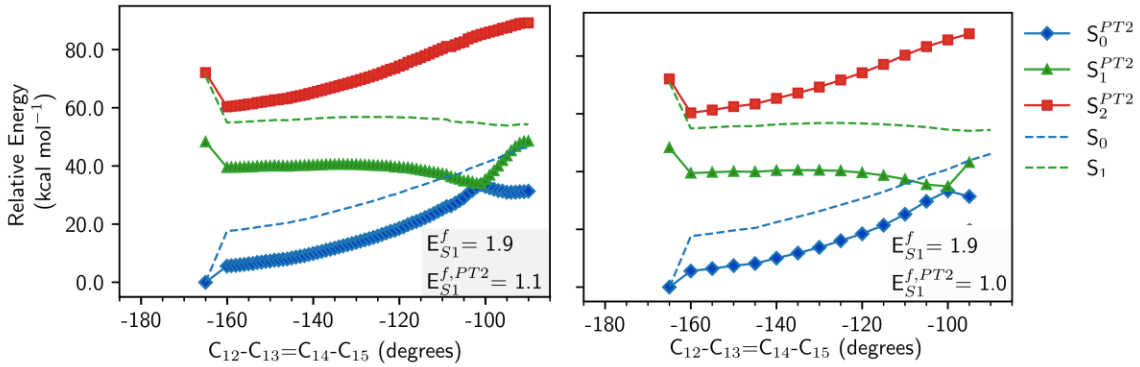

Figure 4.20: Effects of the on the RS of the step size, relative to the constrained isomerization torsional angle variation. The RSs for Arch7 were computed by using a step size of (left) 1 degree (72 constrained optimizations) and (right) 5 degrees (16 constrained optimizations).

ergies for each computed structure along the isomerization path and 3) the graphical analysis of the produced results.

- *Flexible setup:* The user can easily modify some default calculation parameters, such as the number of states to be evaluated in the SA CASSCF optimizations or CASPT2 energy correction, as well as the step size for changing the reactive torsional dihedral, by including optional arguments in the command-line interface (see Figure A.9).
- *Performance:* The module allows parallel screening and classification of large arrays of rhodopsins in a reasonable time.
- *Low dependency of the path profile and E^{f}_{S1} on the step size:* Benchmark calculations have demonstrated that the best compromise between quality of the computed RS and consumption of resources is a step size of 5 degrees. As illustrated in Figure 4.20 for the specific case of Arch7, the use of a step size of 5 degrees (16 constrained optimizations) vs a step size of 1 degree (72 constrained optimization) provides the same results at a much cheaper computational cost.

Limitations and pitfalls:

- *Discontinuity:* In certain cases, the S_1 isomerization path displays discontinuities related to a too fast change in the reaction coordinate, with consequently sudden changes in the energy profiles. This is due to the fact that the Relaxed scans are not computed in rigorous terms, such as it would be possible when employing an Intrinsic Reaction Coordinate. However, this is a forced choice necessary to expedite the computation and avoid the difficulties of the required transition state optimization. Remarkably, the discontinuities are often observed when the analyzed system presents a S_1-S_2 states mixing.

Finally, I remark that most of the methodological pitfalls exhibited for each phase are successfully covered by the application of a subsequent phase. Therefore, I conclude that the execution of the whole protocol for each variant would provide complementary information on their fluorescent behavior, rather than the use of every single phase as a classificatory filter. Even though the former choice may increase the computational time, it is not a

"bottleneck" when all the systems are analyzed in parallel.

Chapter 5

Applications: Modeling ground- and excited-state properties

Although in any show (i.e., theater play, tv serial, movie) the actors are the ones who take all the credit for their performance in front of the public, we must not forget that behind the stage there are all the ingenious people who make such show possible. This can be seen as a metaphor for what actually happens in the case of development of computational tools, where the people wearing big glasses (like in my case) and doing theoretical developments and computational implementations are behind the scenes (or the computer screen) making applications possible. During my doctoral studies, I was fortunate to be on both sides of the stage. More specifically, in this Thesis I present not only the methodological/computational development reviewed in previous chapters, but also many applications that require a photochemical interpretation more in line with my background in theoretical and computational chemistry. Let's see the protagonist of our story, the rhodopsins, in action!

Initial personal remarks

In Chapter 3, I have described the main features of both the command-line and Web-interface versions of the *a*-ARM rhodopsin model building protocol. Furthermore, I proposed two ARM-based protocols for specific QM/MM analysis of the produced Ground-state `ARM QM/MM models` and, thus, the study of absorption properties. Likewise, in Chapter 4 I proposed a protocol for modeling excited-states structures that allows the analysis of emission and photodynamics properties. In this Chapter, I will, instead, show that the ground-state and excited-state `ARM QM/MM models` and properties generated with these protocols, are suitable for diverse applications, yielding results that are, in principle, relevant for the development of optogenetics tools. On the other hand, as a final remark, I would like to express my belief that the presented work has created a new automated computational infrastructure that will facilitate the further refinement and accuracy enhancement of ARM models. Accordingly, I believe the presented work constitutes a further step towards the design of an automated computational tool applicable to biologically or technologically important photoresponsive proteins.

In this Chapter, I will report on the performance of the four ARM-based protocols proposed in this Thesis (see Chapters 3 and 4), highlighting both their methodological and scientific capabilities as well as their current limitations. To do so, I and my collaborators have constructed and employed a **benchmark** set of about 150 wild-type and mutant rhodopsins, as well as carried out selected **applications**, directed to the prediction of trends in light-induced properties, including absorption and emission spectra, as well as excited-state molecular dynamics. Such trends unveil different mechanistic aspects of color tuning and fluorescence emission, as well as, more in perspective, the systematic prediction of photoisomerization quantum yields. As it will be specified, during the execution of such research works, I acted as the main researcher, co-supervisor, or collaborator. In this regard, I would like to stress that not all presented results come from the original work done for my doctoral research. However, I mention such works since i) they are a consequence of the development of *a*-ARM and ii) during their execution I brought either scientific or technical support. In order to distinguish between my work and the work performed by my collaborators, I will use a different color label in the "color boxes" to specify my role, as follows:

- Research carried out as original work of this PhD Thesis **(main researcher)**
- Research work carried out as either original work and part of a Master Thesis **(co-supervisor)**
- Research work carried out as part of a Master Thesis **(collaborator)**
- Research work carried out by another researcher **(collaborator)**

The content of the Chapter is divided into three sections, where the different studies are classified into the following categories: i) mutants generation and color tuning analysis (section 5.1), ii) absorption bands and photodynamics simulations (section 5.2) and iii) emission properties simulation and engineering of fluorescent rhodopsins (section 5.3). These three areas represent applications of the *a*-ARM protocol for i) producing reliable QM/MM models, ii) using such models as starting point for ground state populations and subsequent photodynamics, and iii) as starting points for excited states potential energy surfaces exploration, respectively.

5.1 Mutants generation and color tuning analysis

The possibility of performing fast and parallel screenings of large arrays of rhodopsin variants offered by the *a*-ARM protocol, allows different applications of the generated Ground-state `ARM QM/MM models`, including the systematic comparative study of absorption properties. This, in turn, allows to perform rational analysis of the color tuning mechanisms. In the following, I will present the most recent research on color tuning performed in the LCPP, where I and my collaborators have used WT QM/MM models, that reproduce experimental trends in λ^{a}_{max}, for the further *in silico* generation of mutants (see Section 3.1.4) and the posterior analysis of color tuning mechanism (see Section 3.1.3).

5.1.1 Screening Mouse Melanopsin Color-Tuning mutations

Thesis 4

This section presents the main findings of the Master's Thesis carried out by the student Filippo Sacchetta, under the supervision of Professor Massimo Olivucci and co-supervision of Dr. María del Carmen Marín. **(2018)**.

Contribution: I produced the `ARM QM/MM models` for the WT variants and provided the *a*-ARM technology for the generation of the mutants. In addition, I contributed with technical/scientific support during the realization of the project and the discussion of the results.

Recent studies have revealed that Rods and Cones are not the only two types of photoreceptor cells in mammalian retina.[132–134] Indeed, a third type was identified, called intrinsically photosensitive retinal ganglion cells (ipRGCs). These cells express an opsin-like protein called *Melanopsin*, and provide input to regulate circadian activity. Among various functions, Melanopsins control growth processes and support visual functions of the eye by sensing radiance levels (*e.g.*, pupil constriction) and recognizing contrast and brightness.[132–134, 134–136] Therefore, the study of such opsins would allow, *e.g.*, to gain further insights into the understanding of how the endogenous body clock is regulated by light. In fact, experimental investigations support that these biological functions, exhibited by melanopsins, have an impact on optogenetics applications.[137–139]

Accordingly, the research work performed in the master's Thesis of Sacchetta attempted at the QM/MM modeling of Mouse melanopsin (mMeOp) variants. More specifically, the study consisted in the *in silico* searching of mutants featuring "blue-shifted" absorption properties, that are widely desired for the development of optogenetic tools. I would like to remark the fact that, chronologically, this was the first actual application of the *a*-ARM rhodopsin model building, proposed in this doctoral Thesis (see Section 3.1.1). In the following, I will summarize the main achievements of the study.

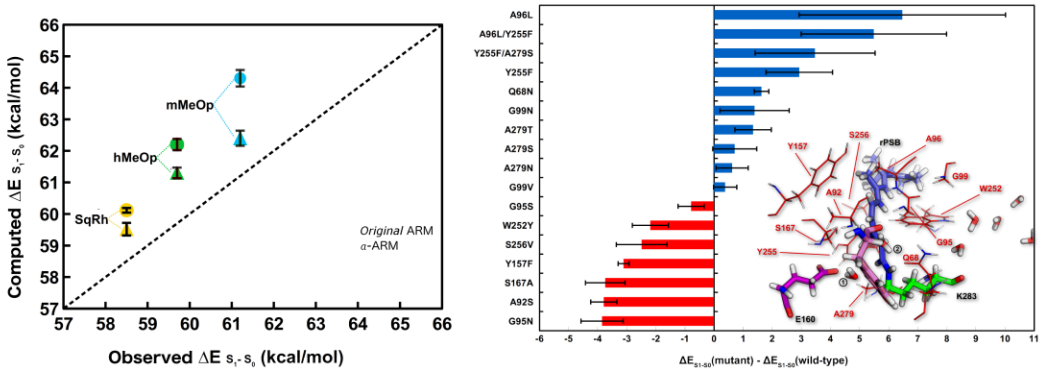

Figure 5.1: Overview of the master's Thesis of Sacchetta et. al. (left) Trends in ΔE^{a}_{S1-S0} using both *original* (circles) and *a*-ARM (triangles) versions of the protocol. (right) Difference of computed ΔE^{a}_{S1-S0} between each of the 17 mutants of mMeOp and the WT value.

In the first part of the work, the quality of the `ARM QM/MM model` for WT mMeOp was assessed by comparing it with both its homologous Human melanopsin (hMeOp)[135] and the invertebrate Squid rhodopsin from *Todarodes pacificus* (SqRh)[1],[85] in terms of the reproduction of the experimental trend in λ^{a}_{max} (*i.e.*, mMeOp = 467 nm < hMeOp= 479 nm < SqRh= 489 nm). The left panel of Figure 5.1 illustrates the trend in λ^{a}_{max} computed with both *original* ARM (see Section 2.2) and *a*-ARM (see Section 3.1.1) versions. As observed, in both cases the trend is reproduced. However, it is possible to notice a slight improvement achieved with *a*-ARM, where the three produced models feature almost the same error bar with respect to experimental ΔE^{a}_{S1-S0}. Moreover, the analysis of the total CASSCF S_0 energies for the QM/MM models suggest that *a*-ARM provides energetically more stable structures. This fact could be attributed to the more rational assigning of the counterion placement, that as reported in paper **[I]**, is based on an energy optimization criterion instead of manual arbitrary placement.

In the second part of the work, an exhaustive study of the chromophore cavity (see inset structure in the right panel of Figure 5.1) for the WT was carried out, in order to identify promising mutations with enhanced sensitivity of the mMeOp *r*PSB receptor to blue light. As a result, a preliminary list of 17 single and double mutants with possible blue-shifting effect, with respect to WT, was identified. Then, as illustrated in Figure 5.1 (right), the corresponding *a*-ARM$_{\text{default}}$ models were generated and the λ^{a}_{max} evaluated, producing 10 variants with blue-shifting from 0.5 to 7.0 kcal mol^{-1}. Such color tuning effect was further evaluated by analyzing the electrostatic and steric contributions to the ΔE^{a}_{S1-S0}, previously introduced in Section 3.1.3. Since during the production of Sacchetta's Thesis the automatic protocol for such analysis was not yet implemented, all the associated calculations were performed manually.

In conclusion, further analysis of the described results would contribute to the understanding of specific "rules" for controlling color tuning mechanisms in melanopsin pho-

[1]Since melanopsins exhibit a great primary sequence similarity with invertebrate rhodopsins, such as Squid rhodospin (SqRh) (39%),[135] both homology models for mMeOp and hMeOp where generated using the X-ray structure of SqRh as a template.[61, 62]

toreceptors. Accordingly, as a perspective, future research will be devoted to the fast and systematic searching of blue-shifted variants for both mMeOp and hMeOp. Notice that, currently, such objective can be easily achieved with the use of the automatic tools implemented in the `ARM` package. This is, certainly, an achievement of the methodological/computational advancements of my doctoral Thesis.

5.1.2 Towards a Comparative Computational Photobiology: Invertebrate Rhodopsin Pigments

Thesis 5

This section presents the main findings of the Master's Thesis carried out by the student Laleh Allahkaram, under the supervision of Professor Massimo Olivucci and co-supervision of Professor Luca De Vico and Dr. María del Carmen Marín.

Contribution: I contributed to the development of the Web-ARM technology employed for the generation of the QM/MM models. In addition, I provided technical/scientific support during the realization of the project and the discussion of the results.

Animals sense light across a broad range of wavelengths, from ultraviolet (UV) to far-red, using visual pigments rhodopsins and related photosentitive proteins[140, 141]. These systems, referred to as G-protein-coupled receptor (GPCR) in Section 1.1.3, act as signal transducer converting external stimuli into cellular responses in either visual and non-visual processes.[142] It is well-known that the the main counterion (MC), a negatively charged amino acid residue that stabilizes a positive charge on the *r*PSB (see Figure 2.1), has an essential structural and electronic role on the primary process of absorption of visible light and, consequently, in the light-induced photoisomerization. Therefore, a thoughtful analysis of the structural differences (*e.g.*, type of amino acid and position) of such counterion in sets of phylogenetical diverse Type-II (*e.g.*, vertebrate and invertebrate) rhodopsins, would provide insights into the molecular evolution of the rhodopsin family.

The work performed in the master's Thesis of Allahkaram consisted on a comparative study of seven phylogenetically diverse WT animal (type II) rhodopsins, selected on the basis of the following three criteria: i) belong to a variety of organisms, ii) feature diverse protein function, and iii) have available experimental λ^{a}_{max}, in a wide range of 473 to 535 nm. Thus, the *set* was composed of the vertebrate Bovine rhodopsin from *Bos taurus* (Rh) and Human melanopsin (hMeOp), and the invertebrate Scallop 1 (Scop1) and 2 (Scop2), Squid rhodopsin (SqRh), Box Jellyfish (JellyOp) and Jumping spider rhodopsin-1 (JSR1). Such a comparison is expected to provide information on the differences in spectral and functional properties of these evolutionary distant pigments.

In the first part of the work, the corresponding `ARM QM/MM models` were constructed

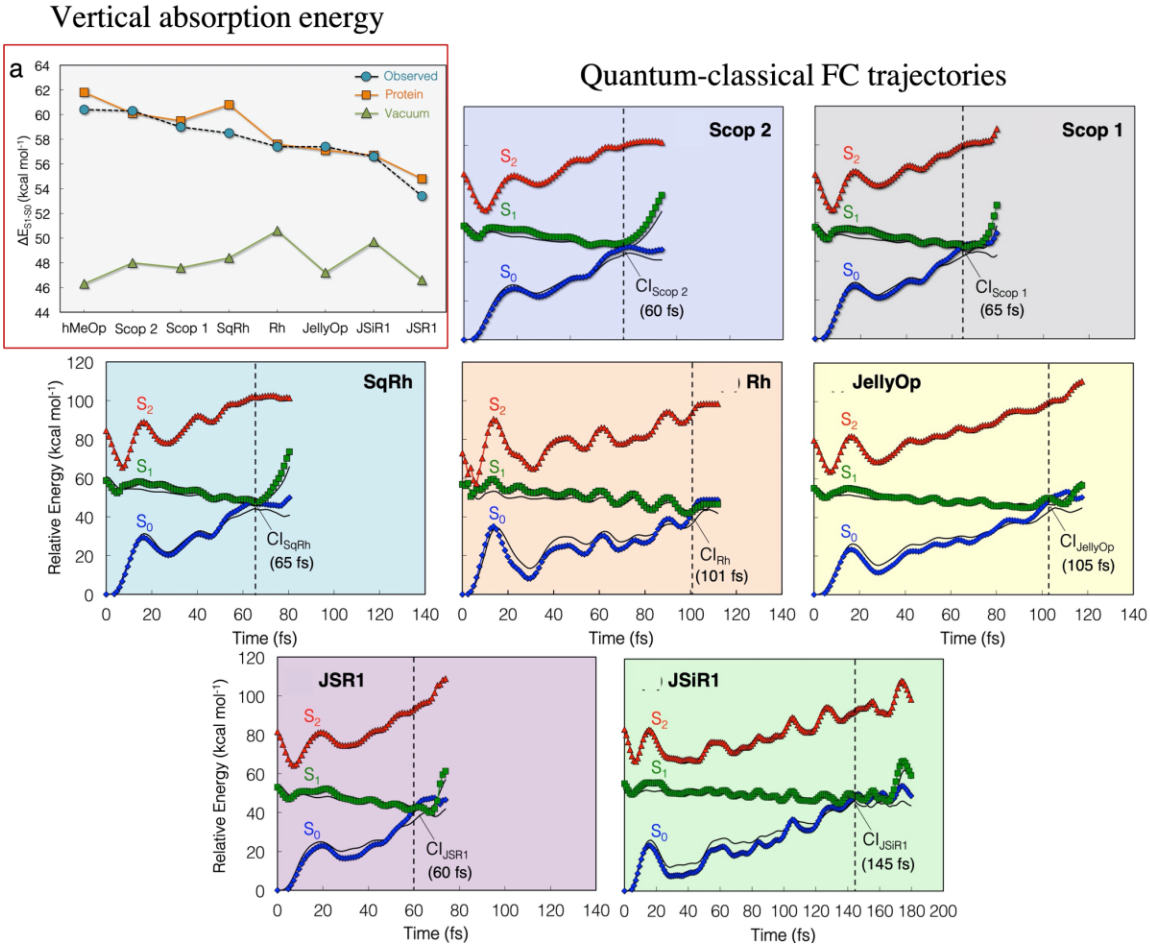

Figure 5.2: Overview of the master's Thesis of Allahkaram (top-left box) Computed trend in Vertical Excitation energy (ΔE^{a}_{S1-S0}); the other panels report the FC trajectory computation on the First excited-state (S_1) state for each members of the *set*, computed at two-root state-averaged-CASSCF/6-31G*/Amber level of theory (black full lines) and corrected at the CASPT2 level. S_0 (blue diamonds), S_1 (green squares), and S_2 (red triangles). The estimated ESL are also presented.

by using the Web-ARM interface, introduced in Section 3.1.2, and the computed Vertical Excitation energy were evaluated.

As observed in Figure 5.2 top-left panel, the reproduction of the experimental trend in ΔE^{a}_{S1-S0} was successfully achieved by using six a-ARM$_{\text{default}}$ models and just one a-ARM$_{\text{customized}}$ model (*i.e.*, JSR1), obtained trough a different setup for the protonation states. These results confirmed that the Web-ARM interface produces `ARM QM/MM models` of the same quality exhibited by the command-line tool.

Subsequently, a color tuning analysis that considered all the residues belonging to the chromophore cavity (as described in Section 3.1.3) was performed on each rhodopsin, while keeping the well-studied Rh as reference. This helped to elucidate the structural and electrostatic effects, due to interactions between the retinal chromophore and the protein environment, on the color of the proteins (*i.e.*, λ^{a}_{max}). Briefly, the results of such analysis suggested that the different colors observed among the rhodopsins of the *set* are a product of a combination of steric and electrostatic effects due to the different protein environments (*i.e.*, large differences in the counter-ion complex of the rPSB), which modulate the absorp-

tion of the retinal chromophore. However, direct and indirect electrostatic components (see Section 3.1.3) dominate over the steric one.

On the other hand, in order to simulate the excited state dynamics and roughly estimate the Excited state lifetime (ESL) for each rhodopsin, quantum-classical Franck-Condon trajectories (see Section 4.1.2) were performed over the S_0 structures. Interestingly, as shown in Figure 5.2, it was found a relationship between the λ^a_{max} and the ESL, where the most blue-shifted variant (hMeOp) exhibits the lowest ESL. As a perspective work, these preliminary results should be verified by the generation of a series of initial conditions to study the population dynamics of each rhodopsin and, thus, obtain a more reliable estimation of the ESL. Although when Allahkaram's Thesis was written such objective was technically difficult to achieve (*i.e.*, manually handled), currently we have available the protocol for the automatic analysis of photodynamics population, implemented in the master's Thesis of Bonfrate. As we will see in Section 5.2.1, such recently implemented computational tool allows the production of the initial conditions and the computation of QY, featuring a one-click architecture.

This work represents another example of why the level of automatization achieved in my doctoral Thesis is relevant for producing standard and congruous studies that rely on QM/MM modeling of rhodopsins.

5.1.3 «Paper [V]» Role of Pro219 as an Electrostatic Color Determinant in the Light-driven Sodium Pump KR2: Combined spectroscopic and QM/MM modelling studies

Submitted 6

The content of this section is an overview of the main findings reported in paper **[V]**. (Manuscript submitted)

Contribution: This is an original result of the research work carried out in this doctoral Thesis. Since this research is a combined spectroscopic and computational study, I remark that my contribution consisted in the design and execution of the computational experiments. More specifically, I carried out i) the methodological development and code implementation of the employed computational protocols (see Sections 3.1.1, 3.1.4, 3.1.3), ii) the corresponding calculations for the QM/MM modeling of WT and mutant variants, and iii) the analysis of the produced results.

In this work, produced as original material of this doctoral Thesis, I and my collaborators have reported on a combined spectroscopic and computational study that consists of an exhaustive investigation of the color tuning induced by point mutations at the P219 position in the light-driven sodium pump Krokinobacter rhodopsin 2 from *Krokinobacter eikastus* (KR2)[21, 78, 143, 144] (see Figure 5.3 (left)). This study was carried out in collabora-

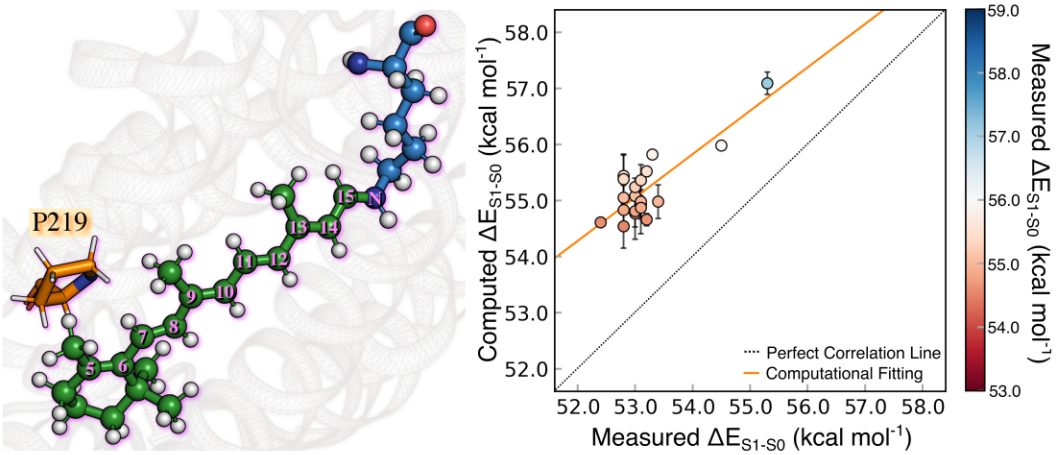

Figure 5.3: Overview of the main results reported in paper [V]. (left) The residue P219 is located near the β-ionone ring. (right) Correlation plot between computed and measured ΔE^{a}_{S1-S0} for WT-KR2 rhodopsin and 19 of its mutants (P219X, with X= A, C, D, E, F, G, H, I, K, L, M, N, Q, R, S, T, V, W, Y).

tion with the experimental research group of Professor Kandori at the Nagoya Institute of Technology.

The aim of the work was to investigate how all possible P219X mutants, where X stands for all alternative 19 natural amino acids, modulate the λ^{a}_{max} value of the reference WT (*i.e.*, 525 nm) and understand the molecular-level mechanism driving color tuning. The first part of the work consisted on the experimental expression of all variants, followed by the subsequent measurement of their corresponding absorption spectra.[2] As shown in Figure 5.3 (right) and further detailed in paper [**V**], the resulting color variations induced by the 19 mutants span a red-to-blue range going from 546 nm to 517 nm.

Remarkably, most of the mutants induce a red-shifting effect, with exception of P219R which λ^{a}_{max} is located at 517 nm. This is a first indication that R (Arg) is positively charged even in the hydrophobic environment surrounding position 219. Indeed, consistently with the color tuning theory mentioned in Section 3.1.3, such positive charge, placed near the β-ionone ring of the protonated *r*PSB chromophore, would cause a spectral blue shift. More interestingly, transport activity assays demonstrated that sodium and proton pump activities characteristic of the WT are maintained for all the red-shifted mutants but not for the only blue-shifted P219R. This experimental fact suggests that the addition of a positive charge at that specific position, near the β-ionone ring of the *r*PSB, is responsible for the lack of ion transport. In addition, both absorption spectra and transport activity assays suggest that other amino acid replacements lead to neutral side-chain. In fact, it is reasonable for D (Asp), E (Glu), and H (His) to be neutral in the hydrophobic environment. In addition, K (Lys) may also be neutral in the P219K mutant. Remarkably, all these assigned protonation states are consistent with those of the `ARM QM/MM models`.

The second part of the work, that corresponds to my actual contribution, consisted of the QM/MM modeling of WT-KR2 and its 19 mutants. To this aim, I used the *a*-ARM$_{\text{customized}}$ approach described in Section 3.1.1 for the generation of the WT-KR2 model, that presents

[2]The experimental part was carried out by the research group of Professor Kandori.

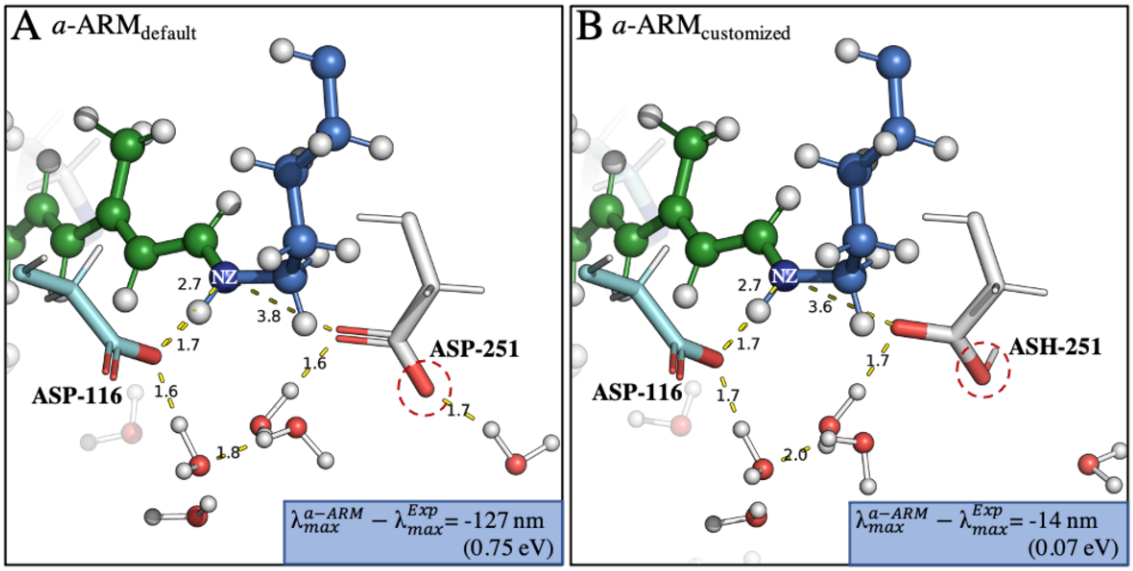

Figure 5.4: KR2 [PDB ID 6REW] model customization. Default model generated with the *a*-ARM$_{default}$ approach (A) and customized model produced with the *a*-ARM$_{customized}$ approach (B). The customization was performed by i) predicting the protonation states for ionizable residues at pH 5.2, and ii) neutralizing the secondary counterion of the *r*PSB (ASP-251 → ASH-251). QM/MM-optimized models, with hydrogen bonds represented as dashed lines.

a $\Delta^{Exp}_{calc}\Delta E^{a}_{S1-S0}$ of just 1.4 kcal mol^{-1}. As further explained in paper [**V**] (Appendix B.5), the customization procedure consisted on changing the protonation states pattern of the counter-ion complex of the *r*PSB, by turning neutral the secondary counterion D251 (see Figure 5.4). The setup of the customized WT-KR2 was employed as a template for the generation of the `ARM QM/MM models` of each of the 19 mutants. As a result, 16 out of the 19 models presented $\Delta^{Exp}_{calc}\Delta E^{a}_{S1-S0}$ inside the error bar of the protocol (*i.e.*, 4.0 kcal mol^{-1}). The other 3 models, that correspond to P219H, P219K and P219R were further customized by changing the protonation states. In summary, the protonation states setup of the twenty QM/MM customized models is: neutral D251, E160 for WT and P219X with X=A, C, F, G, I, L, M, N, Q, S, T, V, W, Y; neutral D116 for P219R; neutral D251, E160, H219 (with hydrogen in the ε nitrogen) for P219H; neutral D251, E160, K219 for P219K; neutral D251, E160, E219 for P219E; and neutral D251, E160, D219 for P219D.

Moreover, the side-chain conformation for each mutant was selected by employing the experimentally-driven automatic procedure described in Section 3.1.4. As mentioned before, that procedure consisted in evaluating the `ARM QM/MM model` performance for a set of automatically chosen rotamers (see Figure 3.9) and select the one that better reproduces the observed ΔE^{a}_{S1-S0} value. Figure 5.5 (top) plots the difference between calculated and observed vertical excitation energy for each of the three produced rotamers, while Figure 5.5 (bottom) shows the selected one. As observed, rotamer 1 is selected for 9 out of the 19 cases. In addition, 7 out of the 10 cases in which either rotamer 2 or 3 is selected, present ΔE^{a}_{S1-S0} values that differ in less than 1.0 kcal mol^{-1} with respect to rotamer 1, suggesting that the eventual choice of the latter could provide results inside the trend in absorption

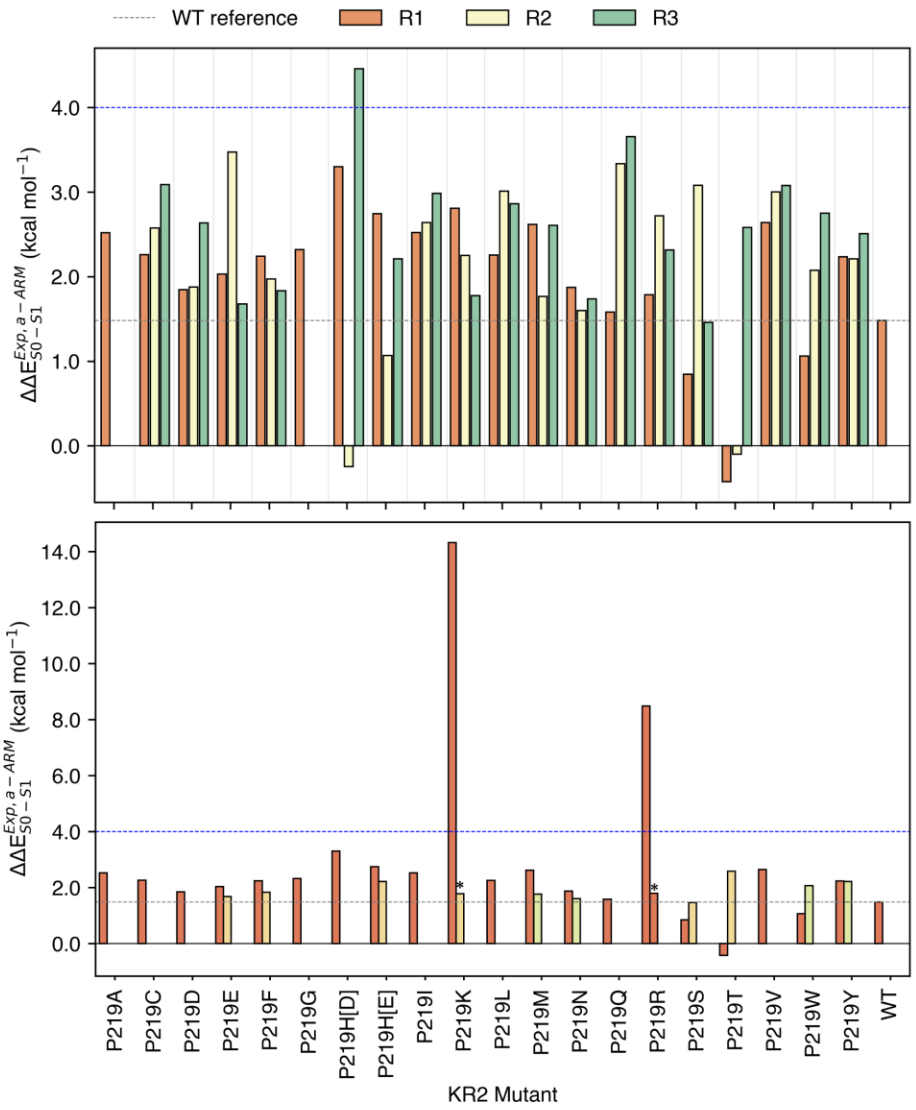

Figure 5.5: Schematic representation of the selection of side-chain conformations for the P219X mutants of KR2. (top) Difference between calculated and observed vertical excitation energy ($\Delta^{Exp}_{calc}\Delta E^{a}_{S1-S0}$) for each of the three default rotamers generated by Modeller, for the 19 P219X mutants of KR2. (bottom) $\Delta^{Exp}_{calc}\Delta E^{a}_{S1-S0}$ for the default (first bar) and customized models (second bar). The default models have the same protonation states than the WT template and their side-chain is modeled with the rotamer 1 (dark orange bars). The side-chain of the customized models could be modeled with the rotamer 2 (green bars) or the rotamer 3 (light orange bars). Most of the customized models exhibit the same protonation states than the WT, with exception of P219K and P219R marked with a star.

energy. However, it is evident that our strategy for rotamers selection is suitable only for cases when experimental data is available. As pointed out in Section 3.1.4, this is one of the main pitfalls of our strategy for mutants generation. For future works aimed to the *in silico* design of mutants, a predictive tool is required.

The ΔE^{a}_{S1-S0} values computed for the above described customized QM/MM models, reproduce the experimental trend in λ^{a}_{max} (see Figure 5.3 (right)), confirming that the P219R variant is the only one causing a blue-shifting effect among the set. In order to elucidate the steric and electrostatic effects modulating the color tuning on the position P219X, I use the automatic protocol introduced in Section 3.1.3 implemented as the `a_arm_ret` and `a_arm_turn_off` modules of the `ARM` package described in Chapter A. As a result, I found

that the 219 position induces color changes through, essentially, electrostatic effects, while steric effects only partially contribute to the single, largely blue-shifting P219R mutant. The analysis of the model shows that the set fully support the color tuning theory revised above and support the hypothesis of the existence of an electrostatically-driven proline-based switch.

In conclusion, the computational analysis of the experimental results via QM/MM modeling suggests that in KR2 the P219X switch operates by inducing variations in the electrostatic interactions between chromophore and environment, while the geometrical distortion of the chromophore caused by the corresponding point mutation does not appear to play a major role. It has also been shown that both direct (the change in the electrostatic field due to the residue replacement at position 219) and indirect effects (the changes due to all other cavity reorganization induced by the replacement and including chromophore reorientation, side-chain and water relocations and the modification of the hydrogen bond network) seem to contribute to such electrostatic determinant of the color tuning.

As a perspective work, similar color tuning analysis over specific amino acids surrounding the *r*PSB, would allow to the *in silico* identification of different color switches determining the color on KR2 and, in turn, useful in optogenetics. A preliminary investigation, consisted in the *a*-ARM QM/MM modeling for a set of 53 KR2 variants (including the 19 analyzed above), following the same computational strategy reported above for the P219X variants. In such application, I have selected variants of six positions (*i.e.*, G153, P219, Y218, D251, D116, D254) with available experimental λ^{a}_{max}.[29] The preliminary results, illustrated in Figure 5.6, show that the *a*-ARM$_{\text{default}}$ is able to reproduce the experimental trend of about of 80% of the cases, whereas the resting 20% are successfully corrected by using the *a*-ARM$_{\text{customization}}$, my modulating either the protonation states or the side-chain conformation of the mutated residue. In future works, such QM/MM models will be used for the analysis of the color tuning mechanism on alternative positions.

5.2 Absorption bands and photodynamics simulations

Another application of the ground-state `ARM QM/MM models` is the possibility of predicting the absorption bands associated with the One-Photon Absorption and Two-Photon Absorption phenomena described in Section 1.1.3. In the following, I will introduce an automatic protocol for the production of OPA spectra and the generation of the initial conditions for the simulation of photodynamics, that allows the estimation of photoisomerization quantum yield. Then, I will show three studies in which the OPA is computed for a representative member of animal, microbials and heliorhodopsin types, respectively. Furthermore, I will present a study where the TPA spectra of bovine rhodopsin is computed.

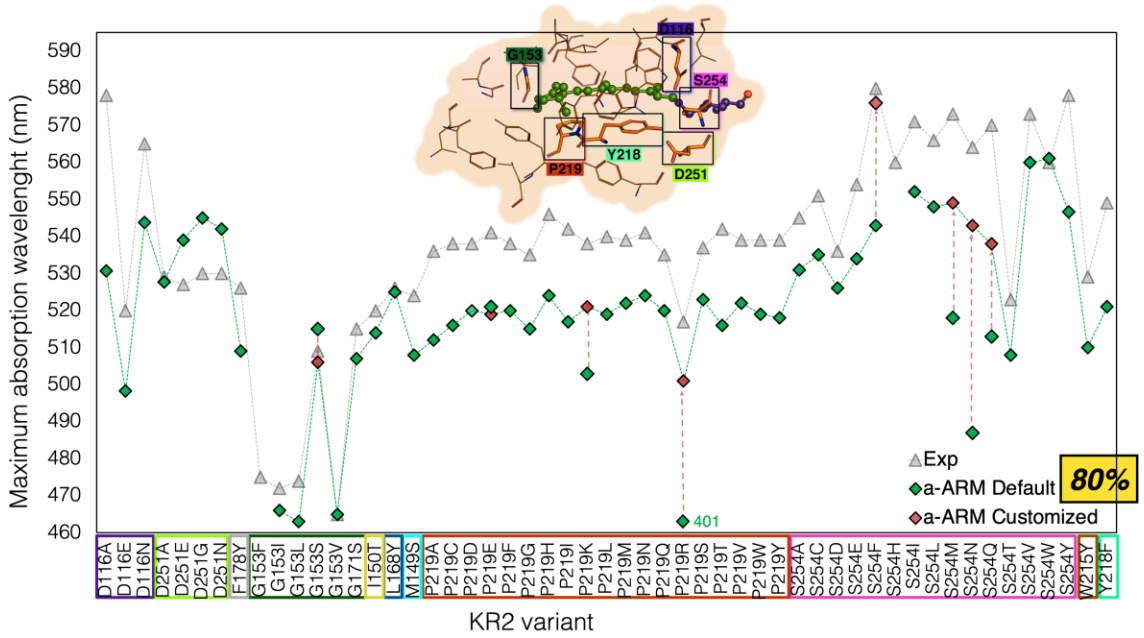

Figure 5.6: Preliminary screening of the KR2 cavity. Trend in Maximum absorption wavelength (λ^a_{max}) for 53 mutants of KR2.

5.2.1 Analysis of Absorption Bands and light-induced Dynamics of Rhodopsins through a QM/MM-based Automatic Protocol

Thesis 7

This section presents the main findings of the Master's thesis carried out by the student Simone Bonfrate, under the supervision of Professor Massimo Olivucci and co-supervision of Ph.D. candidate Laura Pedraza-González (me) and Professor Luca De Vico.

Contribution: I contributed to the methodological design, supervision and execution of the general project. More specifically, I brought technical/scientific support for i) the methodological development and code implementation of the computational protocol presented below, inside the `ARM` package, ii) the corresponding calculations, and iii) the analysis of the produced results.

In the research work carried out during the master's studies of Bonfrate, we have performed the blueprinting, implementation and application of a computational protocol, to handle the automatic simulation of the OPA spectrum (*aka*, absorption band), and the subsequent production of initial conditions for the study of the excited state dynamics, also allowing to estimate the Photoisomerization quantum yield (QY)

The general workflow of the four-phases protocol, illustrated in Figure 5.7, takes inspiration from a methodology previously designed in our laboratory LCPP. The main drawback of such strategy was its actual implementation, consisting in a series of bash-sell scripts not

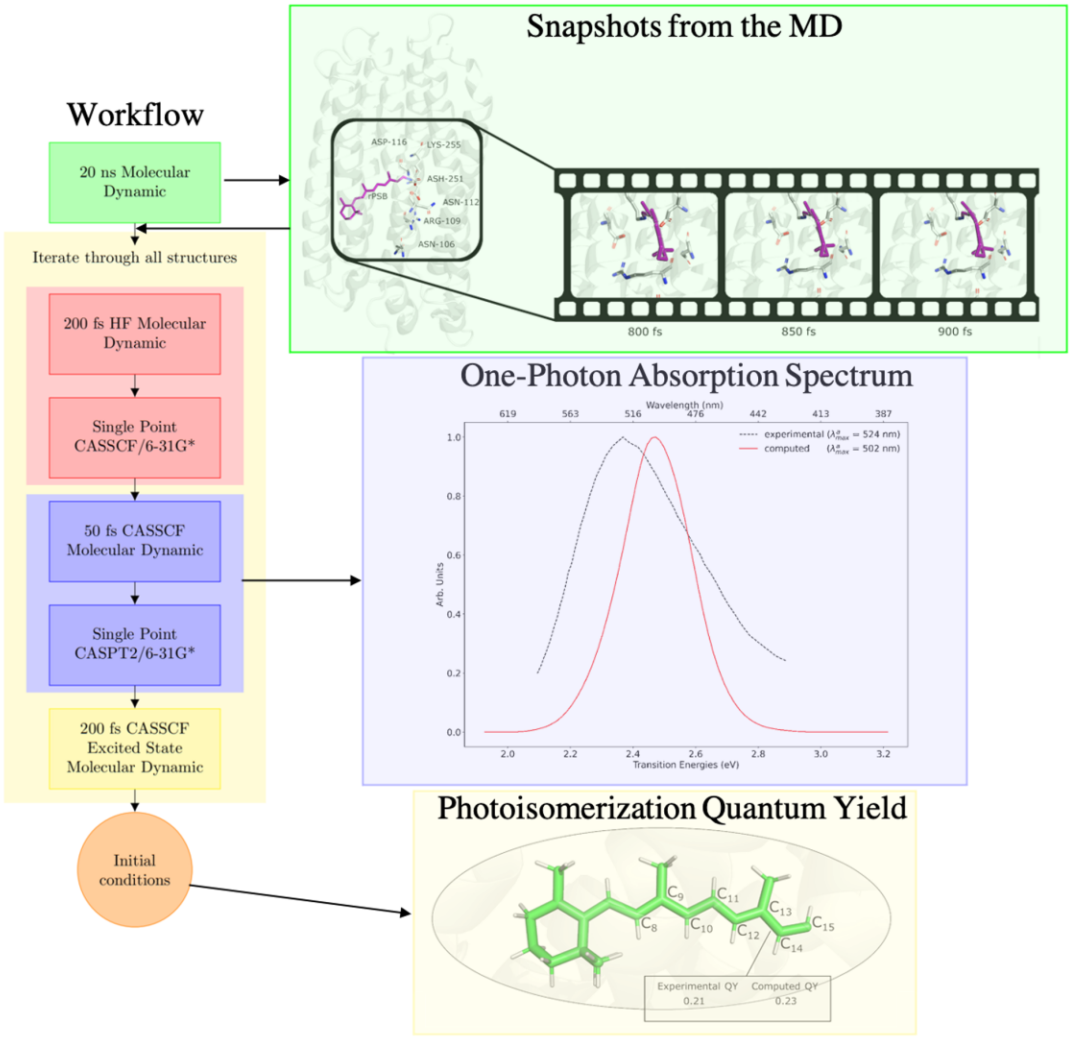

Figure 5.7: `a_arm_opa` driver. The first phase (green box) consists of the generation of a long MD, taking as input the equilibrium S_0 `ARM QM/MM model`, and the subsequent extraction of the representative structures or "snapshots". Then, each of the snapshots (*i.e.*, 200) is processed by the second phase (red boxes) that consists of a Hartree-Fock MD. The output geometries and velocities are used as an input for the third phase that consists of ground-state CASSCF MD that allows the prediction of the OPA. Then, the output geometries and velocities are processed by the fourth phase, where a excited-state MD is performed to generate the initial conditions. Finally, the Photoisomerization quantum yield is computed.

interconnected by a general driver and difficult to manage. As a result, to produce a single OPA spectrum, *e.g*, simulated with 200 initial structures, a specialized user had to handle at least 1000 calculations ensuring the adequate parsing of input/output files through the phases. Moreover, there was no documented benchmarking of the parameters yielding the best compromise between accuracy and computational cost (*e.g.*, number of analyzed structures, wall-time, memory). With its implementation as the `a_arm_opa` driver into the `ARM` package, we have achieved a substantial simplification of the protocol, up to being reduced to the execution of a single command-line, which receives the specification of the required parameters in a single file.

First, each phase of the protocol described in Figure 5.7, was benchmarked to establish the parameters yielding the best compromise between accuracy and computational cost. Consequently, the best setup for the OPA simulation is obtained by employing an ensemble

of 200 snapshots extracted from a MD of 20 ns. Then, the protocol was applied to the microbial rhodopsin WT-KR2, introduced in Section 5.1.3. As shown in the right-middle panel of Figure 5.7, the calculated OPA is in good agreement with the experimental one reported by Tahara et al.. Most importantly, as shown in the right-bottom panel, the simulation of the photoisomerization Photoisomerization quantum yield (QY), considering 300 excited-state trajectories, yielded a value of 0.23, which is in good agreement with experimental data reported by Tahara et al. as 0.20-0.22.[143]

In conclusion, the inclusion of this protocol in the `ARM` package is an important achievement of this doctoral Thesis, since it represents a computational tool for the systematic and automatic study of both absorption bands and photodynamic properties, without human intervention.

As a perspective, this tool will allow to perform complex comparative studies of the photodynamical properties for arrays of rhodopsins, as the one required in Section 5.1.2.

5.2.2 «Paper [IV]» Multi-State Multi-Configuration Quantum Chemical Computation of the Two-Photon Absorption Spectra of Bovine Rhodopsin.

Published 8

Most of the content of this Section is reproduced/adapted with permission from «Gholami et al., *J. Phys. Chem. Lett.* 2019, 10, 20, 6293-6300». Copyright 2019 American Chemical Society.

Contribution: I provided the `ARM QM/MM model` for bovine rhodopsin (Rh) and the computed λ^{a}_{max}. In addition, I contributed to the discussion of the results.

In this work we show that the ground-state *a*-ARM models, as those presented in the benchmark set in Section 3.1.1, can be successfully employed for the further computation of Two-Photon Absorption (TPA) spectra. The interest in achieving a computational tool for the prediction of TPA relies in the fact that, similar to red light absorption, degenerate two-photon absorption[145] provide a way to photoexcite a rhodopsin with infrared (IR) light, leading to an improved light penetration with respect to one-photon absorption with visible light and, possibly, maximizing the use of, for instance, bistable rhodopsin as optogenetics tools.[146] As previously discussed in Section 1.1.3 and illustrated in Figure 1.3, photoexcitation to the S_1 state can be achieved also by simultaneous absorption of two photons featuring the same energy ΔE^{TPA}_{S1-S0} and corresponding to half the energy necessary for One-Photon Absorption.

A complete description of this study is provided in paper **[IV]** and a further analysis is shown in Section 4.3 of paper **[III]**. Briefly, we employed the *a*-$ARM_{default}$ model to compute the TPA cross-section (σ_{TPA}) of wild-type bovine rhodopsin (WT-Rh). To