiting a longer $\lambda^{a}_{\max}$ and/or enhanced fluorescence (i.e., high fluorescence intensity) [52], achieved through the effects of single or multiple amino acid mutations of a template Wild-type (WT) structure.

The rational design of artificially mutated variants is necessary to identify the amino acid replacements that are effective for color tuning and for influencing fluorescent properties. A systematic experimental screening of thousands of possible candidates is not feasible, thus requiring the development of computational approaches for a fast, congruous, and rational design of in silico point mutations, to narrow down the number of tested candidates.

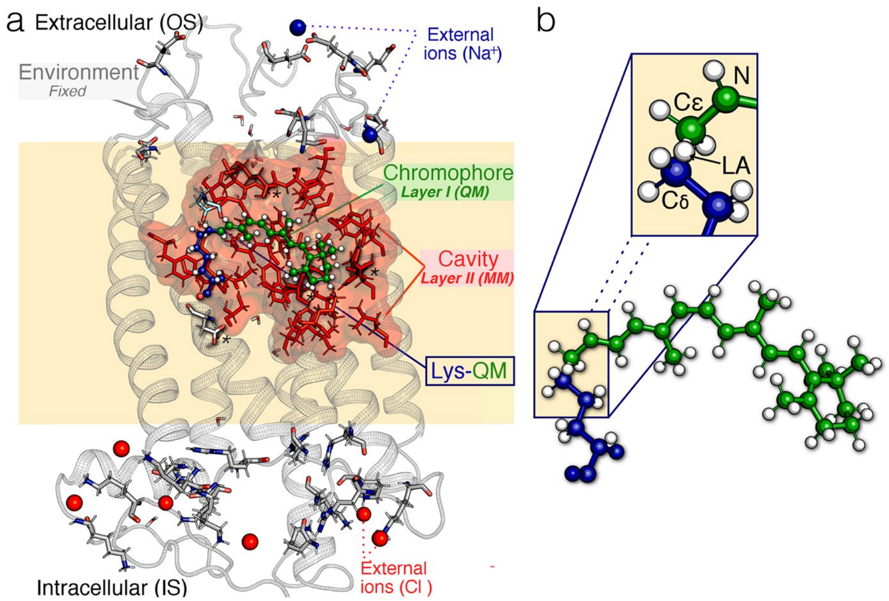


Fig. 5 General structure of a monomeric, gas-phase and globally uncharged Automatic Rhodopsin Modeling (ARM) quantum-mechanics molecular mechanics (QM/MM) model. **a** Relationship between the ARM model's three subsystems and two multiscale layers. Detailed description of the components of the three subsystems (for an animal rhodopsin from the DA state of Bovine rhodopsin). Gray cartoon Environment subsystem, green ball-and-sticks *r*PSB chromophore, blue ball-and-sticks lysine side-chain covalently linked to the *r*PSB chromophore, cyan tubes main chromophore counterion, tubes marked with * residues with non-standard protonation states, red spheres external Cl , blue spheres Na$^+$ counterions, red/white tubes crystallographic water molecules, red frames surface amino acid residues forming the chromophore cavity subsystem, gray tubes external OS and IS charged residues. **b** The *r*PSB chromophore (green) and the linked Lys side-chain fragment (blue) form the Lys-QM subsystem, which includes the H-link atom located along Cε–Cδ connecting blue and green atoms (inset: *LA* H-linked atom), which belong to the MM and QM parts of the model. Adapted with permission from [40]. Copyright 2019 American Chemical Society

3 The *Original* Version of ARM: a Pioneer Technology for Rhodopsin QM/MM Modeling

3.1 State-of-the-Art for QM/MM Modeling of Rhodopsins

The past few decades have witnessed a growing interest in developing hybrid QM/MM approaches to tackle problems in computational photochemistry and photobiology [27]. A remarkable advantage of using hybrid QM/MM methodologies is that in silico models featuring a high-level of complexity can be properly constructed, through the definition of subsystems, each treated at a different level of theory (or layers) according to the required level of accuracy.

As shown in Fig. 5 and further described in detail in Ref. [43], a basic QM/MM model for a biological photoreceptor (e.g., rhodopsin) should feature at least three subsystems: (1) the reactive part of the system, or prosthetic group, that carries the photochemical process (i.e., chromophore), treated with a suitable, and usually computationally expensive, QM method (see Section 2.1.1 in Ref. [43]); (2) the residues that, due to either steric or electrostatic interactions with the prosthetic group, directly influence the role of the reactive part (i.e., amino acids and water molecules forming the chromophore cavity), treated classically with a less expensive (optionally polarizable), MM force field (see Section 2.1.2 in Ref. [43]); and (3) the residues without an evident role on the photochemical process (i.e., protein environment) structurally fixed during the calculation and treated as point charges (see Fig. 5). Usually, QM/MM models are complex and, unfortunately, not univocally defined, thus Ref. [43] collates the different approaches to modeling the photochemical properties of Bovine rhodopsin from *Bos taurus* (Rh), a rhodopsin for which the X-ray structure is available [2, 53, 54] and that, also for this reason, is often taken as a reference for the benchmarking of different QM/MM models. Differences in construction protocols of the QM/MM setup lead to variations in computed $\lambda_{\max}^{a,\text{calc}}$ up to 41 nm [29–31, 55–58].

3.2 ARM Scope

It is important to define a standardized protocol for the fast and automated production of congruous QM/MM models, which can subsequently be replicated in any laboratory. Such a protocol should not strictly aim at the prediction of the absolute values of observable properties, but to the description of their changes along sets of different rhodopsin variants. The ARM protocol described in this work represents our attempt to provide such a tool. The protocol is based on two well-defined phases called “generators”: the (1) input file generator and (2) QM/MM model generator. Sections 3.3 and 3.4 describe the development of the original version of the ARM protocol [59], which provided the QM/MM model generator. Furthermore, Ref. [59] provides a series of instructions for the manual preparation of the input file, which served for the subsequent development of the input file generator presented in Sect. 4.

The original ARM protocol is designed for a semi-automatic, fast and parallel building of congruous sets of QM/MM models of wild-type and mutant rhodopsin-like photoreceptors [59]. Accordingly, as illustrated in Fig. 5, this version provides specialized QM/MM models that, in general, would not be applicable to other (e.g., cytoplasmic) photoresponsive proteins, or even to rhodopsins that contain artificial retinal chromophores. In the general framework of the QM/MM model generator, one needs to consider (1) the wise division of the complex molecular system into different, simpler subsystems, and (2) the definition of particular layers that represent the approaches (i.e., levels of theory) used for the proper description of each of the subsystems.

To assess point (1), one can refer to Sect. 2.1, where it is specified that rhodopsins belonging to the three known families (i.e., animal, microbial and heliorhodopsins) share a common architecture constituted by a protein environment featuring seven transmembrane helices, which form a cavity hosting the retinal chromophore (see Fig. 5a). As illustrated in Fig. 5b, an important feature to be considered is that the chromophore is linked covalently via a specific lysine residue (e.g., located in the middle of helix VII and helix G for animal and microbial rhodopsins, respectively), via an imine (–C=N–) linkage, forming the *r*PSB.

To evaluate point (2), instead, it is crucial to identify the chemical/physical phenomena to be modeled and the target properties to be computed. Section 2.3 shows that the process driving the diverse biological functions of rhodopsins is the photoisomerization of the *r*PSB chromophore, occurring immediately after the absorption of a photon of the appropriate wavelength. As illustrated in Fig. 4, the most relevant properties to be reproduced/predicted are the rhodopsin color (or absorption wavelength), fluorescence, and photochemical reactivity.

The scope of ARM, since inception, has been that of being capable of reproducing experimental trends in rhodopsin series. In other words, it was not designed to be a predictive, but rather an investigative, tool; given a set of rhodopsins, the generated corresponding models should reflect trends in spectroscopic and/or photochemical properties. In turn, the computed models could be used to investigate, at the molecular level, the origin of such property changes.

3.3 Definition of an ARM QM/MM Model

As previously mentioned, it is possible to subdivide the rhodopsin structure into three subsystems. Figure 5a exemplifies how these subsystems are defined, using the case of Rh rhodopsin: the (protein) environment (gray cartoon), the chromophore cavity (red frames/surface), and the Lys-chromophore (blue/green ball-and-sticks). The protein environment sub-system features residues (backbone and side-chain atoms) fixed at the crystallographic or comparative (homology) structure, and incorporates external Cl^{-} and/or Na^{+} counterions (see discussion below) also fixed at pre-optimized positions. The chromophore cavity sub-system, instead, contains residues with fixed backbone and relaxed side-chains. The Lys-QM system contains the atoms of the covalently linked lysine side-chain in contact (through Cδ) with the QM/MM frontier and the entire QM subsystem, which corresponds to a *N*-methylated retinal chromophore. All the Lys-QM atoms are free to relax during the QM/MM calculation.

Accordingly, the ARM QM/MM models illustrated in Fig. 5 can be defined as basic, monomeric, gas-phase and globally uncharged, with electrostatic embedding. The term “gas-phase” refers to the fact that, although rhodopsins are proteins exposed to trans-membrane electrostatic fields (i.e., few tenths of meV) occasioned by an asymmetric distribution of the surface ions, in the ARM basic approach, the interactions between the protein and the membrane, as well as solvation effects, are not modeled explicitly. Instead, the protocol mimics the situation of a globally uncharged protein by adding external Cl^{-} and/or Na^{+} counterions, properly placed near the most positively and/or negatively charged surface amino acids, in both the intracellular (IS) and extracellular (OS) protein surfaces.

To this aim, ARM uses the No (net) Surface Charge (NSC) scheme in which the user must execute the following operations manually: (1) identify the number of positively and negatively charged surface residues, (2) calculate

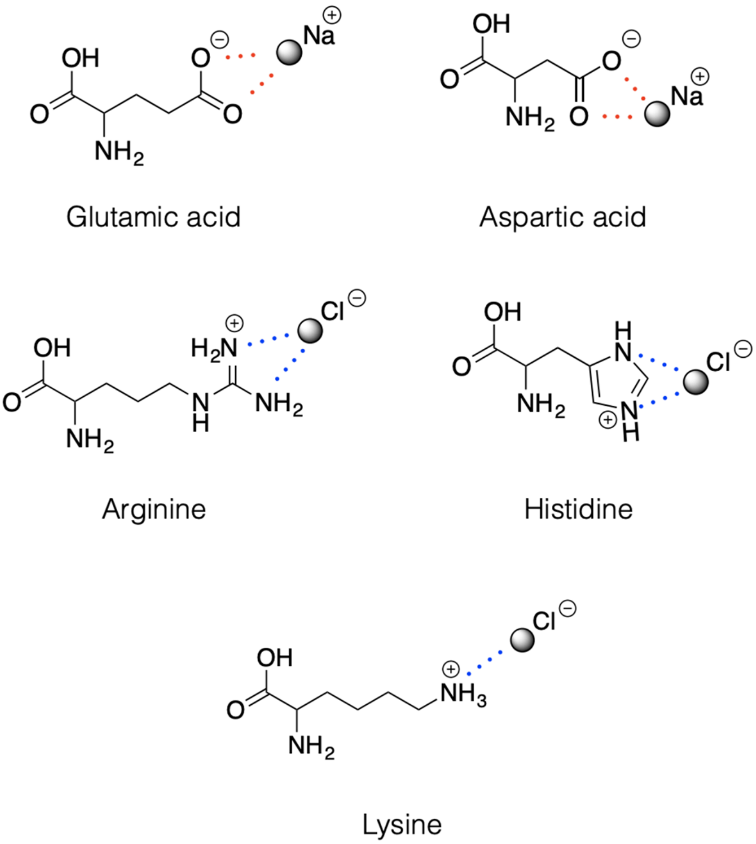


Fig. 6 External counterion positions relative to ionizable residues, according to the No Surface Charge (NSC) scheme used for the ARM QM/MM models Schematic representation of the positions of the external counterions (i.e., Na^+ and Cl) used to neutralize the IS and OS surfaces in the gas-phase ARM QM/MM models. Reproduced with permission from [59] Copyright 2016 American Chemical Society

independently the charge in the IS and OS , (3) determine the number of Cl^- and/or Na^+ to be added, and (4) position the external counterions as illustrated in Fig. 6. As described in Sect. 4, in the most updated version of the protocol, such a procedure is performed automatically by a specific algorithm. Moreover, to account for the water-mediated hydrogen-bond network (HBN) in the protein cavity, the internal crystallographic water molecules are retained in the model, while the water molecules that are not detected experimentally are assumed to be extremely mobile or just absent in the chromophore hydrophobic protein cavity.

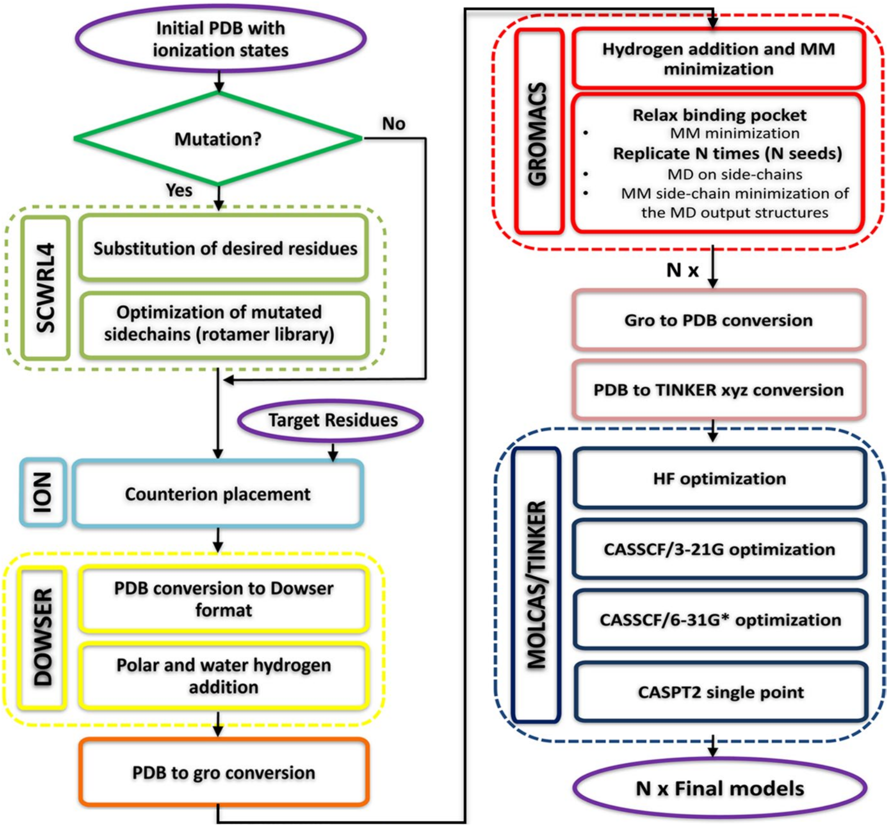

Initial PDB with ionization states
Mutation?
No
Yes
SCWRL4
Substitution of desired residues
Optimization of mutated sidechains (rotamer library)
Target Residues
ION
Counterion placement
DOWSER
PDB conversion to Dowser format
Polar and water hydrogen addition
PDB to gro conversion
GROMACS
Hydrogen addition and MM minimization
Relax binding pocket
• MM minimization
Replicate N times (N seeds)
• MD on side-chains
• MM side-chain minimization of the MD output structures
N x
Gro to PDB conversion
PDB to TINKER xyz conversion
MOLCAS/TINKER
HF optimization
CASSCF/3-21G optimization
CASSCF/6-31G* optimization
CASPT2 single point
N x Final models

Fig. 7 General workflow of the QM/MM model generator developed in the original version of ARM. The procedure starts with the previously prepared input structure and finishes with the S_0 ARM QM/MM model, that consists on 10 replicas of the optimized structure along with the computed Maximum absorption wavelength ($\lambda^a_{\max}$). The different steps of the protocol, as well as the level of theory and the used software are provided Adapted with permission from [59]. Copyright 2016 American Chemical Society

3.4 QM/MM Model Generator

Figure 7 illustrates the general workflow of the QM/MM model generator proposed in Ref. [59] for the semi-automatic building of ground-state ARM QM/MM models of rhodopsins and subsequent computation of the Maximum absorption wavelength, via vertical excitation energy calculations.

3.4.1 Classical Molecular Dynamics Simulations

A preliminary preparation of the input structure is required, which consists of the following steps:

(1) Selection and optimization (i.e., energy criteria) of all the water molecules, using the crystallographic/comparative positions as initial guess.
(2) Addition of hydrogen atoms to polar residues and waters.
(3) Addition of hydrogen atoms to the other residues of the protein environment, chromophore cavity and chromophore subsystems.
(4) MM energy minimization on the added hydrogen atoms.
(5) MM geometry optimization on all the side-chains of the residues of the chromophore cavity subsystem.
(6) Generation of N=10 independent models (replicas) to simulate and explore the possible relative conformational phase space of the cavity residue side-chains and retinal chromophore.

Steps 1 and 2 are performed using the program DOWSER [60]. These steps ensure proper treatment of water and Hydrogen-bond networks (HBN), which affect side-chain conformations and long-range electrostatics, thus ultimately modifying spectral and photochemical properties.

Steps 3–6 are performed using the program GROMACS [61]. Step 6 uses classical molecular dynamics simulations (MD) to perform a simulated annealing relaxation at 298 K on all side-chains of the Lys-QM and cavity subsystem, keeping the backbone fixed at the crystallographic/comparative structure; during the MD computation the retinal chromophore subsystem is also allowed to move. To warrant independent initial conditions, each of the N=10 independent MDs starts with a different, randomly chosen seed. Note that, during the MD run, the chromophore subsystem is represented using an MM parameterization and partial charges computed as AMBER-like Restrained Electrostatic Potential (RESP) charges, which are specific for each used isomer of the chromophore (e.g., 11-*cis*, all-*trans* and 13-*cis* *r*PSB). The corresponding parameterized RESP point charges, currently used in ARM, are reported in the Supplementary Information of Ref. [59]). The default heating, equilibration, and production times for the MD (selected via benchmark calculations in Ref. [59]) are 50, 150, and 800 ps, respectively, for a total length of 1 ns. In each run, the *frame closest to the average* of the 1 ns simulation is then selected as the starting geometry (i.e., guess structure) for constructing the corresponding QM/MM model. Melaccio et al. [59] have shown that, for a set of three phylogenetically diverse rhodopsins (Rh, SqRh and ASR $_{13C}$), N=10 replicas are enough to provide sufficient variability.

3.4.2 QM/MM Calculations

As shown in Fig. 7, each of the 10 replicas generated as guess structures undergoes a series of QM/MM calculations, defining a well-established protocol. In this regard, Fig. 5b illustrates the atoms that are considered as the Lys-QM layer during these calculations, corresponding to the full N-methyl retinal chromophore (*i.e*, QM subsystem; 53 atoms) and its covalently linked lysine side-chain of the MM subsystem (i.e., 9 atoms). The QM/MM frontier is treated within a link atom approach (see Fig. 5b), whose position is restrained according to the Morokuma scheme, and is placed across the lysine Cδ–Cε bond (where Cε is a

QM atom). The lysine charges are modified by setting the Cδ charge to zero to avoid hyperpolarization and to redistribute the residual fractional charge on the most electronegative atoms of the lysine, thus ensuring a +1 integer charge of the Lys-QM layer.

The following QM/MM calculations are performed sequentially, using the program [Open]Molcas/Tinker [62–64]:

(1) Geometry optimization at the HF/AMBER/3-21G level.
(2) Geometry optimization at the 2-roots single-state CASSCF(12,12)/AMBER/3-21G level.
(3) Geometry optimization at the 2-roots single-state CASSCF(12,12)/AMBER/6-31G(d) level.
(4) Inclusion of the electron correlation via a single point energy calculation at the 3-roots state-average CASPT2(12,12)/6-31G(d) level.

The sequential optimizations steps 1–3 aim at a more rapid convergence of both the molecular orbitals and geometry, and use “microiterations” that provide quicker convergence, lower energies, and a more realistic description of chromophore-environment interactions [65]. Besides, suitable level shifting values are used during the CASSCF and CASPT2 calculations, to minimize the possibility of convergence failure due to state mixing and intruder state problems. The CASPT2//CASSCF/6-31G(d)/MM treatment [55] has been investigated extensively for photobiological studies and its limitations are well understood. As previously documented [59], the rather small (ca. 3–4 kcal mol^{-1}) error in excitation energy reported in several studies for this level of theory, is somewhat due to error cancellations associated with the limited quality of single-state CASSCF/AMBER/6-31G(d) equilibrium geometries. Therefore, the different properties computed by ARM are expected to be affected by systematic error cancellations. Nevertheless, the main focus of ARM is the ability to reproduce observed trends in vertical excitation energies (i.e., the sign and magnitude of the individual differences concerning experimental data).

As observed in Fig. 7, the final output consists of 10 replicas of equilibrated QM/MM models of the type described in Sect. 3.3 and, for each replica, the vertical excitation energy values between S_0 and the first two singlet excited states S_1 and S_2 is provided.

3.5 Automation Issues

The *original* ARM protocol provides basic, gas-phase and computationally fast QM/MM models for comparative studies, to predict photochemical property trends (e.g., for large arrays of rhodopsin variants) that fit selected sets of experimental data (mainly λ^a_{max} values), within a well-established error bar. However, the general target in the formulation of an automated modeling of rhodopsins is to achieve a protocol featuring the following features for both the input file generator and QM/MM model generator phases:

(1) **transferability**, so as to properly describe rhodopsins with differences in protein sequence (i.e., organism belonging to different life domains and kingdoms; see Sect. 2.1), and different configurations of the *r*PSB;
(2) **documented accuracy**, so as to be able to translate results obtained in silico into hypothesis that can be proved experimentally;
(3) **reproducibility**, so as to be reproduced in any laboratory starting the model building from scratch;
(4) **speed** and **parallelization**, so as to achieve the fast generation of large arrays of rhodopsins variants (i.e., wild-type and mutants) simultaneously;
(5) **automation**, so as to reduce building errors (i.e., human factors) and avoid biased QM/MM modeling.

In order to assess whether or not the original ARM satisfies each of the points (1)–(5) described above, the protocol was benchmarked on the prediction of trends in $\lambda^a_{\max}$ for a limited set of 10 wild-type and 17 mutant rhodopsins [59]. Accordingly, points (1) and (2) were successfully achieved, while points (3), (4) and (5) were accomplished only partially. More specifically, point (1) was achieved since the predicted trends in $\lambda^a_{\max}$ presented a good agreement with experimental data (i.e., error bar of ca. 4.0 kcal mol^{-1}). Point (2) has been recently demonstrated via collaborative experimental and computational studies attempting enhancement of either color tuning [42] and fluorescence [66] properties for microbial rhodopsins, achieving novel applications in optogenetics. Further applications can be found in [43]. The main drawback of the original ARM is that it does not include a computational tool for the automatic, or even semi-automatic, generation of the input files (i.e., no input file generator). Input file generation is, instead, achieved through a manual manipulation of the template structure [59]. Such an input file is based on a X-ray crystallographic structure or comparative model of the protein in PDB (Protein Data Bank) format [67, 68], which contains the information specified in the caption of Fig. 5 and summarized as follows:

(i) the selected monomeric chain structure, including the *r*PSB chromophore, crystallographic/comparative water molecules, and excluding membrane lipids and non functional ions;
(ii) a list of residues forming the chromophore cavity;
(iii) the protonation states for all the internal and surface ionizable amino acid residues;
(iv) suitable external counterions (Cl^{-}/Na^{+}) needed to neutralize both IS and OS protein surfaces.

Due to possible different user choices (e.g., during the placement of IS and OS counter-ions; see Fig. 5a), reproducibility of the results described (point (3)) cannot be guaranteed. In addition, the input preparation required a few hour user's manipulation of the template protein structure (i.e., a skilled user completes the preparation of an ARM input for a new rhodopsin in not less than 3 h and after taking a series of decisions based on their chemical/physical knowledge and intuition). Such limitations, added to the human error factor, represent a serious issue when the target is the generation of hundreds rhodopsin models. Therefore, due to manual interventions of the user in the input generation, speed and parallelization (point (4)) are not guaranteed. Of course, the lacking of an automatic input file generator and the current methodological issues of the QM/MM model generator described below, made the protocol semi-automatic rather than automatic (i.e., also point (5) was not accomplished).

Additionally, as specified by Melaccio et al. [59], the code was written as a series of independent bash-shell scripts that are not interconnected by a general driver. Therefore, as explained in the Supporting Information in [59], for each step of the protocol, the user should execute each script manually and make a series of choices via a command-line assisted tool. This features the QM/MM model generator as a semi-automatic rather than an automatic tool. The following sections will show how each of these problems have been overcome.

4 *a*-ARM: the First Major Update Towards Automation

(Most of the content of the following four sections is reproduced/adapted with permission from [69], copyright 2019 American Chemical Society, while content of Sect. 4.5 is reproduced/adapted with permission from [35], open access under a CC BY license (Creative Commons Attribution 4.0 International License)).

4.1 Methodological Aspects

This section presents the *advanced a*-ARM [69], an updated version of the *original* semi-automatic ARM protocol [59] (Sect. 3) that, as main novelty, features an input file generator phase for either the fully automatic or semi-automatic computer-aided building of the ARM input. This updated version overcomes most of the automation issues of the original ARM, highlighted in Sect. 3.5, by including methodological improvements that lead to more reliable and reproducible QM/MM models.

The first aim of the *a*-ARM model building is to be consistent (in terms of output) with the original protocol, as described in Sect. 3.3 and shown in Fig. 5. Figure 8 shows the general workflow of *a*-ARM, which encapsulates two well-defined and automated phases (panels a and b), hereafter referred to as Phase I and Phase II, respectively. Whereas the latter is substantially the same QM/MM model generator phase reviewed in Sect. 3.3 (i.e., in terms of methodology, although not of implementation), the former is the new input file generator phase.

The combined use of the two phases achieves the automatic building of Ground-state ARM QM/MM models, starting from the structure of a rhodopsin in PDB format, either as PDB code or as a comparative homology model. As observed in Fig. 8, the initial structure is processed by Phase I to obtain the ARM input, which is subsequently processed by Phase II to obtain the ARM QM/MM model (i.e., gas-phase equilibrated optimized S_0 structure) and the predicted average $\lambda^a_{\max}$. As further explained in Ref. [69], the input file generator is implemented as a user-friendly command-line interface, where the researcher

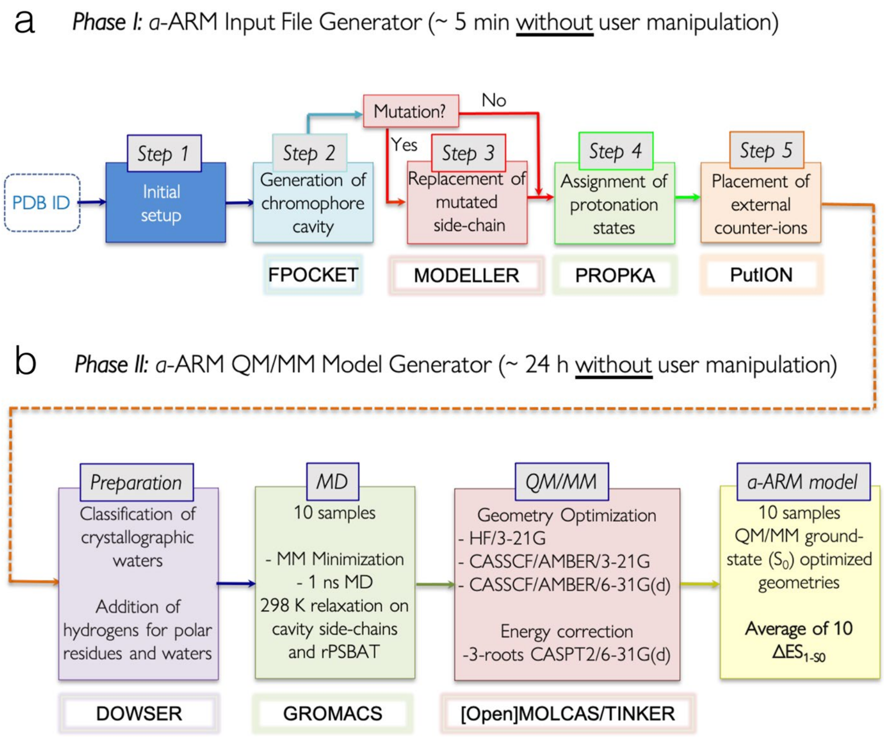

a
Phase I: a-ARM Input File Generator (~ 5 min without user manipulation)
PDB ID
Step 1
Initial setup
Step 2
Generation of chromophore cavity
FPOCKET
Mutation?
No
Yes
Step 3
Replacement of mutated side-chain
MODELLER
Step 4
Assignment of protonation states
PROPKA
Step 5
Placement of external counter-ions
PutION
b
Phase II: a-ARM QM/MM Model Generator (~ 24 h without user manipulation)
Preparation
Classification of crystallographic waters
Addition of hydrogens for polar residues and waters
DOWSER
MD
10 samples
- MM Minimization
- 1 ns MD
298 K relaxation on cavity side-chains and rPSBAT
GROMACS
QM/MM
Geometry Optimization
- HF/3-21G
- CASSCF/AMBER/3-21G
- CASSCF/AMBER/6-31G(d)
Energy correction
-3-roots CASPT2/6-31G(d)
[Open]MOLCAS/TINKER
a-ARM model
10 samples
QM/MM ground-state (S0) optimized geometries
Average of 10
ΔES1-S0

Fig. 8 General workflow of the two phases of the *a*-ARM rhodopsin model building protocol. **a** Input file generator phase. **b** QM/MM model generator phase

interacts with the program by typing information directly in the computer terminal, without the need to manipulate text files or visualize chemical structures, as previously necessary in the original “manual” strategy. Refs. [69] (see Section S1 in the SI) and [43] (see Section 3.1 in this reference) report a detailed description of both manual (i.e., *original* ARM) and automatic procedures used to pursue steps 1–5 of Fig. 8a, with particular emphasis on the improvements achieved with *a*-ARM, as well as in its higher level of automation. Figure 9 presents an overview of such improvements.

One of the most remarkable features of the new protocol is that, given the options (i.e., parameters) selected in steps 1–5 of phase I (Fig. 8a), *a*-ARM allows either the automatic or semi-automatic computer-aided production of the ARM input. Accordingly, *a*-ARM is sub-divided in *a*-ARM$_{\text{default}}$ (see Section 3.1 of Ref. [69]) and *a*-ARM$_{\text{customized}}$ (see Section 3.2 of Ref. [69]) approaches. The former refers to a fully automatic input generation, which uses default parameters as suggested by the code (i.e., chain, rotamers or side-chain conformations, pH, protonation states, residues forming the chromophore cavity), whereas the latter allows the computer-aided customization of some of such parameters when the default choices are not suitable.

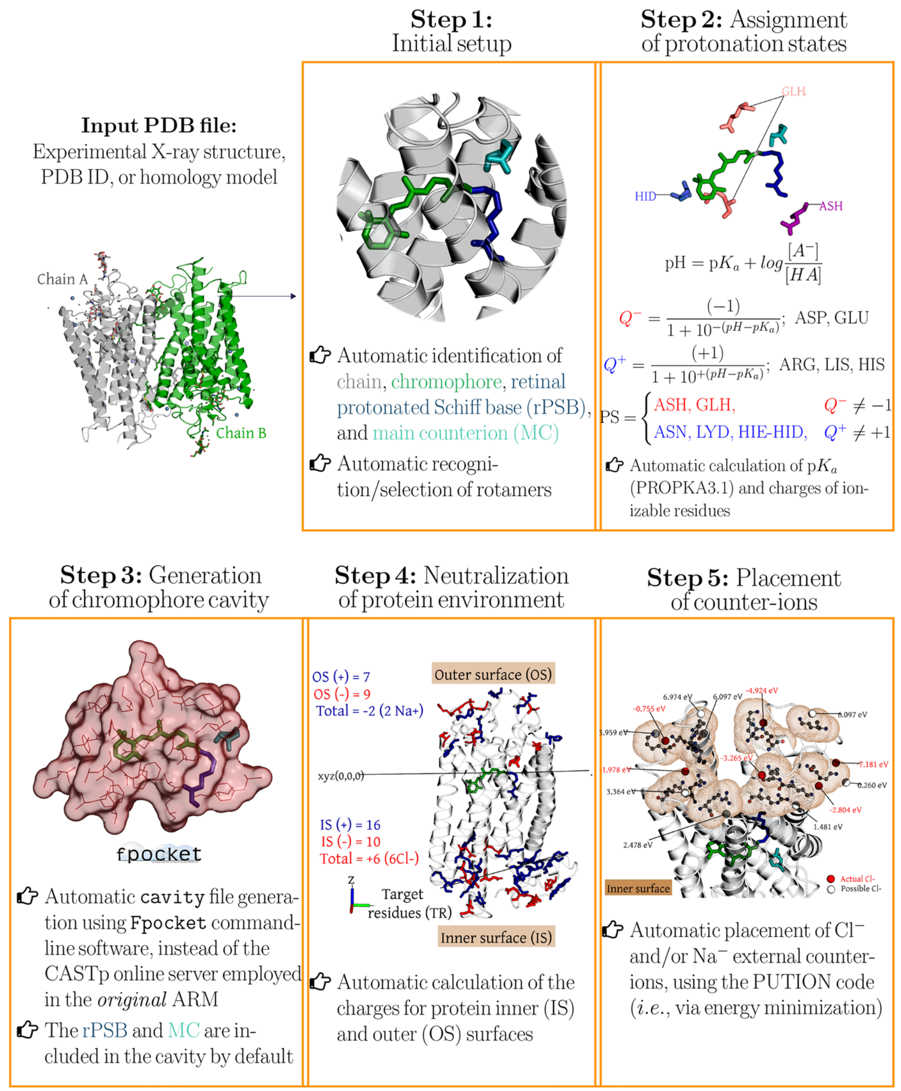

Step 1: Initial setup
Step 2: Assignment of protonation states
Input PDB file: Experimental X-ray structure, PDB ID, or homology model
Chain A
Chain B
Automatic identification of chain, chromophore, retinal protonated Schiff base (rPSB), and main counterion (MC)
Automatic recognition/selection of rotamers
GLH
HID
ASH
$\mathrm{pH} = \mathrm{p}K_a + log\frac{[A^-]}{[HA]}$
$Q^- = \frac{(-1)}{1+10^{-(pH-pK_a)}}$; ASP, GLU
$Q^+ = \frac{(+1)}{1+10^{+(pH-pK_a)}}$; ARG, LIS, HIS
$\mathrm{PS} = \begin{cases} \text{ASH, GLH,} & Q^- \neq -1 \\ \text{ASN, LYD, HIE-HID,} & Q^+ \neq +1 \end{cases}$
Automatic calculation of pK_a (PROPKA3.1) and charges of ionizable residues
Step 3: Generation of chromophore cavity
Step 4: Neutralization of protein environment
Step 5: Placement of counter-ions
fpocket
Automatic cavity file generation using Fpocket command-line software, instead of the CASTp online server employed in the *original* ARM
The rPSB and MC are included in the cavity by default
OS (+) = 7
OS (-) = 9
Total = -2 (2 Na+)
Outer surface (OS)
xyz(0,0,0)
IS (+) = 16
IS (-) = 10
Total = +6 (6Cl-)
z
Target residues (TR)
Inner surface (IS)
Automatic calculation of the charges for protein inner (IS) and outer (OS) surfaces
6.974 eV
6.097 eV
-4.924 eV
-0.755 eV
3.959 eV
6.097 eV
-3.265 eV
-1.978 eV
7.181 eV
3.364 eV
0.260 eV
-2.804 eV
1.481 eV
2.478 eV
Inner surface
● Actual Cl-
○ Possible Cl-
Automatic placement of Cl$^-$ and/or Na$^-$ external counter-ions, using the PUTION code (*i.e.*, via energy minimization)

Fig. 9 Overview of the most relevant features of the input file generator, introduced in the *a*-ARM version of the protocol. Methodological and automation improvements achieved with the input file generator, in terms of: initial setup; automatic strategy adopted for the assignment of protonation states for ionizable residues; replacement of the software (i.e., fpocket instead of CASTp) for the automatic generation of the chromophore cavity; automatic approach to define the charge of the IS and OS surfaces and automatic counterion placement based on energy minimization

The customized approach is used in cases where the default choices produce outlying models that fail to reproduce trends in absorption properties. Several examples are presented in Refs. [35, 43, 69]. For instance, Fig. 10 illustrates how the customization, in terms of either selection of side-chain conformations and protonation states, is achieved for the case of the microbial rhodopsin

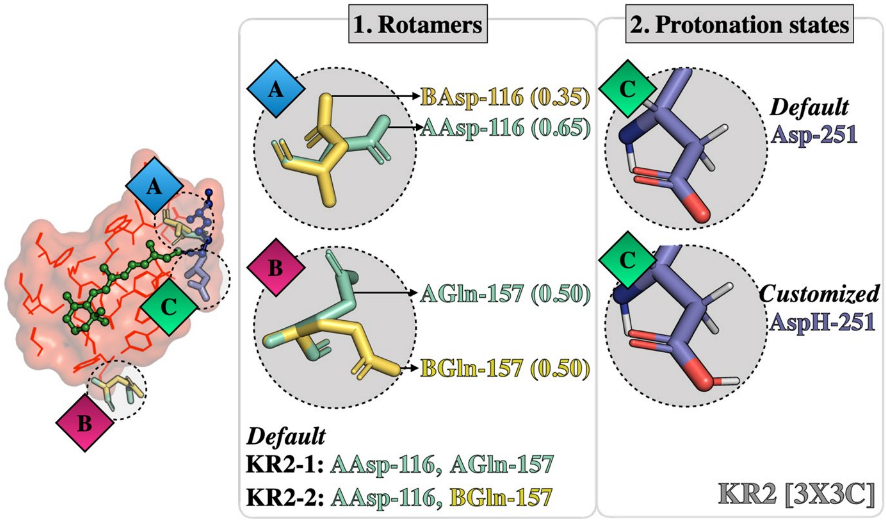


Fig. 10 Default and customized *a*-ARM models for Krokinobacter rhodopsin 2 from *Krokinobacter eikastus* (KR2) [PDB ID 3X3C [70]]. Left: conformational (the occupancy factor of the rotamers Asp-116 and Gln-157 are presented in parentheses). Right: ionization state variability

Krokinobacter rhodopsin 2 from *Krokinobacter eikastus* (KR2) [35, 69, 70]. As observed, in the 3X3C X-ray structure [70], the residue Asp-116, considered as the main counterion (MC) of the *r*PSB, exhibits two side-chain conformations, namely, AAsp and BAsp, labeled with occupancy numbers 0.65 and 0.35, respectively. Moreover, the residue Gln-157 that is part of the environment subsystem (i.e., fixed during the QM/MM calculations) presents two conformations (AGln and BGln) both with 0.5 occupancy. According to the occupancy numbers, *a*-ARM$_{\text{default}}$ selects the rotamer AAsp-116 and generates two models relative to Gln-157: KR2-1, which includes AAsp-116 and AGln-157, and KR2-2, which includes AAsp-116 and BGln-157. The computed ΔE^{a}_{S1-S0} for both default models, presented below in Fig. 11, features an error of about 15.0 kcal mol^{-1} with respect to experimental data. Since the default models are unable to provide values inside the experimental trend, the *a*-ARM$_{\text{customized}}$ approach is necessary. As shown in the right panel of Fig. 10, such customization is performed through a more rational assignment of the protonation states of the two aspartic acid residues forming the counterion complex of the *r*PSB, namely Asp-116 and Asp-251 [35]. The default model predicts that both aspartic acids are negatively charged. However, as further discussed in Refs. [35, 43, 69], the presence of these two negative charges would outbalance the single positive charge of the *r*PSB, generating the large blue-shifted effect mentioned above. Accordingly, in the customized model the secondary counterion (SC) Asp-251 is, instead, protonated (i.e., neutral) to counterbalance the charge in the vicinity of the *r*PSB. As can be seen in Fig. 11, such customization provides a model with a small error bar of about 1.5 kcal mol^{-1}. An updated ARM model of KR2 was recently reported [35]. Although such a model was constructed starting from a different

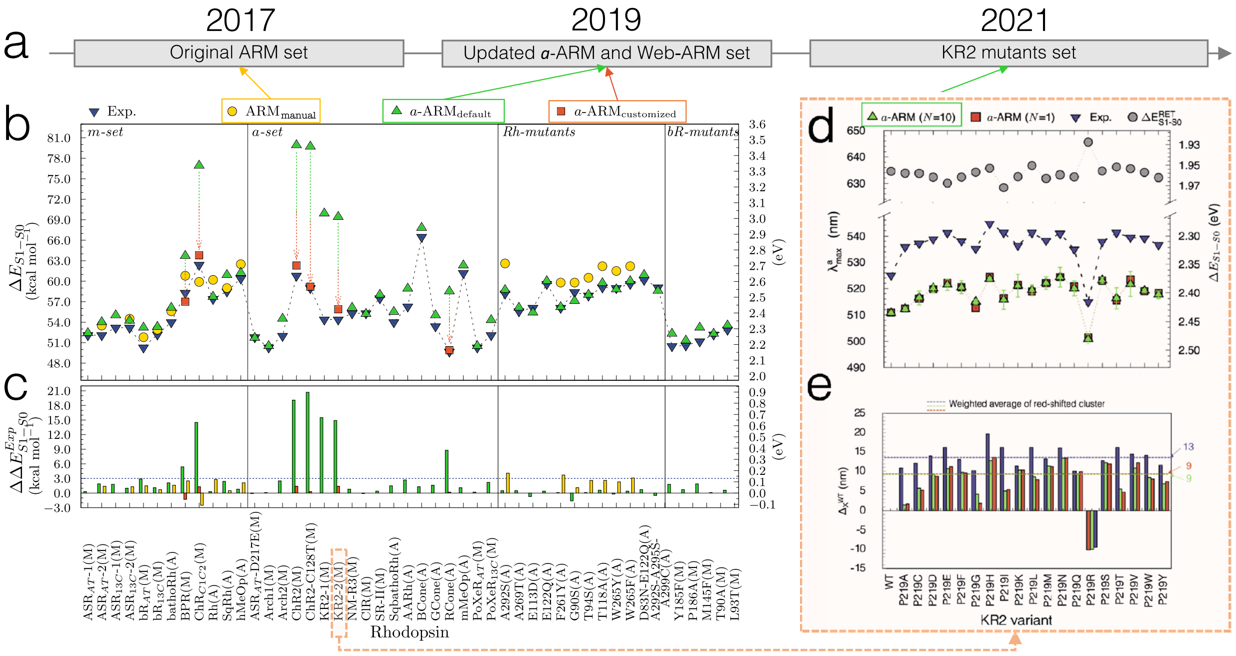


Fig. 11 Evolution and benchmarking of the ARM protocol over time, in terms of reproduction of experimental trends in λ^a_{max}. **a** Timeline for the benchmark sets used for testing each version of ARM over the years. **b** Comparison between vertical excitation energies (ΔE_{S1-S0}) computed with either *a*-ARM$_{default}$ (up green triangles) or *a*-ARM$_{customized}$ (red squares) [69], and ARM$_{original}$ [59] (yellow circles) and experimental data (down blue triangles), along with the **c** differences between computed and experimental $\Delta E_{S1\text{-}S0}$ ($\Delta\Delta E^{Exp}_{S1\text{-}S0}$). The *m*-set corresponds to wild-type (WT) rhodopsins forming the original benchmark set for the ARM protocol; *a*-set introduces new rhodopsins to the benchmark set of the *a*-ARM protocol; *Rh − mutants* set contains mutants of bovine Rhodopsin (Rh) belonging either to the benchmark set of the ARM or *a*-ARM protocols (see Refs. [59, 69]); and *bR − mutants* set contains mutants of bR evaluated with the Web-ARM interface (see Sect. 5 and Ref. [74]). **d** Comparison between ΔE_{S1-S0} computed with *a*-ARM, average value (up green triangles) and representative seed (red squares), and experimental data (down blue triangles) for WT and P219X mutants of KR2 [6REW], along with **e** the differences between these values for each mutant with respect to the WT (see Ref. [35]). Adapted with permission from [69]. Copyright 2019 American Chemical Society and [35], open access under a CC BY license (Creative Commons Attribution 4.0 International License)

template X-ray structure [PDB ID 6REW [1]], the same customized setup for the protonation states of the counterion complex of the *r*PSB (i.e., deprotonated Asp-116, protonated Asp-251) was found. Remarkably, such *a*-ARM $_{\text{customized}}$ model was used as a starting structure for modeling a set of 19 mutants and for reproducing not only experimental trends in ΔE^{a}_{S1-S0}, but also giving further insights into the mechanism of color tuning in the position Pro-219 of KR2 [35].

Notice that in the *a*-ARM benchmark set (Fig. 11) a similar large blue-shifting has been observed in ΔE^{a}_{S1-S0} for all *a*-ARM$_{\text{default}}$ models that predict two negative charges near the *r*PSB (see Tables S2 and S3 in Ref. [69]). The customization procedure (see below) to obtain a ΔE^{a}_{S1-S0} within the error bar of the protocol, always implies the neutralization of one of these two charges. This is, in fact, one particularity of the simplified structure/scheme of the *a*-ARM models.

In general, protonation state assignment for ionizable residues remains a basic issue for current QM/MM protein modeling. No robust method is available that guarantees a correct choice of a pKa value, due to the complexity of the protein environments and its interconnected local effects. Furthermore, a residue may be present as an equilibrium between ionized and not-ionized forms, hence carrying only a partial charge. These issues are a current research matter [35, 71–73].

Despite the difficulties mentioned above, *a*-ARM adopts the following guiding customization procedure. As shown in the latter example, customized ARM QM/MM models can be constructed according to well-defined operations that can be easily replicated. Ref. [69] proposes to focus on the selection of the ionization states and side-chain conformations only. Accordingly, the customization of the protonation states involves three phases: (1) at pH > 6 the ionization states are modified by setting the pH to 5.2 in step 4 (see Fig. 8b); (2) the protonation state of the main and secondary counterions of the rPSB are checked, and if the analysis shows them both ionized the secondary counterion is neutralized; (3) in case the model generated in step (2) does not reproduce the experimental absorption maxima, then the secondary and main counterion ionization states are exchanged (see also the Supporting Information of Ref. [35] for further details on, e.g., the pH value choice). Note that in QM/MM modeling, it is a common practice to evaluate the protonation states of the *r*PSB counterion complex by looking, as a guide, at the reproducibility of the experimental $\lambda^{a}_{\max}$ (see, for instance, Refs. [35, 71–73])

Indeed, the novelty of the default and customized approaches is that, regardless of the user or computational facility, reproducible inputs, and consequently reproducible ARM QM/MM models, are guaranteed when the same parameters are used. This represents an advancement with respect to the original version since it allows for the models to be reproduced in any laboratory and by any user, even when starting building an ARM QM/MM model from scratch (see point (3) of Sect. 3.5).

4.2 Software Implementation Aspects

The computational implementation of both the input file generator and the QM/MM model generator phases as Python-based, modular codes boosted the building of the `PyARM` software package that will be introduced in Sect. 6. Moreover, their ease of transferability, allowed their use behind the Web-ARM web page, which will be described in Sect. 5.

Furthermore, although *a*-ARM can presently build only rhodopsin models (i.e., with natural retinal), it provides a template for the development and generation of an automatic QM/MM building strategy for other, more general, systems such as rhodopsins incorporating artificial (i.e., unnatural) chromophores. This is straightforwardly achieved given the modular architecture of the `PyARM` package and the fact that any chromophore can be treated when using the appropriate force field.

Finally, it is worth stressing that the new protocol achieves all the features (1)–(5) described in Sect. 3.5, overcoming the automation limits of the original version.

4.3 Benchmark, Validation and Application Aspects

Figure 11 shows the current validation of the *a*-ARM protocol through the prediction of trends in $\lambda^a_{\max}$, performed using a benchmark set of 44 animal and microbial rhodopsin variants (i.e., 25 wild type and 19 mutants) that come from different organism and are phylogenetically diverse [69, 74] (see Fig. 11b). The full benchmark set features values ranging from 458 nm (62.4 kcal mol^{-1}, 2.71 eV) to 575 nm (49.7 kcal mol^{-1}, 2.15 eV). Such a relatively wide range provides information on the method accuracy, while the rhodopsin diversity provides information on the transferability and general applicability of the generated models. Figure 11b, c are divided in four different regions: *m*-set, *a*-set, $Rh - mutants$ set, and $bR - mutants$ set. The *m*-set and $Rh - mutants$ set are used to compare the performance of original ARM and *a*-ARM versions, while the remaining sets focus exclusively on the performance of *a*-ARM.

The *a*-ARM$_{\text{default}}$ approach proved to be capable of reproducing the ΔE^a_{S1-S0} values for 86% of cases (38/44), with an error lower than 4.0 kcal mol^{-1} (0.13 eV), whereas the other 14% cases were successfully obtained with the *a*-ARM$_{\text{customized}}$ approach (i.e., changing the side-chain conformation and/or protonation states pattern). A detailed description of the customization procedure used for reproducing the experimental $\lambda^a_{\max}$ values of KR2, BPR, ChR2-C128T, ChR2 and ChR$_{\text{C1C2}}$ is provided in Section 3.2 of Ref. [69] and Section 3.3 of Ref. [43].

Recently, *a*-ARM was applied to (1) reproduce the $\lambda^a_{\max}$ of WT KR2 and 19 mutants [35] (see Fig. 11a) and (2) to gain further insights into the origin of red- or blue-shifting. As observed in Fig. 11d, e, the performance of *a*-ARM reported for the benchmark set (see above), is maintained when modeling a set of mutants that feature $\lambda^a_{\max}$ spanning a red-to-blue range going from 545 nm (54.5 kcal mol^{-1}, 2.36 eV) to 515 nm (57.0 kcal mol^{-1}, 2.47 eV). Furthermore, *a*-ARM demonstrated to be useful to generate models that reproduce blue- or red-shifting effects observed experimentally (see Ref. [35]). The final S$_0$ optimized equilibrium structures can be then used for further excited-state optimizations (see Sect. 6.2). Indeed, some of the ARM QM/MM models produced have been used as input for sophisticated constant-pH dynamics [75], the simulation of one-/two-photon absorption spectra [40], and in combined computational/experimental studies on color tuning possibilities in KR2 [35].

Accordingly, and as stated in Sect. 3.2, we claim that the *a*-ARM protocol, in its current version, does not represent a predictive tool, but rather is designed to produce models for rhodopsins, which structure was obtained from either X-ray crystallography or comparative modeling, useful for reproducing and explaining the origin of trends in spectroscopic/photochemical properties (e.g., between sequence variability and function) appearing from sets of experimental data.

4.4 Limitations and Pitfalls of *a*-ARM

Despite the encouraging outcome of the photochemical studies based on *a*-ARM (as previously mentioned), additional work is necessary to generate a tool that can be systematically applied to larger arrays of rhodopsins. The following main issues, in part anticipated above, have to be tackled to improve the input file generator phase:

- Assignment of the protonation states: there are two main aspects that limit the confidence in the automation of the ionizable state assignment described in Refs. [43, 69]. The first is that, due to the fact that the information provided by PROPKA [76] is approximated, the computed pK_a^{Calc} value may, in certain cases, be not sufficiently realistic. The second aspect regards the assignment of the correct tautomer of histidine. *a*-ARM uses as default the histidine dipeptide (HID) tautomer (deprotonated δ-nitrogen) for the automatic assignment, or allows the user to choose between the three possible tautomers for a “not-automated selection. Therefore, when possible, the user should collect the available experimental data and/or inspect the chemical environment of the ionizable residues including the histidines, and propose the appropriate tautomer [69]. Alternatively, one has to systematically examine all sensible choices, which may not always be feasible.
- Automatic construction of comparative models: since rhodopsin structural data are rarely available, it would be important to investigate the possibility of building, automatically, the corresponding comparative models. With such an additional tool, one could achieve a protocol capable of producing QM/MM models starting directly from the constantly growing repositories of rhodopsin amino acid sequences. This target is currently pursued in our laboratory.
- Automatic prediction of side-chain conformation for mutants: recent efforts have been directed to achieve a successful technology for systematically predicting mutant structures, which provides a superior level of accuracy of the *a*-ARM models than that proposed in Ref. [69].

  More specifically, the mutations routine that used a backbone-dependent rotamer library (i.e., SCWRL4 [77]) was replaced by a software based on comparative modeling (i.e., MODELLER [78]) (see Fig. 8b). The description of the new approach as well as an example that illustrates its effectiveness for mutating a specific position with each of the 20 essential amino acids, are provided in Ref. [35] and summarized in Sect. 4.5.

- Insufficient description of possible cavity rearrangements after mutation: the updated procedure described in Ref. [35] for modeling the side-chain conformation (see point above) comprises a short MD, where the introduced side-chain is allowed to relax, whereas the rest of the cavity residues, water molecules, chromophore and protein environment remain fixed at the crystallographic/comparative structure (see Supplementary Note 13 in Ref. [35]).

  Notwithstanding the following, more sophisticated MD step related to the cavity residues (see Sect. 3.4.1), a proper description of the impact of the new side-chain on the protein environment is lacking, due to a not sufficient description of possible local steric/electronic rearrangements of those residues of the chromophore cavity surrounding the mutated one.
- Mutations only allowed in the chromophore cavity: currently, *a*-ARM only allows mutations of residues that belong to the chromophore cavity sub-system, as well as backbone relaxation is not allowed. The latter is to ensure that, during the QM/MM model generator phase, the geometry of the new modeled side-chain as well as the sidechain of its neighbors (belonging to the chromophore cavity) can be readjusted during the 1 ns GROMACS MD step, while assuming that the general structure of the protein is conserved.
- Lack of a predictive tool for mutants generation: the fact that the mutants generator relies on the use of experimental data to select the correct rotamer limits the usability of the protocol, which cannot be considered as a predictor tool.

4.5 Recent Updates and Improvements

In silico modeling of point mutations in proteins relies on the selection of a robust methodology for the prediction of the side-chain conformation of the replaced amino acid [77, 79–90]. Both *original* [59] and *advanced* [43, 64, 69, 91] versions of the ARM protocol use the software SCWRL4 [77] to predict the side-chain conformation of the mutated residues. This approach is based on backbone-dependent rotamer libraries (from public databases of experimentally resolved protein structures), and was found adequate for the production of single, double and triple point rhodopsin mutants. This was demonstrated by studies carried out by some of the authors on mutants of bovine rhodopsin (Rh) [59, 69], Anabaena Sensory rhodopsin (ASR) [42, 59], bacteriorhodopsin (bR) [74] and KR2 rhodopsins [92].

Recently, some of the authors reported the first attempt to use *a*-ARM model building to systematically and exhaustively mutate a single residue [35]. More specifically, they attempted, unsuccessfully, to perform single point mutations of KR2 at the P219 location near the *β*-ionone ring of the *r*PSB via SCWRL4 modelling. Despite the encouraging results reported in Ref. [42] for the cases P219A, P219G and P219T, the authors found that, for larger side-chains SCWRL4 generated conformers sterically clashing with either the *r*PSB or neighboring amino acids.

After examining the tools available for side-chain predictions (see, for instance, Ref. [79]) and evaluating them in terms of performance and accessibility as command-line tools, the authors of Ref. [35] modified the mutations routine (see Section 2.2.5. in Ref. [69]) by substituting SCWRL4 with MODELLER [78]. This alternative approach allows the production of mutants suitable for the prediction of absorption wavelengths in either an automatic or a computer-aided semi-automatic fashion.

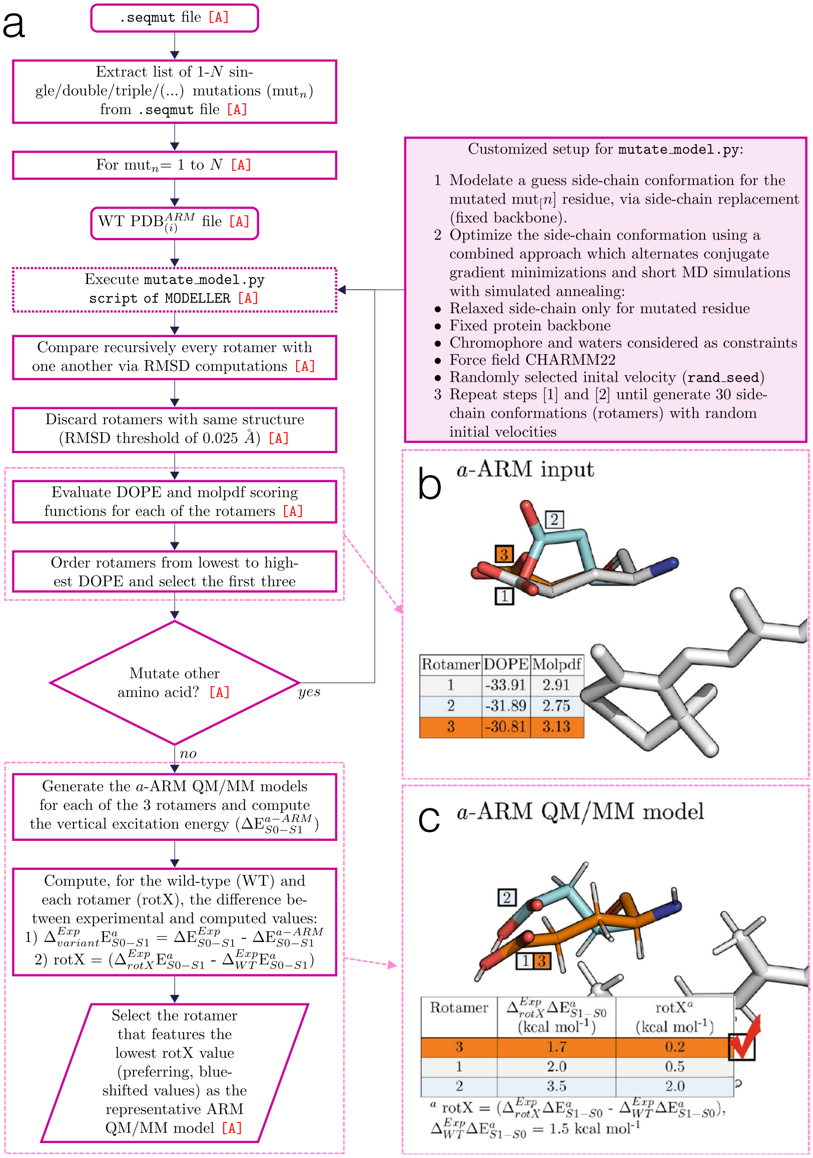
Figure 12a illustrates the general workflow of the proposed subroutine, which replaces Step 3 of the input file generator phase of *a*-ARM (see Fig. 8a). At input level, each point mutation is generated via a customized version of the `mutate_model.py` routine implemented in Modeller, where the conformation of the modeled side-chain is optimized using a conjugate gradient method, and refined using a short MD ($^{\mathrm{mod}}$).

Briefly, as reported in Ref. [78], the `mutate_model.py` script has been designed to model point mutations via side-chain replacement in a fixed environment, assuming that single mutations do not generally determine deep conformational changes of the protein backbone. Accordingly, and consistently with the structurally "conservative" approach of the *a*-ARM protocol (see Sect. 3.3), our methodology replaces only the side-chains of the mutated residues keeping the backbone atoms at fixed positions. In order to sample the conformational space of a mutated residue more extensively. and evaluate its effect on the vertical excitation energy ($\Delta \mathrm{E}^{a}_{S1-S0}$), the new customized setup produces 30 rotamers of the same mutated side-chain by providing the script with different initial seeds (i.e., initial velocities) for the MD$^{\mathrm{mod}}$ run. The obtained rotamer structures are compared with each other, in terms of root mean square deviation, and discarded if found less than 0.025 Å from another. Although not particularly efficient, this procedure allows for the quick selection of a set of non-redundant rotamers for a single mutant, which are evaluated using the scoring function discrete optimized protein energy (DOPE) [93], implemented by MODELLER, and ranked from lowest to highest. The ARM input for the three highest DOPE scored mutated side-chain rotamers is completed by phase I of the *a*-ARM protocol, and their ARM QM/MM models are produced using phase II (see Fig. 8). The corresponding computed $\Delta \mathrm{E}^{a}_{S1-S0}$ is then used to evaluate the performance of different rotamers of the mutated side-chain in reproducing the experimental trend in line with the WT, leading to the selection to the conformer (rotamer) that better agrees with experimental data. Figure 12, panels b and c illustrate an example of the procedure for selecting a rotamer from three evaluated models. Further deatils can be found in Supplementary Note 13 in the supporting information of Ref. [35].

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 QM/MM models.

5 Web-ARM, a Web-Based Interface to ARM

(Most of the content of this section is reproduced/adapted with permission from [74]. Copyright 2020 American Chemical Society).


a
.seqmut file [A]
Extract list of 1-N single/double/triple/(...) mutations (mut$_n$) from .seqmut file [A]
For mut$_n$= 1 to N [A]
WT PDB$^{ARM}_{(i)}$ file [A]
Execute mutate_model.py script of MODELLER [A]
Compare recursively every rotamer with one another via RMSD computations [A]
Discard rotamers with same structure (RMSD threshold of 0.025 Å) [A]
Evaluate DOPE and molpdf scoring functions for each of the rotamers [A]
Order rotamers from lowest to highest DOPE and select the first three
Mutate other amino acid? [A]
yes
no
Generate the a-ARM QM/MM models for each of the 3 rotamers and compute the vertical excitation energy (ΔE^{a-ARM}_{S0-S1})
Compute, for the wild-type (WT) and each rotamer (rotX), the difference between experimental and computed values:
1) $\Delta^{Exp}_{variant}E^{a}_{S0-S1} = \Delta E^{Exp}_{S0-S1}$ - ΔE^{a-ARM}_{S0-S1}
2) rotX = ($\Delta^{Exp}_{rotX}E^{a}_{S0-S1}$ - $\Delta^{Exp}_{WT}E^{a}_{S0-S1}$)
Select the rotamer that features the lowest rotX value (preferring, blue-shifted values) as the representative ARM QM/MM model [A]
b
a-ARM input
2
3
1
c
a-ARM QM/MM model
2
1
3

Customized setup for mutate_model.py:

1 Modelate a guess side-chain conformation for the mutated mut$_{[n]}$ residue, via side-chain replacement (fixed backbone).
2 Optimize the side-chain conformation using a combined approach which alternates conjugate gradient minimizations and short MD simulations with simulated annealing:
- Relaxed side-chain only for mutated residue
- Fixed protein backbone
- Chromophore and waters considered as constraints
- Force field CHARMM22
- Randomly selected inital velocity (rand_seed)
3 Repeat steps [1] and [2] until generate 30 side-chain conformations (rotamers) with random initial velocities

Rotamer	DOPE	Molpdf
1	-33.91	2.91
2	-31.89	2.75
3	-30.81	3.13

Rotamer	$\Delta^{Exp}_{rotX}\Delta E^{a}_{S1-S0}$ (kcal mol^{-1})	rotX^a (kcal mol^{-1})
3	1.7	0.2 ✓
1	2.0	0.5
2	3.5	2.0

^{a} rotX = ($\Delta^{Exp}_{rotX}\Delta E^{a}_{S1-S0}$ - $\Delta^{Exp}_{WT}\Delta E^{a}_{S1-S0}$), $\Delta^{Exp}_{WT}\Delta E^{a}_{S1-S0}$ = 1.5 kcal mol^{-1}

Fig. 12 General workflow of the novel side-chain generator. **a** Modified routine for the mutants generator of *a*-ARM, based on Modeller. This procedure is used to model, e.g., the side-chain of the E219 residue of the KR2 rhodopsin, as shown in panels **b**, **c**. **b** First, the discrete optimized protein energy (DOPE) and molpdf scoring functions for all the possible rotamers are evaluated and the three best values are ranked. **c** Then, the a-ARM QM/MM model for each rotamer is generated and the rotamer model featuring the lowest difference in vertical excitation energy with respect to experimental data (rotamer 3) is selected Adapted with permission from [35], open access under a CC BY license (Creative Commons Attribution 4.0 International License)

5.1 Interface Features

The *a*-ARM protocol (described in Sect. 4) and its most updated version `PyARM`, a Python-based software package (that will be presented in Sect. 6) represent an easy-to-use command-line interface directed to users (i.e., researchers, undergraduate students) familiar with the Linux environment. The latter is not due to a fact of usability, but rather to the technically complex initial setup of the package, since it requires the prior installation of several software and python dependencies. Moreover, it is recommended to install the `PyARM` package in a high-performance computer cluster, which is usually required to run the underlying computationally intensive tasks (i.e., MM, MD, QM/MM), rather than on a local personal machine.

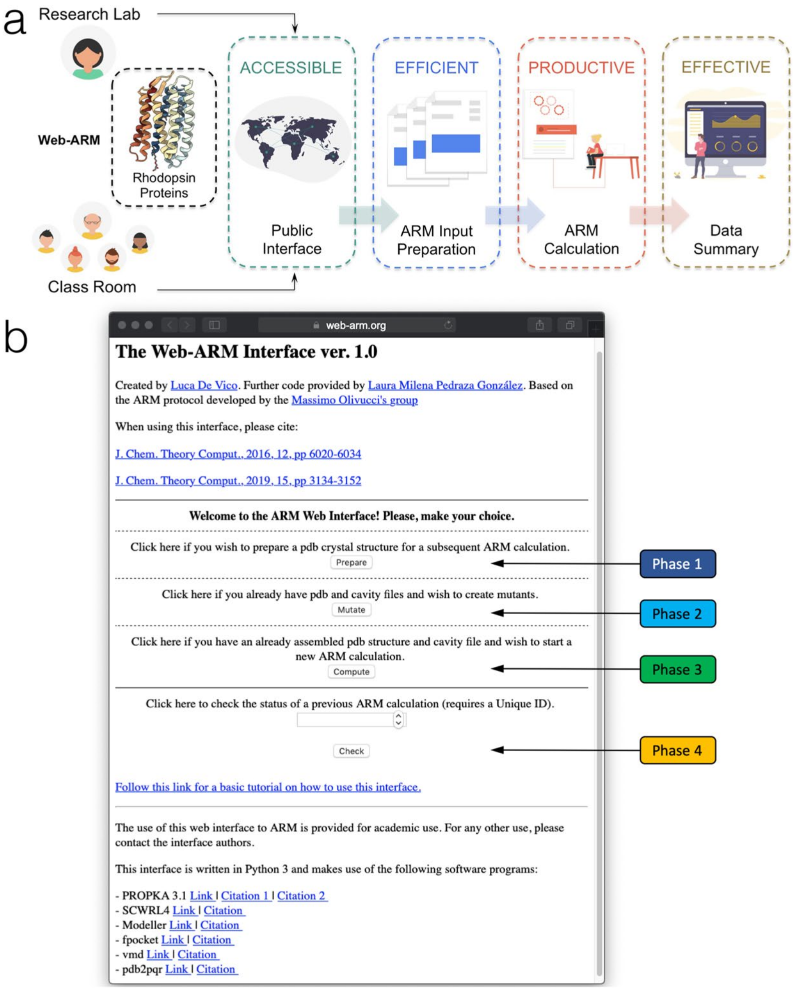
The use of a user-friendly interface accessible through the web and together with the provider computational resources, avoids dealing with complicate installations and the need for local computer facilities. This would mean accessing a simple, computationally fast and automated construction and analysis of rhodopsin QM/MM models, fit for an interdisciplinary community that is interested mostly in actual applications, rather than methodological development.

Accordingly, Ref. [74] reports the web version of the *a*-ARM protocol, that is the Web-ARM interface. Web-ARM is a user-friendly interface written using Python 3, available at the following address: web-arm.org. Therefore, the potential user only needs to use an up-to-date browser on any operating system (i.e., Linux, macOS, Windows, Android, iOS) and platform to take advantage of the *a*-ARM protocol.

In order to generate an ARM QM/MM model, Web-ARM goes through the four phases explained in Section 2.1 of Ref. [74], and illustrated in Fig. 13. As described for the command-line version, the procedure starts with the initial structure of the rhodopsin variant and finishes with the generation of the S_0 equilibrium geometry along with the calculations on absorption properties. During such a procedure, the interface gives the user enough flexibility to generate either *a*-ARM $_{\text{default}}$ or *a*-ARM $_{\text{customized}}$ inputs (see Sect. 4.1), the former automatically and the latter by modifying some of the default choices. This is made on top of the implementation of the input file generator inside the framework of the web interface. Then, the so-generated ARM input is used to compute a QM/MM model, by using the QM/MM model generator. The Web-ARM internal driver takes care of performing all the necessary steps, as well as submitting the calculations to the dedicated computational facilities.

One feature of the interface is that, once a QM/MM model is generated, the user is provided with a summary of all the relevant data (i.e., energetics, oscillator strengths), along with a downloadable file (in compressed format) containing the major output files. Further information, and a complete walk-through, are provided in a Tutorial that can be accessed/downloaded from the Web-ARM main web page.

Web-ARM is intended as both a research, as well as a teaching tool. Ref. [74] shows that the interface can systematically screen rhodopsin variants, and thus obtain a qualitative check prior to, e.g., an experimental study. The interface can also be used successfully in teaching and learning activities, e.g., to introduce students to the idea of QM/MM models and corresponding computed data. Therefore, Web-ARM is envisioned as a tool used in teaching and training, as well as by non-experienced users, as previously noted, mostly for bulk production. However,



The Web-ARM Interface ver. 1.0

Created by Luca De Vico. Further code provided by Laura Milena Pedraza González. Based on the ARM protocol developed by the Massimo Olivucci's group

When using this interface, please cite:

J. Chem. Theory Comput., 2016, 12, pp 6020-6034

J. Chem. Theory Comput., 2019, 15, pp 3134-3152

Welcome to the ARM Web Interface! Please, make your choice.

Click here if you wish to prepare a pdb crystal structure for a subsequent ARM calculation.

Click here if you already have pdb and cavity files and wish to create mutants.

Click here if you have an already assembled pdb structure and cavity file and wish to start a new ARM calculation.

Click here to check the status of a previous ARM calculation (requires a Unique ID).

Follow this link for a basic tutorial on how to use this interface.

The use of this web interface to ARM is provided for academic use. For any other use, please contact the interface authors.

This interface is written in Python 3 and makes use of the following software programs:

- PROPKA 3.1 Link | Citation 1 | Citation 2
- SCWRL4 Link | Citation
- Modeller Link | Citation
- fpocket Link | Citation
- vmd Link | Citation
- pdb2pqr Link | Citation

Fig. 13 General overview of the Web-ARM interface. **a** Main features and **b** home page of the Web-ARM interface Adapted with permission from [74]. Copyright 2019 American Chemical Society

also an experienced computational chemist can take advantage of the web interface, to produce rhodopsin QM/MM models in a standardized manner, being aware of the documented accuracy and rate of success. Of course, one, possibly very useful, application of such a model is to provide high quality guesses for more sophisticated subsequent calculations, e.g., as a starting substrate to which apply further, high-level refinement methods.

In conclusion, by using Web-ARM both junior researchers and trainees will be able to perform meaningful QM/MM calculations focusing on the underling research targets, methodological concepts, and data analysis, while remaining confident that the calculations are internally consistent.

5.2 Limitations and Future Development of Web-ARM

- Limited computational resources: before using the Web-ARM interface, the user is asked to provide an email address to be registered into our database. Registered users are allowed to build as many concurrent ARM QM/MM models as wished (default 10). However, given to the present threshold in the host available computational resources, guest users are allowed to build only one ARM QM/MM model model at a time on the developer's dedicated resources.
- Technical issues: the Tutorial of the Web-ARM interface reports on possible errors or issues in the execution of the interface, and how to solve or avoid them.
- Current and future implementation: presently, the capability of the Web-ARM interface is limited to the construction of ground-state models. Future work will implement inside the interface all of the features of the `PyARM` software package, illustrated in Fig. 14.

6 PyARM

6.1 Package Description

`PyARM` is a user-friendly, open-source Python-based software tool designed to facilitate the systematic, reproducible and congruous QM/MM modeling/analysis of photoexcited states of rhodopsin proteins. More specifically, `PyARM` is a development platform that implements high-level algorithms associated to specialized QM/MM protocols, under the framework of the *a*-ARM protocol [43, 59, 64, 69, 74, 91].

The package structure represents a collection of hierarchical instances here defined as: scripts, basic low-level functions, general high-level functions, modules, drivers and templates. Table 1 provides a technical definition of each of these terms.

Figure 14 depicts the general overview of `PyARM`, showing how all components of the package are modular, i.e., independent of the context in which they are used. The characteristics of the drivers and package create an application-programming tool, capable of making complex workflows/protocols available to users with minimal programming knowledge. Indeed, the inclusion of new protocols (i.e., modules and drivers) can be achieved easily with only a minimal modification of the code. In this regard, the package includes a programming interface for parsing user input, and retrieving and storing specific data. Therefore, a user can simply use a module to perform a given type of application (e.g., perform a geometry optimization) or use a driver to launch a protocol connecting different modules and, thus, executing more complex applications (e.g., generate a S_0 or S_1 QM/MM model, locate a fluorescent rhodopsin, compute an absorption band, etc.). Nevertheless, an experienced user can also create a new driver, capable of performing a new type of application/analysis based on *a*-ARM models and, therefore, within the limits of their accuracy [59, 69, 74].

Furthermore, the modular framework of the package, which was envisioned initially for the study of natural rhodopsins, presents a flexible architecture

Table 1 Different instances of the PyARM framework

Instance	Definition
Script	Python file that is intended to be executed directly. This means that scripts will often contain code written outside the scope of any classes or functions
Basic low-level function	Basic, importable function that performs a single, simple and generic task. Example: Copy/move files, create folders, calculate energy gaps, generate input file
General high-level function	General function that makes use of basic low-level functions to perform a more complex task. Can be used either as importable or as stand-alone by parsing arguments in the terminal. Example: quantum chemistry calculation
Module	Script that makes use of both basic low-level and general high-level functions to perform a complex task (i.e., semi-protocol) and that requires different stages. Can be used either as importable for the drivers or as stand-alone by parsing arguments in the terminal. Example: Input file generator << `a_arm_input_generator` >>
Driver	Script that makes use of the modules to implement a complex protocol. Works as stand-alone by parsing arguments in the terminal. Example: *a* -ARM model building << `a_arm_protocol` >>
Package	Collection of related modules that work together to provide certain functionality. These modules are contained within a folder (`/src/`) and can be imported just like any other modules. This folder contains a special `__init__`.py file that tells Python it's a package, potentially containing more modules nested within subfolders. Example: << `PyARM` >>

suitable to be generalized to other photoreceptors including synthetic light-responsive proteins. As a first step in such a direction, the current implementation is suitable for the study of rhodopsins featuring artificial retinal-like chromophores, which is currently pursued in our group.

6.2 Current ARM-Based QM/MM Protocols

Presently, PyARM contains four different ARM-based fully automatic protocols, each implemented as a general driver (i.e., user-friendly one-click command-line interface), for the QM/MM modeling of rhodopsin electronically excited states. A brief description of each of these protocols, illustrated in Fig. 14, is provided below.

(1) Protocol: *a*-ARM model building protocol: ARM with chromophore cavity generation, ionization state selection, and external counterion placement.

- Driver: `a_arm_qmmm_protocol`
  - Phase I: Input file generator. Module: `a_arm_input_generator` (see Fig. 8a)

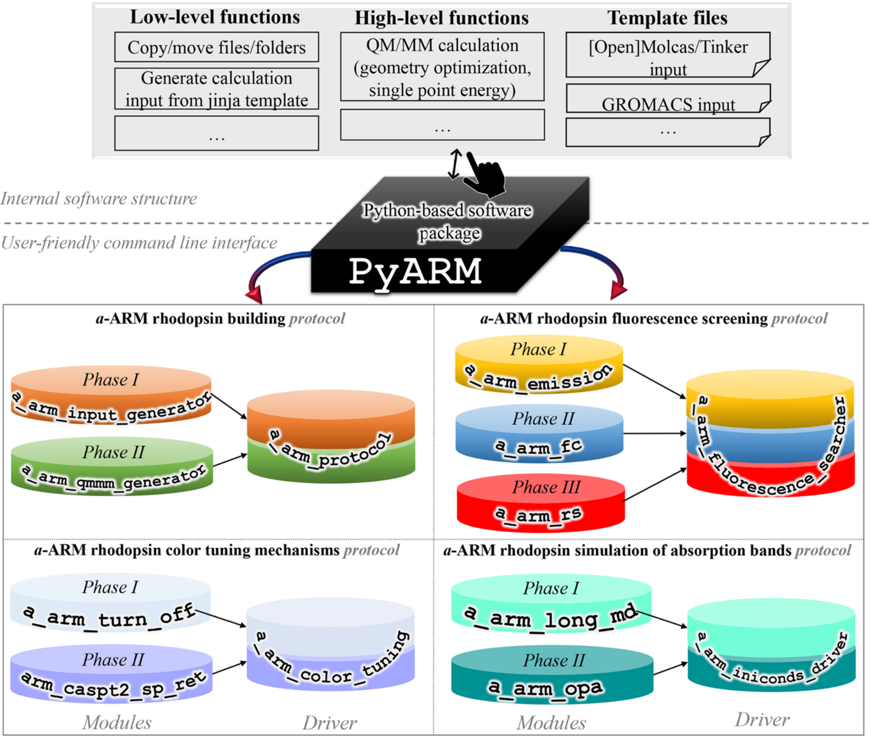

Low-level functions
Copy/move files/folders
Generate calculation input from jinja template
…
High-level functions
QM/MM calculation (geometry optimization, single point energy)
…
Template files
[Open]Molcas/Tinker input
GROMACS input
…
Internal software structure
User-friendly command line interface
Python-based software package
PyARM
a-ARM rhodopsin building protocol
Phase I
a_arm_input_generator
Phase II
a_arm_qmmm_generator
a_arm_protocol
a-ARM rhodopsin fluorescence screening protocol
Phase I
a_arm_emission
Phase II
a_arm_fc
Phase III
a_arm_rs
a_arm_fluorescence_searcher
a-ARM rhodopsin color tuning mechanisms protocol
Phase I
a_arm_turn_off
Phase II
arm_caspt2_sp_ret
a_arm_color_tuning
a-ARM rhodopsin simulation of absorption bands protocol
Phase I
a_arm_long_md
Phase II
a_arm_opa
a_arm_iniconds_driver
Modules
Driver
Modules
Driver

Fig. 14 Representation of the contents of the PyARM software package. The package contains various modules, which can be steered using a driver. Modules make use of high- and low-level functions, as well as template files that are also an integral part of the package

  - Phase II: QM/MM model generator. Module: a_arm_qmmm_generator (see Fig. 8b)

- Description: The first step towards the standardization and full automation of the *original* ARM was the development of *a*-ARM. This updated version not only overcame the automation drawbacks of the original version, but also included significant methodological/computational improvements. The achieved level of automation is accompanied by other features, such as speed in preparing the model building input, and standardization and reproducibility of the final model when operated by different users. In fact, the time required for preparing the input for the QM/MM model construction is reduced from around 3 h to less than 5 min (user time), with respect to the original ARM protocol. This is a consequence of the automation of the different preparatory steps, thus avoiding the user manipulation of text files and/or visualization of chemical structures. More specifically, a Python-based automatic input file generator subroutine was written to automate the assignment of the residues defining the chromophore cavity, including the chromophore linker and counter-ions, the protonation state of ionizable residues and, finally, the unambiguous placement of cytoplasmic and extracellular counter-ions. The benchmark calculations demonstrated that the resulting models for several rhodopsin sets are accurate enough for reproducing experimental trends in vertical excitation energy (i.e., the computed vertical excitation energies have an error bar of less than 4.0 kcal mol^{-1} blue-shifted), as well as transferability to rhodopsins of very different sequence. This protocol has been further described in Sect. 4.

(2) Protocol: Automated Analysis of Color Tuning Mechanisms in Rhodopsins.

- Driver: `a_arm_color_tuning`
  - Phase I: Electrostatic effects. Module: `a_arm_turn_off` (see Sects. 6.4.3 and 6.4.4)
  - Phase II: Steric effects. Module: `arm_caspt2_sp_ret` (see Sect. 6.4.2)
- Description: This protocol provides a tool to study how the protein environment (i.e., protein sequence) modulates the absorption wavelength of the retinal chromophore and, in turn, the color of the protein. In other words, it performs rhodopsin color tuning analysis so as to reveal steric and electrostatic effects between the retinal chromophore and the surrounding cavity amino acid residues. This tools aids the mechanistic description of the ways amino acid residues influence the photophysical properties of the retinal chromophore. This, in turn, builds up a reference book for helping the tuning of rhodopsin cavities (i.e., via point mutations) towards a desired effect (e.g., blue or red shifted wavelength). This protocol will be further described in Sect. 6.4.

(3) Protocol: Automated Simulation of Absorption Bands and light-induced Dynamics of Rhodopsins.

- Driver: `a_arm_iniconds_driver`
  - Phase I: Long MD from a-ARM QM/MM model. Module: `a_arm_long_md`
  - Phase II: Simulation of one photon absorption spectrum and light-induced dynamics. Module: `a_arm_opa`
- Description: This protocol allows the automatic simulation of one photon absorption (OPA) spectra, and the subsequent calculation of the initial conditions necessary for the study of the excited state dynamics, also allowing estimation of the photoisomerization quantum yield (QY). In other words, this tools provides another link between experimental quantities and simulated results. The photoisomerization QY, in particular, is of importance when interested in devising rhodopsins that are either particularly reactive or, contrarily, extremely fluorescent. This protocol will be the subject of a future publication.

(4) Protocol: Automated QM/MM Model Screening of Rhodopsin Variants Displaying Enhanced Fluorescence.

- Driver: `a_arm_fluorescence_searcher`

  - Phase I: Location of the S_1 fluorescent excited-state (FS). Module: `a_arm_emission`
  - Phase II: Computation of Franck–Condon (FC) trajectories. Module: `a_arm_fc`
  - Phase III: Calculation of the S_1 photoisomerization path (RS). Module: `a_arm_rs`

- Description: This protocol is designed for gaining insights into the possible molecular-level mechanisms behind fluorescence enhancement, when the fluorescent species correspond to the initial dark state of the rhodopsin photocycle. It is composed of three different phases able to categorize a rhodopsin as dim- or enhanced-fluorescent, with respect to a reference. This allows the fast and systematic in silico screening of potentially hundreds of rhodopsin mutants. Furthermore, each phase provides information (i.e., properties such as emission wavelength, excited state lifetime (ESL), energy isomerization barrier (E^f_{S1}), fluorescence quantum yield (ϕ^f) and structural parameters) that allows to elucidate and understand the factors determining rhodopsin fluorescence and, possibly, learn how to modulate it, with the ultimate goal of designing fluorescent candidates in silico for applications in, e.g., optogenetics [94]. This protocol will be the subject of a future publication.

6.3 PyARM Technical Details

6.3.1 PyARM Default Parameters

The setup for all the MD (see Sect. 3.4.1) and QM/MM (see Sect. 3.4.2) calculations performed by the different modules and drivers of PyARM (Sect. 6.2), is consistent with that reported for the *a*-ARM protocol in Refs. [43, 59, 64, 69] (see also Supplementary Notes 2 and 3 of Ref. [35]).

As described in Sect. 3.4.2, the default parameters for QM/MM calculations correspond to 2-roots single-state CASSCF(12,12)/AMBER optimization and state-average 3-roots CASPT2(12,12)/6-31G(d) energy correction. Such values for the active space and number of roots were selected based on benchmark calculations [59]. While the latter is not a modifiable parameter, the former could be modified via command-line arguments; however, it is strongly recommended to keep the default values.

As shown in Sect. 6.2, each instance of PyARM (see Table 1) is executed via a command-line interface, which contains a help menu facility that provides descriptions and syntax for the default parameters and how to modify them. Some of these parameters are listed below. Note that it is possible (but not recommended) to customize all the parameters for the MD run.

```
MD setup
-n    Number of independent replicas of Molecular Dynamics
      [default 10]
-ns   Specific seed for MD
      [default random number]
-mdt  Production temperature of the MD simulations (Kelvin)
      [default 298 K]
-mdh  Simulation time for the MD heating phase (ps)
      [default 50 ps]
-mde  Simulation time for the MD equilibration phase (ps)
      [default 150 ps]
-mdp  Simulation time for the MD production phase (ps)
      [default 800 ps]
QM/MM setup
-nr   Number of roots SA-CASSCF
      [default 2]
-lr   Number of roots CASPT2
      [default 3]
```

6.3.2 PyARM Installation

The PyARM code can be obtained by contacting the authors. PyARM relies on the following programs being already present and properly installed: Molcas ver. 8.4 or OpenMolcas [62, 64]; TINKER ver. 6.3 (with specific QM/MM patches) [63]; PROPKA ver. 3.1 [76]; PutIon [59]; GROMACS ver. 4.5.5 [61]; DOWSER [60]; PDB2PQR [95, 96]; Fpocket [97]; Modeller ver. 9.18 [78]. Furthermore, PyARM necessitates of a Python 3.x installation including the following modules: OpenBabel [98]; MDAnalysis [99, 100]; Matplotlib [101]; pandas [102]; NumPy [103]; SciPy [104]; Jinja2; Python-crontab; PyYAML; TextTable; cclib [105], GromacsWrapper.

PyARM contains a configuration file, the contents of which point to the installation path of the aforementioned programs. Submission template files are also provided for each type of calculation. These templates have to be tailored to the specifics of the computer cluster queuing system. Once these files are in place, and the user has write permissions for the Python installation, PyARM is installed as a normal Python 3.x package.

6.3.3 PyARM Tailoring

The modular nature of PyARM allows the relatively easy implementation of changes in the various drivers, as well as introducing a new driver. As a hypothetical example, *a*-ARM models have been used to compare excitation energy data obtained using CASPT2, MS-CASPT2 and MC-PDFT levels of theory [106]. The introduction of a different level of theory to evaluate excitation energies would imply using a new high-level function (Table 1) that performs such calculation, which in turn is based on a new template input file.

The new high-level function would be used by one or more modules, which in turn would belong to any number of drivers. If the user would like to simply swap level of theory, they would need to change the call to the original high-level function in each module with the newly made one. Otherwise, the user can simply create copies of the desired modules and drivers, which makes use of the new high-level function.

Similarly, a new driver could be implemented, taking advantage of the already present modules and functions as much as possible. For example, one could introduce a new, hypothetical, driver to locate the possible photoproduct structure of a given rhodopsin. Such a driver would combine the machinery of the previously mentioned `a_arm_fluorescence_searcher`, but extending the `a_arm_rs` module (i.e., making a new module) to perform a ca. 180 degrees isomerization, followed by a S_0 geometry optimization and energetic evaluation. In this case, all necessary high-level functions and templates are already present, they just have to be combined in a different way.

6.4 Color Tuning Analysis in Terms of Steric and Electrostatic Effects

(Part of the content of the next three sections is reproduced/adapted with permission from [35], open access under a CC BY license (Creative Commons Attribution 4.0 International License)).

As an example of a possible usage of a driver and a module as found inside `PyARM`, this section presents the driver (`a_arm_color_tuning`) behind a common kind of analysis [35, 42], which aims at discerning the contribution of nearby cavity residues onto the computed excitation energy ΔE^{a}_{S1-S0} of the *r*PSB, and assign them as blue- or red-shifting. The shifting contributions are assessed in terms of both electrostatic effects onto the charged *r*PSB (Sect. 6.4.3), and steric effects distorting the geometry of the retinal chromophore (Sect. 6.4.2). This assessment is performed based on quantities defined in Sect. 6.4.1. Figure 15 presents the results of using this analysis, as found while performing all possible mutations (P219X) for the KR2 rhodopsin [35]. Finally, Sect. 6.4.4 presents a simpler possible use of the module `a_arm_turn_off` to analyze electrostatic contributions from cavity residues onto the *r*PSB.

6.4.1 Computed Quantities

In Ref. [35], the authors define three fundamental quantities ($\Delta\Delta E^{\mathrm{TOT}}_{\mathrm{S1\text{-}S0}}$, $\Delta\Delta E^{\mathrm{STR}}_{\mathrm{S1\text{-}S0}}$, and $\Delta\Delta E^{\mathrm{OFF}}_{\mathrm{S1\text{-}S0}}$) 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, using as an illustrative example the mutants P219X of the KR2 rhodopsin (Fig. 11c).

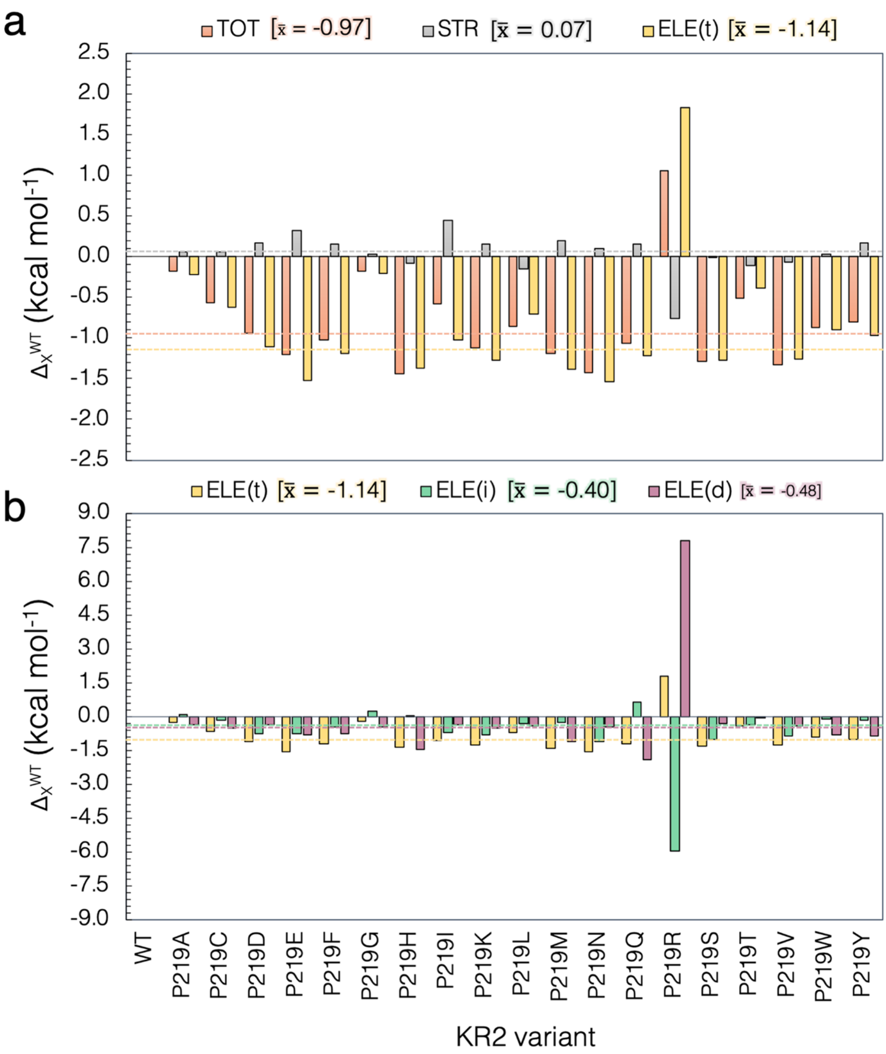


Fig. 15 Steric and electrostatic contributions to the vertical excitation energy. **a** Total ($\Delta\Delta E^{TOT}_{S1\text{-}S0}$), steric ($\Delta\Delta E^{STR}_{S1\text{-}S0}$) and electrostatic ($\Delta\Delta E^{ELE(t)}_{S1\text{-}S0}$) contributions to the interaction of the retinal with the protein environment for the possible P219X mutations. **b** Decomposition of the total electrostatic effects on its indirect ($\Delta\Delta E^{ELE(i)}_{S1\text{-}S0}$) and direct ($\Delta\Delta E^{ELE(d)}_{S1\text{-}S0}$) components. The dashed lines and corresponding numerical values refer to the weighted average values ($\bar{x}$) of the 18 residues of the red-shifted cluster exclusively (that is, excluding P219R), presented in square parenthesis. Reproduced with permission from [35], open access under a CC BY license (Creative Commons Attribution 4.0 International License)

- $\Delta\Delta E^{TOT}_{S1\text{-}S0}$ is the "Total" excitation energy change. It is computed directly as the difference between the QM/MM computed vertical excitation energies of mutant and WT models.
- $\Delta\Delta E^{STR}_{S1\text{-}S0}$ is the "Steric component" of $\Delta\Delta E^{TOT}_{S1\text{-}S0}$. It is computed directly as the difference between the QM/MM vertical excitation energies of the iso-

lated retinal chromophores, however retaining the geometries as if inside the protein environment, between mutant and WT.

- $\Delta\Delta E^{OFF}_{S1\text{-}S0}$ is used to quantify the "indirect" electrostatic component (see below) of $\Delta\Delta E^{TOT}_{S1\text{-}S0}$. It is computed as the difference between the vertical excitation energy of the mutant and WT obtained after having switched off (turned to zero) the charges of residue 219. $\Delta\Delta E^{OFF}_{S1\text{-}S0}$ can also be used directly to analyze, qualitatively, the red- or blue-shifting role of each of the residues of the chromophore cavity for a determined rhodopsin (Turn-Off Module). In such a case, it is computed as the differences between the vertical excitation energy of the rhodopsin before and after having switched off (turned to zero) the charges of specific residues.

As we will see in the following, using these quantities, we can compute three additional components.

- $\Delta\Delta E^{ELE(t)}_{S1\text{-}S0}$ is the "Total electrostatic component" that is indirectly computed as the difference between the Total and the Steric components above ($\Delta\Delta E^{ELE(t)}_{S1\text{-}S0} = \Delta\Delta E^{TOT}_{S1\text{-}S0} - \Delta\Delta E^{STR}_{S1\text{-}S0}$) for each mutant. As specified in Sect. 6.4.2, $\Delta\Delta E^{ELE(t)}_{S1\text{-}S0}$ can be decomposed into two parts:
  - $\Delta\Delta E^{ELE(i)}_{S1\text{-}S0}$ 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 obtained after having switched off (turned to zero) the charges of residue 219, and then by subtracting from such difference the steric effect $\Delta\Delta E^{STR}_{S1\text{-}S0}$ defined above.
  - Finally, $\Delta\Delta E^{ELE(d)}_{S1\text{-}S0}$ is the "Direct electrostatic component" that is computed indirectly as $\Delta\Delta E^{ELE(d)}_{S1\text{-}S0} = \Delta\Delta E^{ELE(t)}_{S1\text{-}S0} - \Delta\Delta E^{ELE(i)}_{S1\text{-}S0}$.

6.4.2 Steric Effects

In the context of the protocol, 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. In turn, this change in geometry of the *r*PSB could be induced by both steric and electrostatic factors. Such effects are investigated by analyzing how the retinal chromophore is structurally modified by mutations near the *β*-ionone or near the Schiff base linkage (see Fig. 5).

As discussed in Ref. [35], such structural rearrangements of the *r*PSB may be due to different factors, such as a simple effect induced by the side-chain replacement, a different charge distribution due to changes in protonation states for ionizable residues, as well as water molecules addition/removal. Notice that steric effects were evaluated through an "atomistic" approach, focused on the changes in the *r*PSB geometrical and electronic structure, and therefore not directly related to steric effects evaluated on the basis of the changes in residue volume [35, 92].

6.4.3 Electrostatic Effects

As mentioned above, the total electrostatic effect ($\Delta\Delta E^{\mathrm{ELE(t)}}_{\mathrm{S1\text{-}S0}}$) can be decomposed in two parts: (1) the first can be considered as a direct component ($\Delta\Delta E^{\mathrm{ELE(d)}}_{\mathrm{S1\text{-}S0}}$) due to the variation in number, magnitude, and position of the point charges of mutated residue caused by the P to X replacement, and (2) a more indirect component ($\Delta\Delta E^{\mathrm{ELE(i)}}_{\mathrm{S1\text{-}S0}}$) 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 in Ref. [35], possible changes in protonation states of conserved residues, induced by P to X replacement, have a major contribution to the indirect component. Figure 15 shows the contributions due to different effects (steric and electrostatic) and components.

6.4.4 Turn-Off Module

As a final example, we present a possible usage of the `a_arm_turn_off` module, which is part of the `a_arm_color_tuning` driver (Phase I). As previously stated (Table 1), modules can be used also independently from the parent driver. When used as a stand-alone script, the command-line menu of the module provides four different options to turn-off residues in the chromophore cavity, with different scopes, as follows:

1. < `cav` >: Turn off the charges of EACH of the residues in the cavity, independently. For instance, if the cavity is formed by *N* residues, the output will be *N* single calculations.
2. < `cav_all` >: Turn off the charges for ALL the residues in the cavity, simultaneously. Therefore, the output will be a single calculation.
3. < `single` >: Turn off the charges for ONE specific residue given by the user.
4. < `multiple` >: Turn off the charges for a list consisting of MULTIPLE residues, given by the user, simultaneously. Therefore, the output will be a single calculation.

As an example of a possible analysis, we envision the use of the module with the first option (< `cav` >), to map out the expected electrostatic influence of cavity residues onto the ΔE^{a}_{S1-S0} of the *r*PSB. Such analysis could drive subsequent mutation tests, such as all possible mutations performed onto the residue responsible for, e.g., the largest or smallest shift.

6.5 `PyARM` Current Accuracy and Drawbacks

A number of limitations of *a*-ARM have already been described in Sects. 3.5, 4.4 and 5.2. Due to their simplified definition, ARM models are more exposed to potential pitfalls than more complex QM/MM models. Such possible pitfalls, that concern the QM/MM model generator phase, can be summarized as:

(1) lack of a proper description of the protein environment (membrane + explicit solvent),
(2) rigid protein backbone and non-cavity side-chains,
(3) approximated protonation states for ionizable residues, and
(4) missing description of any mutual polarization effects between the QM and MM sub-systems, that can be accounted for by polarizable embedding using a polarizable force field. Since polarizable force fields are technologies still under development in the QM/MM area (see, for instance Ref. [107]), we have not adopted/benchmarked them in this version of our specialized QM/MM models.

When considering points 1–4, the different properties computed by ARM are expected to be affected by a systematic error. Our current research is aimed at dealing with those points, while maintaining reasonable computational costs, or estimating the errors due to them. Nevertheless, according to the philosophy of the ARM protocol (Sect. 3.2), the main focus of ARM is the ability to reproduce property (especially $\lambda^{a}_{\mathrm{max}}$) and explain trends, rather than predicting their absolute values. The default model generation and subsequent customization revised above clearly shows that the models can fit the experimental results. In other words, the customization protocol searches for possible sources of systematic errors (i.e., different protonation state, different rotamer), due to points 1–4.

A trend deviation factor was computed [69], to evaluate the accuracy of the *a*-ARM data in a given set of rhodopsins with available experimental data. Thus, a mean absolute error of 2.5 kcal mol^{-1} was found (Table S4 in the supporting information of Ref. [69]), when considering *a*-ARM$_{\mathrm{default}}$ results for the benchmark set of Fig. 11b. A smaller value of 0.4 kcal mol^{-1} (Supplementary Table 5 of Ref. [35]) was found for the coherent set of KR2 mutants presented in Fig. 11d. These numbers show the expected accuracy of `PyARM` in evaluating trends of photophysical properties of rhodopsin sets. Logically, a coherent set of mutants of the same rhodopsin shows a smaller deviation, with respect to a set of rhodopsins spanning different organisms.

Although outside the *a*-ARM scope, it is possible to evaluate the accuracy of `PyARM` with respect to computing absolute absorbed wavelengths.[4] For example, with respect to the data behind Fig. 11b (Table S3 in the supporting information of Ref. [69]), we found an average accuracy for the *a*-ARM$_{\mathrm{default}}$ protocol of 90%.

[4] We here define accuracy as the absolute difference of computed and experimental absorbed wavelength, divided by the experimental number, in percentage. This is done for each of the N=10 replicas of a rhodopsin (Sect. 3.4.1) and then averaged. The final number is the average of the accuracy of all the considered rhodopsins.

7 Outlook and Concluding Remarks

In this review, we presented the ARM protocol, along with its past and current development and achievements. After a brief introduction to rhodopsins, their importance and current and potential technological uses, we presented the development of ARM. We saw how the protocol was thought from inception to be accurate enough to reproduce trends in photochemical properties (mainly ΔE^{a}_{S1-S0}), that is points (1) and (2) of Sect. 3.5.

Most of the remaining points (3)–(5) were addressed through the subsequent development of the Updated version of the Automatic Rhodopsin Modeling protocol *a*-ARM, presented in Sect. 4. The major achievement of *a*-ARM has been the complete automation of the input file generator and QM/MM model generator, and the corresponding coding as python, user-friendly, command-line interfaces. Such tools were also included in Sect. 5. The method was benchmarked (Fig. 11), and found reliable in obtaining photochemical properties of different rhodopsins, obtained from various life domains. Furthermore, the usage of the input file generator ensured reproducibility of the results, even when considering the two approaches *a*-ARM$_{\text{default}}$ and *a*-ARM$_{\text{customized}}$.

Finally, Sect. 6 presented the latest development of the ARM protocol, namely `PyARM`. As shown in Fig. 14, `PyARM` consists of different drivers capable of performing various actions and automatic analyses, combining several python modules, which in turn are composed of functions and templates (Table 1). The automation level achieved by `PyARM` allows the efficient generation of, e.g., all 19 possible mutants of a given residue (Fig. 11c). The concurrent preparation of many QM/MM models allows, in principle, the study of rhodopsin mutants arrays. The completely modular architecture of `PyARM` will permit scaling up the code (i.e., introducing novel kinds of analyses), through the implementation of new drivers, alongside those described in Sect. 6.2.

Despite the encouraging outcome of the applications of the protocol, additional work is required, for providing the scientific community with a robust tool that can be applied systematically to the study and design of sizable rhodopsin arrays. Much of its future success will depend on further improvements in the construction of the *a*-ARM model and, in particular, of rhodopsin mutants models. Some of the envisioned improvements are as follows:

- Since rhodopsin structural data are rarely available and still difficult to obtain experimentally, it would be important to integrate, in an automatic fashion, comparative modeling technologies in the protocol for *a*-ARM model building. It is possible that with such tool one could achieve a protocol capable of producing more accurate QM/MM models starting directly from the constantly growing repositories of rhodopsin sequences. This target is currently pursued in our laboratory.
- The above item is also related to the improvement of mutant models generation and screening. As a perspective of this work, we suggest to introduce a mutant

generator routine using proper comparative (homology) modeling, instead of just modifying the mutated side-chain conformation locally, as seen in Sect. 4.5.
- Both previous items could also benefit from the introduction of a machine-learning or artificial intelligence (AI)-assisted protocol for more accurate structures or overall structure predictions.
- Methods for improving the prediction of the residue protonation states during the construction of ARM QM/MM models appear of capital importance for increasing the accuracy (e.g., the percentage of success in reproducing trends in spectroscopic properties) of the final models. These are being presently investigated in our lab (see for instance a preliminary study in Ref. [75]).
- Currently, the *a* ARM model building protocol is the only ARM-based tool implemented in the Web-ARM interface and, therefore, accessible through the web. Future efforts will be devoted to the implementation of the four protocols described in Sect. 6.2 as utilities of the Web-ARM interface.

Finally, long-term development goals include introducing new drivers for `PyARM` (possibly re-using some of the already existing modules and scripts) for novel kinds of analyses. It is our intention to pursue also the possibility of simulating the use of different retinal chromophores, either natural or synthetic. We foresee the possibility of having a generalized method capable of handling any kind of chromophore. Last but not least, we would like to extend the applicability of the protocol beyond rhodopsins, and be able to apply it to many (any) photoactive protein complex.

Acknowledgements Thank to all current and past group members that have contributed to the development of the ARM protocol: Federico Melaccio, Alessio Valentini, Yoelvis Orozco-Gonzalez, Nicolas Ferré, Samira Gholami, Fabio Montisci, Silvia Rinaldi, Marco Cherubini, Xuchun Yang, Michael Stenrup, Hoi-Ling Luk, María Del Carmen Marín, Simone Bonfrate and Riccardo Palombo.

Author Contributions LPG and LDV prepared text and figures, all authors participated to their revision

Funding We acknowledge funding from the following entities: NSF (grant CHE-CLP-1710191), NIH (grant 1R15GM126627 01), Banca D'Italia, MIUR (Dipartimento di Eccellenza, 2018–2022), MUR (Rita Levi Montalcini grant, 2021-2024)

Code availabiity The code can be obtained by contacting the authors.

Declarations

Conflict of interest All authors declare that they have no competing interests.

Ethics approval Not applicable

Consent to participate Not applicable

Consent for publication Not applicable

References

1. Kovalev K, Polovinkin V, Gushchin I, Alekseev A, Shevchenko V, Borshchevskiy V, Astashkin R, Balandin T, Bratanov D, Vaganova S et al (2019) Structure and mechanisms of sodium-pumping KR2 rhodopsin. Sci Adv 5(4):2671
2. Okada T, Sugihara M, Bondar AN, Elstner M, Entel P, Buss V (2004) The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. J Mol Biol 342(2):571–583
3. Shihoya W, Inoue K, Singh M, Konno M, Hososhima S, Yamashita K, Ikeda K, Higuchi A, Izume T, Okazaki S et al (2019) Crystal structure of heliorhodopsin. Nature 574(7776):132–136
4. Braslavsky SE (2007) Glossary of terms used in photochemistry, (IUPAC Recommendations 2006). Pure Appl Chem 79(3):293–465
5. Ernst OP, Lodowski DT, Elstner M, Hegemann P, Brown LS, Kandori H (2014) Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. Chem Rev 114(1):126–163
6. Govorunova EG, Sineshchekov OA, Li H, Spudich JL (2017) Microbial rhodopsins: diversity, mechanisms, and optogenetic applications. Annu Rev Biochem 86:845–872
7. Kandori H (2020) Retinal proteins: photochemistry and optogenetics. Bull Chem Soc Jpn 93(1):76–85
8. Kurihara M, Sudo Y (2015) Microbial rhodopsins: wide distribution, rich diversity and great potential. Biophys Psychobiol 12:121–129
9. Kojima K, Shibukawa A, Sudo Y (2020) The unlimited potential of microbial rhodopsins as optical tools. Biochemistry 59(3):218–229
10. Kojima K, Kurihara R, Sakamoto M, Takanashi T, Kuramochi H, Zhang XM, Bito H, Tahara T, Sudo Y (2020) Comparative studies of the fluorescence properties of microbial rhodopsins: spontaneous emission versus photointermediate fluorescence. J Phys Chem B 124(34):7361–7367
11. Needham DM, Yoshizawa S, Hosaka T, Poirier C, Choi CJ, Hehenberger E, Irwin NA, Wilken S, Yung C-M, Bachy C et al (2019) A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. Proc Natl Acad Sci USA 116(41):20574–20583
12. Bratanov D, Kovalev K, Machtens J-P, Astashkin R, Chizhov I, Soloviov D, Volkov D, Polovinkin V, Zabelskii D, Mager T et al (2019) Unique structure and function of viral rhodopsins. Nat Commun 10(1):4939
13. Pushkarev A, Béjà O (2016) Functional metagenomic screen reveals new and diverse microbial rhodopsins. ISME J 10(9):2331–2335
14. Luk HL, Melaccio F, Rinaldi S, Gozem S, Olivucci M (2015) Molecular bases for the selection of the chromophore of animal rhodopsins. Proc Natl Acad Sci USA 112(50):15297–15302
15. Pushkarev A, Inoue K, Larom S, Flores-Uribe J, Singh M, Konno M, Tomida S, Ito S, Nakamura R, Tsunoda SP et al (2018) A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. Nature 558(7711):595–599
16. Lenahan C, Sanghavi R, Huang L, Zhang JH (2020) Rhodopsin: a potential biomarker for neurodegenerative diseases. Front Neurosci 14:14
17. Tsujimura M, Ishikita H (2020) Insights into the protein functions and absorption wavelengths of microbial rhodopsins. J Phys Chem B 124(52):11819–11826
18. Tahara S, Singh M, Kuramochi H, Shihoya W, Inoue K, Nureki O, Béjà O, Mizutani Y, Kandori H, Tahara T (2019) Ultrafast dynamics of heliorhodopsins. J Phys Chem B. 123(11):2507–2512
19. Tanaka T, Singh M, Shihoya W, Yamashita K, Kandori H, Nureki O (2020) Structural basis for unique color tuning mechanism in heliorhodopsin. Biochem Biophys Res Commun 533(3):262–267

20. Kim S-H, Chuon K, Cho S-G, Choi A, Meas S, Cho H-S, Jung K-H (2021) Color-tuning of natural variants of heliorhodopsin. Sci Rep 11(1):1–9
21. Karasuyama M, Inoue K, Nakamura R, Kandori H, Takeuchi I (2018) Understanding colour tuning rules and predicting absorption wavelengths of microbial rhodopsins by data-driven machine-learning approach. Sci Rep 8(1):15580
22. Harris A, Lazaratos M, Siemers M, Watt E, Hoang A, Tomida S, Schubert L, Saita M, Heberle J, Furutani Y, Kandori H, Bondar AN, Brown LS (2020) Mechanism of inward proton transport in an Antarctic microbial rhodopsin. J Phys Chem B 124(24):4851–4872
23. Kandori H, Shichida Y, Yoshizawa T (2001) Photoisomerization in rhodopsin. Biochemistry (Moscow) 66(11):1197–1209
24. Mai S, González L (2020) Molecular photochemistry: recent developments in theory. Angew Chem Int Ed 59(39):16832–16846
25. Luecke H, Schobert B, Lanyi JK, Spudich EN, Spudich JL (2001) Crystal structure of sensory rhodopsin: insights into color tuning and transducer interaction II at 2.4 Angstroms. Science 293(5534):1499–1503
26. Hoffmann M, Wanko M, Strodel P, König PH, Frauenheim T, Schulten K, Thiel W, Tajkhorshid E, Elstner M (2006) Color tuning in rhodopsins: the mechanism for the spectral shift between bacteriorhodopsin and sensory rhodopsin II. J Am Chem Soc 128(33):10808–10818
27. Wanko M, Hoffmann M, Frauenheim T, Elstner M (2006) Computational photochemistry of retinal proteins. J Comput Aided Mol Des 20(7–8):511–518
28. Fujimoto K, Hasegawa J-Y, Hayashi S, Kato S, Nakatsuji H (2005) Mechanism of color tuning in retinal protein: SAC-CI and QM/MM study. Chem Phys Lett 414(1–3):239–242
29. Fujimoto K, Hayashi S, Hasegawa JY, Nakatsuji H (2007) Theoretical studies on the color-tuning mechanism in retinal proteins. J Chem Theory Comput 3(2):605–618
30. Altun A, Yokoyama S, Morokuma K (2008) Mechanism of spectral tuning going from retinal in vacuo to bovine rhodopsin and its mutants: multireference ab initio quantum mechanics/molecular mechanics studies. J Phys Chem B 112(51):16883–16890
31. Altun A, Yokoyama S, Morokuma K (2008) Spectral tuning in visual pigments: an ONIOM(QM:MM) study on bovine rhodopsin and its mutants. J Phys Chem B 112(22):6814–6827
32. Kim SY, Waschuk SA, Brown LS, Jung KH (2008) Screening and characterization of proteorhodopsin color-tuning mutations in *Escherichia coli* with endogenous retinal synthesis. Biochim Biophys Acta Bioenerg 1777(6):504–513
33. Palczewska G, Vinberg F, Stremplewski P, Bircher MP, Salom D, Komar K, Zhang J, Cascella M, Wojtkowski M, Kefalov VJ et al (2014) Human infrared vision is triggered by two-photon chromophore isomerization. Proc Natl Acad Sci USA 111(50):5445–5454
34. Engqvist MKM, McIsaac RS, Dollinger P, Flytzanis NC, Abrams M, Schor S, Arnold FH (2015) Directed evolution of *Gloeobacter violaceus* rhodopsin spectral properties. J Mol Biol 427(1):205–220
35. Nakajima Y, Pedraza-González L, Barneschi L, Inoue K, Olivucci M, Kandori H (2021) Pro219 is an electrostatic color determinant in the light-driven sodium pump KR2. Commun Biol 4(1185):1–15
36. Birge RR, Murray LP, Pierce BM, Akita H, Balogh-Nair V, Findsen LA, Nakanishi K (1985) Two-photon spectroscopy of locked-11-cis-rhodopsin: evidence for a protonated schiff base in a neutral protein binding site. Proc Natl Acad Sci USA 82(12):4117–4121
37. Birge RR (1986) Two-photon spectroscopy of protein-bound chromophores. Acc Chem Res 19(5):138–146
38. Swartz TE, Szundi I, Spudich JL, Bogomolni RA (2000) New photointermediates in the two photon signaling pathway of sensory rhodopsin-i. Biochemistry 39(49):15101–15109
39. Ehrenberg D, Varma N, Deupi X, Koyanagi M, Terakita A, Schertler GF, Heberle J, Lesca E (2019) The two-photon reversible reaction of the bistable jumping spider rhodopsin-1. Biophys J 116(7):1248–1258
40. Gholami S, Pedraza-González L, Yang X, Granovsky AA, Ioffe IN, Olivucci M (2019) Multistate multiconfiguration quantum chemical computation of the two-photon absorption spectra of bovine rhodopsin. J Phys Chem Lett 10(20):6293–6300
41. Deisseroth K (2011) Optogenetics. Nat Methods 8(1):26–29

42. Marín MdC, Agathangelou D, Orozco-Gonzalez Y, Valentini A, Kato Y, Abe-Yoshizumi R, Kandori H, Choi A, Jung KH, Haacke S, Olivucci M (2019) Fluorescence enhancement of a microbial rhodopsin via electronic reprogramming. J Am Chem Soc 141(1):262–271
43. Pedraza-González L, Marín MdC, De Vico L, Yang X, Olivucci M (2020) On the automatic construction of QM/mm models for biological photoreceptors: rhodopsins as model systems. QM/MM studies of light-responsive biological systems. Springer, Berlin, pp 1–75
44. Bouas-Laurent H, Dürr H (2001) Organic photochromism (iupac technical report). Pure Appl Chem 73(4):639–665
45. Mendes HF, Van Der Spuy J, Chapple JP, Cheetham ME (2005) Mechanisms of cell death in rhodopsin retinitis pigmentosa: implications for therapy. Trends Mol Med 11(4):177–185
46. Mendes HF, Cheetham ME (2008) Pharmacological manipulation of gain-of-function and dominant-negative mechanisms in rhodopsin retinitis pigmentosa. Hum Mol Genet 17(19):3043–3054
47. Athanasiou D, Aguila M, Bellingham J, Li W, McCulley C, Reeves PJ, Cheetham ME (2018) The molecular and cellular basis of rhodopsin retinitis pigmentosa reveals potential strategies for therapy. Prog Retin Eye Res 62:1–23
48. Skulachev VP, Bogachev A (1988) Membrane bioenergetics. Springer, Berlin
49. Klapoetke NC, Murata Y, Kim SS, Pulver SR, Birdsey-Benson A, Cho YK, Morimoto TK, Chuong AS, Carpenter EJ, Tian Z, Wang J, Xie Y, Yan Z, Zhang Y, Chow BY, Surek B, Melkonian M, Jayaraman V, Constantine-Paton M, Wong GKS, Boyden ES (2014) Independent optical excitation of distinct neural populations. Nat Methods 11(3):338–346
50. Bogomolni RA, Spudich JL (1987) The photochemical reactions of bacterial sensory rhodopsin-I. Flash photolysis study in the one microsecond to eight second time window. Biophys J 52(6):1071–1075
51. Béja O, Spudich EN, Spudich JL, Leclerc M, DeLong EF (2001) Proteorhodopsin phototrophy in the ocean. Nature 411(6839):786–789
52. Romei MG, Lin CY, Mathews II, Boxer SG (2020) Electrostatic control of photoisomerization pathways in proteins. Science 367(6473):76–79
53. Okada T, Fujiyoshi Y, Silow M, Navarro J, Landau EM, Shichida Y (2002) Functional role of internal water molecules in rhodopsin revealed by x-ray crystallography. Proc Natl Acad Sci USA 99(9):5982–5987
54. Teller DC, Okada T, Behnke CA, Palczewski K, Stenkamp RE (2001) Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). Biochemistry 40(26):7761–7772
55. Andruniów T, Ferré N, Olivucci M (2004) Structure, initial excited-state relaxation, and energy storage of rhodopsin resolved at the multiconfigurational perturbation theory level. Proc Natl Acad Sci USA 101(52):17908–17913
56. Tomasello G, Gloria OG, Altoè P, Stenta M, Luis SA, Merchán M, Orlandi G, Bottoni A, Garavelli M (2009) Electrostatic control of the photoisomerization efficiency and optical properties in visual pigments: on the role of counterion quenching. J Am Chem Soc 131(14):5172–5186
57. Bravaya K, Bochenkova A, Granovsky A, Nemukhin A (2007) An opsin shift in rhodopsin: Retinal S0–S1 excitation in protein, in solution, and in the gas phase. J Am Chem Soc 129(43):13035–13042
58. Valsson O, Campomanes P, Tavernelli I, Rothlisberger U, Filippi C (2013) Rhodopsin absorption from first principles: bypassing common pitfalls. J Chem Theory Comput 9(5):2441–2454
59. Melaccio F, Marín MdC, Valentini A, Montisci F, Rinaldi S, Cherubini M, Yang X, Kato Y, Stenrup M, Orozco-Gonzalez Y, Ferré N, Luk HL, Kandori H, Olivucci M (2016) Toward automatic rhodopsin modeling as a tool for high-throughput computational photobiology. J Chem Theory Comput 12(12):6020–6034
60. Zhang L, Hermans J (1996) Hydrophilicity of cavities in proteins. Proteins J Chem Theory Comput Bioinf 24(4):433–438
61. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Van Der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29(7):845–854
62. ...Aquilante F, Autschbach J, Carlson RK, Chibotaru LF, Delcey MG, De Vico L, Fdez Galván I, Ferré N, Frutos LM, Gagliardi L, Garavelli M, Giussani A, Hoyer CE, Li Manni G, Lischka H, Ma D, Malmqvist PÅ, Müller T, Nenov A, Olivucci M, Bondo Pedersen T, Peng D, Plasser F, Pritchard B, Reiher M, Rivalta I, Schapiro I, Segarra-Martí J, Stenrup M, Truhlar DG, Ungur L, Valentini A, Vancoillie S, Veryazov V, Vysotskiy VP, Weingart O, Zapata F, Lindh R (2016)

Molcas8: new capabilities for multiconfigurational quantum chemical calculations across the periodic table. J Comput Chem 37(5):506–541

63. Rackers JA, Wang Z, Lu C, Laury ML, Lagardére L, Schnieders MJ, Piquemal J-P, Ren P, Ponder JW (2018) Tinker 8: software tools for molecular design. J Chem Theory Comput 14(10):5273–5289
64. ...Aquilante F, Autschbach J, Baiardi A, Battaglia S, Borin VA, Chibotaru LF, Conti I, De Vico L, Delcey M, Fdez Galván I, Ferré N, Freitag L, Garavelli M, Gong X, Knecht S, Larsson E, Lindh R, Lundberg M, Malmqvist P-A, Nenov A, Norell J, Odelius M, Olivucci M, Pedersen T, Pedraza-González L, Phung Q, Pierloot K, Reiher M, Schapiro I, Segarra-Martí J, Segatta F, Seijo L, Sen S, Sergentu D-C, Stein C, Ungur L, Vacher M, Valentini A, Veryazov V (2020) Modern quantum chemistry with [Open] Molcas. J Chem Phys 152(21):214117
65. Melaccio F, Olivucci M, Lindh R, Ferré N (2011) Unique QM/MM potential energy surface exploration using microiterations. Int J Quantum Chem 111(13):3339–3346
66. Inoue K, Ito S, Kato Y, Nomura Y, Shibata M, Uchihashi T, Tsunoda SP, Kandori H (2016) A natural light-driven inward proton pump. Nat Commun 7(1):1–10
67. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The protein data bank: a computer-based archival file for macromolecular structures. Eur J Biochem 80(2):319–324
68. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28(1):235–242
69. Pedraza-González L, De Vico L, Marín MdC, Fanelli F, Olivucci M (2019) a-ARM: automatic rhodopsin modeling with chromophore cavity generation, ionization state selection, and external counterion placement. J Chem Theory Comput 15(5):3134–3152
70. Kato HE, Inoue K, Abe-Yoshizumi R, Kato Y, Ono H, Konno M, Hososhima S, Ishizuka T, Hoque MR, Kunitomo H, Ito J, Yoshizawa S, Yamashita K, Takemoto M, Nishizawa T, Taniguchi R, Kogure K, Maturana AD, Iino Y, Yawo H, Ishitani R, Hideki K, Nureki O (2015) Structural basis for Na(+) transport mechanism by a light-driven Na(+) pump. Nature 521(7550):48–53
71. Broser M, Spreen A, Konold PE, Peter E, Adam S, Borin V, Schapiro I, Seifert R, Kennis JT, Sierra YAB et al (2020) Neor, a near-infrared absorbing rhodopsin. Nat Commun 11(1):5682
72. Adam S, Wiebeler C, Schapiro I (2021) Structural factors determining the absorption spectrum of channelrhodopsins: a case study of the chimera c1c2. J Chem Theory Comput 17(10):6302–6313
73. Kaufmann JC, Krause BS, Adam S, Ritter E, Schapiro I, Hegemann P, Bartl FJ (2020) Modulation of light energy transfer from chromophore to protein in the channelrhodopsin reachr. Biophys J 119(3):705–716
74. Pedraza-González L, Marín MdC, Jorge AN, Ruck TD, Yang X, Valentini A, Olivucci M, De Vico L (2020) Web-ARM: a web-based interface for the automatic construction of QM/MM models of rhodopsins. J Chem Inf Model 60(3):1481–1493
75. Pieri E, Ledentu V, Sahlin M, Dehez F, Olivucci M, Ferré N (2019) CpHMD-then-QM/MM identification of the amino acids responsible for the anabaena sensory rhodopsin pH-dependent electronic absorption spectrum. J Chem Theory Comput 15(8):4535–4546
76. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput 7(2):525–537
77. Krivov GG, Shapovalov MV, Dunbrack RL (2009) Improved prediction of protein side-chain conformations with SCWRL4. Proteins Struct Funct Bioinf 77(4):778–795
78. Webb B, Sali A (2016) Comparative protein structure modeling using MODELLER. Curr Protoc Bioinform 54(1):5–6
79. Ochoa R, Soler MA, Laio A, Cossio P (2018) Assessing the capability of in silico mutation protocols for predicting the finite temperature conformation of amino acids. Phys Chem Chem Phys 20(40):25901–25909
80. Ignatov A (2021) Statistical analysis of protein side-chain conformations. J Phys Conf Ser 1740:012013
81. Xiang Z, Honig B (2001) Extending the accuracy limits of prediction for side-chain conformations. J Mol Biol 311(2):421–430
82. Wilson C, Gregoret LM, Agard DA (1993) Modeling side-chain conformation for homologous proteins using an energy-based rotamer search. J Mol Biol 229(4):996–1006
83. Dunbrack RL Jr, Karplus M (1993) Backbone-dependent rotamer library for proteins application to side-chain prediction. J Mol Biol 230(2):543–574
84. Vasquez M (1996) Modeling side-chain conformation. Curr Opin Struct Biol 6(2):217–221
85. Kono H, Doi J (1996) A new method for side-chain conformation prediction using a hopfield network and reproduced rotamers. J Comput Chem 17(14):1667–1683

86. Canutescu AA, Shelenkov AA, Dunbrack RL Jr (2003) A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci 12(9):2001–2014
87. Peterson LX, Kang X, Kihara D (2014) Assessment of protein side-chain conformation prediction methods in different residue environments. Proteins Struct Funct Bioinf 82(9):1971–1984
88. Nagata K, Randall A, Baldi P (2012) Sidepro: a novel machine learning approach for the fast and accurate prediction of side-chain conformations. Proteins Struct Funct Bioinf 80(1):142–153
89. Liang S, Zheng D, Zhang C, Standley DM (2011) Fast and accurate prediction of protein side-chain conformations. Bioinformatics 27(20):2913–2914
90. Dunbrack RL Jr (2002) Rotamer libraries in the 21st century. Curr Opin Struct Biol 12(4):431–440
91. Mroginski M-A, Adam S, Amoyal GS, Barnoy A, Bondar A-N, Borin VA, Church JR, Domratcheva T, Ensing B, Fanelli F et al (2021) Frontiers in multiscale modeling of photoreceptor proteins. Photochem Photobiol 97(2):243–269
92. Inoue K, Marín MdC, Tomida S, Nakamura R, Nakajima Y, Olivucci M, Kandori H (2019) Red-shifting mutation of light-driven sodium-pump rhodopsin. Nat Commun 10(1):1993
93. Shen M-Y, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Sci 15(11):2507–2524
94. Kandori H (2020) Retinal proteins: photochemistry and optogenetics. Bull Chem Soc Jpn 93(1):76–85
95. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic Acids Res 32:66–667
96. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res 35(suppl–1):522–525
97. Le Guilloux V, Schmidtke P, Tuffery P (2009) Fpocket: an open source platform for ligand pocket detection. Bioinformatics 10(1):168–179
98. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open babel: an open chemical toolbox. J Cheminform 3:33
99. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O (2011) MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J Comput Chem 32(10):2319–2327
100. Gowers RJ, Linke M, Barnoud J, Reddy TJ, Melo MN, Seyler SL, Domanski J, Dotson DL, Buchoux S, Kenney IM, Beckstein O (2016) MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. In: Sebastian B, Scott R (eds) Proceedings of the 15th Python in science conference, pp 98–105
101. Hunter JD (2007) Matplotlib: a 2d graphics environment. Comput Sci Eng 9(3):90–95
102. McKinney W (2010) Data structures for statistical computing in Python. In: van der Walt S, Millman J (eds) Proceedings of the 9th Python in science conference, pp 56–61
103. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE (2020) Array programming with NumPy. Nature 585(7825):357–362
104. ...Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P (2020) SciPy 1.0 contributors: SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17:261–272
105. O'boyle NM, Tenderholt AL, Langner KM (2008) cclib: a library for package-independent computational chemistry algorithms. J Comput Chem 29(5):839–845
106. Marín MdC, De Vico L, Dong SS, Gagliardi L, Truhlar DG, Olivucci M (2019) Assessment of MC-PDFT excitation energies for a set of QM/MM models of rhodopsins. J Chem Theory Comput 15(3):1915–1923
107. Loco D, Lagardère L, Caprasecca S, Lipparini F, Mennucci B, Piquemal J-P (2017) Hybrid qm/mm molecular dynamics with amoeba polarizable embedding. J Chem Theory Comput 13(9):4025–4033

Authors and Affiliations

Laura Pedraza-González[1,3] · Leonardo Barneschi[1] · Daniele Padula[1] · Luca De Vico[1] · Massimo Olivucci[1,2]

Leonardo Barneschi
leonardo.barneschi@student.unisi.it

Daniele Padula
daniele.padula@unisi.it

[1] Dipartimento di Biotecnologie, Chimica e Farmacia, Università degli Studi di Siena, Via Aldo Moro 2, 53100 Siena, Italy

[2] Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403, USA

[3] Present Address: Department of Chemistry and Industrial Chemistry, University of Pisa, Via Moruzzi 13, 56124 Pisa, Italy